Flipper Games for Monadically Stable Graph Classes

Jakub Gajarský
University of Warsaw, Poland

Nikolas Mählmann
Universität Bremen, Germany

Rose McCarty
Princeton University, NJ, USA

Pierre Ohlmann
University of Warsaw, Poland

Michał Pilipczuk
University of Warsaw, Poland

Wojciech Przybyszewski
University of Warsaw, Poland

Sebastian Siebertz
Universität Bremen, Germany

Marek Sokołowski
University of Warsaw, Poland

Szymon Toruńczyk
University of Warsaw, Poland

Abstract

A class of graphs $\mathcal{C}$ is *monadically stable* if for every unary expansion $\widehat{\mathcal{C}}$ of $\mathcal{C}$, one cannot encode – using first-order transductions – arbitrarily long linear orders in graphs from $\widehat{\mathcal{C}}$. It is known that nowhere dense graph classes are monadically stable; these include classes of bounded maximum degree and classes that exclude a fixed topological minor. On the other hand, monadic stability is a property expressed in purely model-theoretic terms that is also suited for capturing structure in dense graphs.

In this work we provide a characterization of monadic stability in terms of the *Flipper game*: a game on a graph played by *Flipper*, who in each round can complement the edge relation between any pair of vertex subsets, and *Localizer*, who in each round is forced to restrict the game to a ball of bounded radius. This is an analog of the *Splitter game*, which characterizes nowhere dense classes of graphs (Grohe, Kreutzer, and Siebertz, J. ACM '17).

We give two different proofs of our main result. The first proof is based on tools borrowed from model theory, and it exposes an additional property of monadically stable graph classes that is close in spirit to definability of types. Also, as a byproduct, we show that monadic stability for graph classes coincides with monadic stability of existential formulas with two free variables, and we provide another combinatorial characterization of monadic stability via forbidden patterns. The second proof relies on the recently introduced notion of *flip-flatness* (Dreier, Mählmann, Siebertz, and Toruńczyk, arXiv 2206.13765) and provides an efficient algorithm to compute Flipper's moves in a winning strategy.

2012 ACM Subject Classification Theory of computation → Finite Model Theory; Mathematics of computing → Graph theory

Keywords and phrases Stability theory, structural graph theory, games

Digital Object Identifier 10.4230/LIPIcs.ICALP.2023.128

Category Track B: Automata, Logic, Semantics, and Theory of Programming

Related Version *Full Version*: `https://arxiv.org/abs/2301.13735` [11]

Funding This work is a part of projects CUTACOMBS (R. McCarty) and BOBR (J. Gajarský, P. Ohlmann, M. Pilipczuk, W. Przybyszewski, M. Sokołowski and S. Toruńczyk) that have received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreements No. 714704 and 948057, respectively). Nikolas Mählmann and Sebastian Siebertz are supported by the German Research Foundation (DFG) with grant agreement No. 444419611. Rose McCarty is also supported by National Science Foundation (NSF) grant No DMS-2202961.

1 Introduction

Monadic stability is a notion of logical tameness for classes of structures. Introduced by Baldwin and Shelah [3] in the context of model theory[1], it has recently attracted attention in the field of structural graph theory. We recall the definition below. One of the main contributions of this paper is to provide purely combinatorial characterizations of monadically stable classes of graphs via games and via forbidden patterns. Our game characterization is effective, and can be employed in algorithmic applications, as we explain later.

In this paper we focus on (undirected, simple) graphs, rather than arbitrary structures. A graph is modelled as a relational structure with one symmetric binary relation signifying adjacency. By a *class* of graphs we mean any set of graphs. For a class of graphs $\mathcal{C}$, a *unary expansion* of $\mathcal{C}$ is any class $\widehat{\mathcal{C}}$ of structures such that each $\widehat{G} \in \widehat{\mathcal{C}}$ is obtained from some graph in $G \in \mathcal{C}$ by adding some unary predicates. Thus, the elements of $\widehat{\mathcal{C}}$ can be regarded as vertex-colored graphs from $\mathcal{C}$. A class of graphs $\mathcal{C}$ is called *monadically stable* if one cannot interpret, using a fixed formula $\varphi(\bar{x}, \bar{y})$ of first-order logic, arbitrarily long linear orders in any unary expansion $\widehat{\mathcal{C}}$ of $\mathcal{C}$. More precisely, for every unary expansion $\widehat{\mathcal{C}}$ and formula $\varphi(\bar{x}, \bar{y})$ with $|\bar{x}| = |\bar{y}|$ (over the signature of $\widehat{\mathcal{C}}$) there is a bound ℓ such that there is no structure $\widehat{G} \in \widehat{\mathcal{C}}$ and tuples $\bar{a}_1, \ldots, \bar{a}_\ell \in V(\widehat{G})^{\bar{x}}$ such that $\widehat{G} \models \varphi(\bar{a}_i, \bar{a}_j)$ if and only if $i \leqslant j$. More generally, $\mathcal{C}$ is *monadically dependent* (or *monadically NIP*) if one cannot interpret, using a fixed formula $\varphi(\bar{x}, \bar{y})$ of first-order logic, all finite graphs in any unary expansion of $\mathcal{C}$. Thus, from the model-theoretic perspective, the intuition is that being monadically dependent is being non-trivially constrained: for any fixed interpretation, one cannot interpret arbitrarily complicated structures in vertex-colored graphs from the considered class. On the other hand, graphs from monadically stable classes are "orderless", in the sense that one cannot totally order any large part of them using a fixed first-order formula.

Baldwin and Shelah proved that in the definitions, one can alternatively consider only formulas $\varphi(x, y)$ with just a pair of free variables, instead of a pair of tuples of variables [3, Lemma 8.1.3, Theorem 8.1.8]. Moreover, they proved that monadically stable theories are *tree decomposable* [3, Theorem 4.2.17], providing a structure theorem for such theories, although one of a very infinitary nature. A more explicit, combinatorial structure theorem for monadically stable and monadically dependent is desirable for obtaining algorithmic results for the considered classes, as we discuss later.

On the other hand, Braunfeld and Laskowski [6] very recently proved that for *hereditary* classes of structures $\mathcal{C}$ that are not monadically stable or monadically dependent, the required obstructions (total orders or arbitrary graphs) can be exhibited by a boolean combination of existential formulas $\varphi(\bar{x}, \bar{y})$ in the signature of $\mathcal{C}$, without any additional unary predicates. Among other things, this shows that for hereditary classes of structures, the notions of monadic stability coincides with the more well-known notion of stability, and similarly, monadic dependence coincides with dependence (NIP). Furthermore, since the formulas are existential, this result can be seen as a combinatorial non-structure theorem for hereditary classes that are not monadically stable (resp. monadically dependent). Still, they do not provide explicit structural results for classes that are monadically stable or monadically dependent.

[1] Formally, Baldwin and Shelah [3], as well as Braunfeld and Laskowski [6], study monadically dependent and monadically stable *theories*, rather than classes of structures. Some of their results transfer to the more general setting of monadically dependent/stable classes of structures.

Explicit, combinatorial and algorithmic structural results for monadically dependent and monadically stable classes are not only desired, but also expected to exist, based on the known examples of such classes that have been studied in graph theory and computer science. As observed by Adler and Adler [2] based on the work of Podewski and Ziegler [16], all *nowhere dense* graph classes are monadically stable. A class $\mathcal{C}$ is nowhere dense if for every fixed $r \in \mathbb{N}$, one cannot find r-subdivisions of arbitrarily large cliques as subgraphs of graphs in $\mathcal{C}$. In particular, every class excluding a fixed topological minor (so also the class of planar graphs, or the class of subcubic graphs) is monadically stable. In fact, it follows from the results of Adler and Adler [2] and of Dvořák [10] that monadic stability and monadic dependence are both equivalent to nowhere denseness when considering only sparse classes of graphs (formally, classes of graphs that excludes a fixed biclique as a subgraph). However, monadic stability and monadic dependence are not bound to sparsity; they can be used to understand and quantify structure in dense graphs as well.

The pinnacle of the theory of nowhere dense graph classes is the result of Grohe, Kreutzer, and Siebertz [14] that the model-checking problem for first-order logic is fixed-parameter tractable on any nowhere dense class of graphs.

▶ **Theorem 1** ([14]). *For every nowhere dense graph class $\mathcal{C}$, first-order sentence φ, and $\varepsilon > 0$, there exists an algorithm that given an n-vertex graph $G \in \mathcal{C}$ decides whether $G \models \varphi$ in time $\mathcal{O}_{\mathcal{C},\varphi,\varepsilon}(n^{1+\varepsilon})$.*

Here, and in the following, the notation $\mathcal{O}_p(\cdot)$ hides multiplicative factors that depend only on the the parameter p.

Monadically dependent classes include all monadically stable classes, in particular all nowhere dense classes, but also for instance all classes of bounded twin-width [5]. An analogous result, with $1+\varepsilon$ replaced by 3, holds for all classes $\mathcal{C}$ of *ordered* graphs[2] of bounded twin-width [4].

In light of the discussion above, monadic stability and monadic dependence seem to be well-behaved generalizations of nowhere denseness that are defined in purely model-theoretic terms; hence these concepts may be even better suited for treating the model-checking problem for first-order logic. This motivated the following conjecture [1], which has been a subject of intensive study over the last few years[3].

▶ **Conjecture 2.** *Let $\mathcal{C}$ be a monadically dependent graph class. There exists a constant $c \in \mathbb{N}$ depending only on $\mathcal{C}$ and, for every first-order sentence φ, an algorithm that, given a n-vertex graph $G \in \mathcal{C}$, decides whether $G \models \varphi$ in time $\mathcal{O}_{\mathcal{C},\varphi}(n^c)$.*

Conjecture 2 is not even resolved for monadically stable classes. To approach this conjecture, it is imperative to obtain explicit, combinatorial structure theorems for monadically stable and in monadically dependent graph classes, with a particular focus on finding analogs of the tools used in the proof of Theorem 1. Our work contributes in this direction. We provide certain recursive tree-like decompositions for graphs in monadically stable graph classes, which can be most intuitively explained in terms of games. On the one hand, our decompositions generalize a similar result for nowhere dense classes, recalled below. On the other hand, they are remininiscent of the tree decomposability property proved by Baldwin

[2] *Ordered graphs* are graphs equipped with a total order.

[3] To the best of our knowledge the conjecture was first explicitly discussed during the open problem session of the Algorithms, Logic and Structure Workshop in Warwick, in 2016, see [1].

and Shelah, but are more explicit and finitary in nature. Furthermore, we provide a characterization of monadic stability via forbidden patterns, similar to the known characterization of nowhere denseness.

Splitter game. The cornerstone of the proof of Theorem 1 is a game-theoretic characterization of nowhere denseness, through the *Splitter game*. This game has a fixed radius parameter $r \in \mathbb{N}$ and is played on a graph G between two players, *Splitter* and *Localizer*, who make moves in rounds alternately. In each round, first Splitter chooses any vertex u and removes it from the graph. Next, Localizer selects any other vertex v, and the game gets restricted to the subgraph induced by the ball of radius r with center at v. The game ends with Splitter's victory when there are no vertices left in the graph.

▶ **Theorem 3** ([14]). *A class $\mathcal{C}$ of graphs is nowhere dense if and only if for every $r \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that for every $G \in \mathcal{C}$, Splitter can win the radius-r Splitter game on G within k rounds.*

Very roughly speaking, Theorem 3 shows that any graph from a nowhere dense class can be hierarchically decomposed into smaller and smaller parts so that the decomposition has height bounded by a constant k depending only on the class and the locality parameter r. This decomposition is used in the algorithm of Theorem 1 to guide model-checking.

Flipper game. In this work we introduce an analog of the Splitter game for monadically stable graph classes: the *Flipper game*. Similarly to before, the game is played on a graph G and there is a fixed radius parameter $r \in \mathbb{N}$. There are two players, *Flipper* and *Localizer*, which make moves in rounds alternately. In each round, first Flipper selects any pair of vertex subsets A, B (possibly non-disjoint) and applies the *flip* between A and B: inverts the adjacency between any pair (a, b) of vertices with $a \in A$ and $b \in B$. Then Localizer, just as in the Splitter game, selects a ball of radius r, and the game is restricted to the subgraph induced by this ball. The game is won by Flipper once there is only one vertex left. See Figure 1 for an illustration.

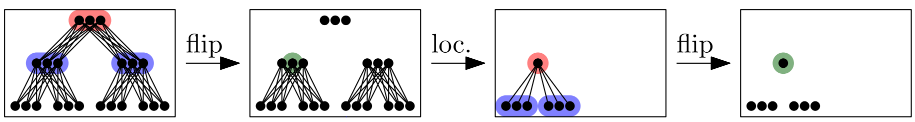

Figure 1 An example play of the radius-1 Flipper game. Taking turns, Flipper flips the red set with the blue set and Localizer restricts to the radius-1 ball centered at the green vertex.

We remark that the Flipper game is a radius-constrained variant of the natural game for graph parameter *SC-depth*, which is functionally equivalent to *shrubdepth*, in the same way that the Splitter game is a radius-constrained variant of the natural game for treedepth. SC-depth and shrubdepth were introduced and studied by Ganian et al. in [13, 12].

Our main result is the following analog of Theorem 3 for monadically stable classes.

▶ **Theorem 4.** *A class $\mathcal{C}$ of graphs is monadically stable if and only if for every $r \in \mathbb{N}$ there exists $k \in \mathbb{N}$ such that for every graph $G \in \mathcal{C}$, Flipper can win the radius-r Flipper game on G within k rounds.*

Let us compare Theorem 4 with another recent characterization of monadic stability, proposed by Gajarský and Kreutzer, and proved by Dreier, Mählmann, Siebertz, and Toruńczyk [9], through the notion of *flip-flatness*. This notion is an analog of *uniform* *quasi-wideness*, introduced by Dawar [7]. Without going into technical details, a class of graphs $\mathcal{C}$ is *uniformly quasi-wide* if for any graph $G \in \mathcal{C}$ and any large enough set of vertices A in G, one can find many vertices in A that are pairwise far from each other after the removal of a constant number of vertices from G. As proved by Nešetřil and Ossona de Mendez [15], a class of graphs is uniformly quasi-wide if and only if it is nowhere dense. The definition of flip-flatness is obtained from uniform quasi-wideness similarly as the Flipper game is obtained from the Splitter game: by replacing the concept of deleting a vertex with applying a flip; see Definition 7 for a formal definition. The fact that monadic stability is equivalent to flip-flatness (as proved in [9]) and to the existence of a short winning strategy in the Flipper game (as proved in this paper) suggests the following: the structural theory of monadically stable graph classes mirrors that of nowhere dense graph classes, where the flip operation is the analog of the operation of removing a vertex.

We give two very different proofs of Theorem 4. The first proof is based on elementary model-theoretic techniques, and it provides new insight into the properties of monadically stable graph classes. As a side effect, it gives a new (though non-algorithmic) proof of the main result of [9]: equivalence of monadic stability and flip-flatness. On the other hand, the second proof relies on the combinatorial techniques developed in [9]. It has the advantage of being effective, and provides an efficient algorithm for computing Flipper's moves in a winning strategy.

Forbidden patterns. A class $\mathcal{C}$ of graphs is nowhere dense if for every fixed $r \in \mathbb{N}$ the exact r-subdivision of some clique K_n is not a subgraph of any $G \in \mathcal{C}$, which can be understood as a forbidden pattern characterization. Our model-theoretic proof of Theorem 4 uncovers a similar characterization of monadically stable classes, providing a strong combinatorial non-structure theorem. We prove that a class $\mathcal{C}$ of graphs is monadically stable if and only if there exists a fixed $\ell \in \mathbb{N}$ such that all graphs from $\mathcal{C}$ exclude a ladder of length ℓ as a semi-induced subgraph (see Section 2 for a formal definition), and $\mathcal{C}$ is *pattern-free*. A class $\mathcal{C}$ of graphs is *not* pattern-free if for some $r \geqslant 1, k \in \mathbb{N}$ the exact r-subdivision of every clique K_n can be obtained from an induced subgraph H of some $G \in \mathcal{C}$ by first partitioning $V(H)$ into k parts, and then either flipping the edges, removing all the edges, or inserting all the edges between some pairs of the partition. Equivalently $\mathcal{C}$ is *not* pattern-free if, using a quantifier-free formula $\varphi(x, y)$, one can encode (more formally *transduce*) the class of all r-subdivided cliques for a fixed $r \geqslant 1$.

Model-theoretic proof. The following statement lists properties equivalent to monadic stability uncovered in our model-theoretic proof of Theorem 4. Each condition is shortly explained below the theorem and formally defined in the full version of the paper [11].

▶ **Theorem 5.** *Let $\mathcal{C}$ be a class of graphs. Then the following conditions are equivalent:*

1. *$\mathcal{C}$ is monadically stable.*
2. *$\mathcal{C}$ has a stable edge relation and is monadically dependent with respect to existential formulas $\varphi(x, y)$ with two free variables.*
3. *$\mathcal{C}$ has a stable edge relation and is pattern-free.*
4. *For every $r \in \mathbb{N}$ every model G of the theory of $\mathcal{C}$, every elementary extension H of G, and every vertex $v \in V(H) - V(G)$, there is a finite set $S \subseteq V(G)$ that r-separates v from G.*
5. *For every $r \in \mathbb{N}$ there is $k \in \mathbb{N}$ such that Flipper wins the Flipper Game with qf-definable separation of radius r on every $G \in \mathcal{C}$ in at most k rounds.*

6. *For every $r \in \mathbb{N}$ there is $k \in \mathbb{N}$ such that Flipper wins the Flipper game of radius r on every $G \in \mathcal{C}$ in at most k rounds.*
7. *$\mathcal{C}$ is flip-flat.*

Note that Theorem 4 is the equivalence (1)↔(6). Let us give a brief overview of the presented conditions.

Conditions (1) and (2), respectively, are monadic stability and a weak form of *existential monadic stability*. Recall that Baldwin and Shelah proved that it is sufficient to consider formulas $\varphi(x, y)$ with two free variables in the definition of monadic stability (instead of formulas $\varphi(\bar{x}, \bar{y})$). Braunfeld and Laskowski proved that it is sufficient to consider boolean combinations of existential formulas $\varphi(\bar{x}, \bar{y})$ that do not involve additional unary predicates. The condition (2) lies somewhere in between: it implies that it is sufficient to consider existential formulas $\varphi(x, y)$ with two variables, possibly involving additional unary predicates. In particular, it implies the result of Baldwin and Shelah (in the case of graph classes) and is incomparable with the result of Braunfeld and Laskowski. Our proof uses different techniques.

Condition (3) concerns the combinatorial notion of pattern-freeness discussed earlier.

Condition (4) is phrased in the language of model theory and serves a key role in our proof. It resembles a fundamental property called "definability of types", and in essence it says the following: whenever working with a model G of the theory of $\mathcal{C}$, every element of any elementary extension of G can be robustly "controlled" by a finite subset of G. We believe that the new notion of r-separation used here is of independent interest. It refers to non-existence of short paths after applying some flips governed by S.

Conditions (5) and (6) assert the existence of a short winning strategy in two variants of the Flipper game.

Finally, condition (7) is the notion of flip-flatness, whose equivalence with monadic stability was proved by Dreier et al. [9].

Algorithmic proof. We also give a purely combinatorial proof of the forward implication of Theorem 4, which in particular provides a way to efficiently compute Flipper's moves in a winning strategy. Formally, we show the following.

▶ **Theorem 6.** *Let $\mathcal{C}$ be a monadically stable class of graphs. Then for every radius $r \in \mathbb{N}$ there exist $k \in \mathbb{N}$ and a Flipper strategy* $\mathrm{flip}^\star$ *such that the following holds:*

- *When playing according to* $\mathrm{flip}^\star$ *in the Flipper game of radius r on any graph $G \in \mathcal{C}$, Flipper wins within at most k rounds.*
- *Each move of* $\mathrm{flip}^\star$ *on an n-vertex graph $G \in \mathcal{C}$ can be computed in time $\mathcal{O}_{\mathcal{C},r}(n^2)$.*

The main idea behind the proof of Theorem 6 is to rely on the result of Dreier et al. that monadically stable graph classes are flip-flat [9]. Using the combinatorial tools developed in [9], we strengthen this property: we prove that the set of flips F whose application uncovers a large *scattered set* Y (a set of vertices that are pairwise far from each other) can be selected in a somewhat canonical way, so that knowing any 5-tuple of vertices in Y is enough to uniquely determine F. We can then use such strengthened flip-flatness to provide a winning strategy for Flipper; this roughly resembles the Splitter's strategy used by Grohe et al. in their proof of Theorem 3, which in turn relies on uniform quasi-wideness.

Theorem 6, the algorithmic version of Theorem 4, is the key to algorithmic applications of the Flipper game. In particular, it was very recently used by Dreier, Mählmann, and Siebertz [8] to approach the first-order model checking problem on monadically stable graph classes and prove that it is fixed-parameter tractable on structurally nowhere dense classes, an important subclass of monadically stable classes.

Organization. After introducing monadic stability and the Flipper game in the next section, we give an outline of the model theoretic proof (Section 3) and the algorithmic proof (Section 4). We refer to the appended full version for details.

2 Preliminaries

All graphs in this paper are simple and loopless but not necessarily finite. For a vertex v of a graph G, we write $N(v)$ for the (open) neighborhood of v in G; so $N(v) \coloneqq \{u \in V(G) \mid uv \in E(G)\}$. For two sets $X, Y \subseteq V(G)$ the bipartite graph *semi-induced* by X and Y in G, denoted $G[X,Y]$, is the bipartite graph with parts X and Y, and edges uv for $u \in X$, $v \in Y$ with $uv \in E(G)$. For vertices $a, b \in V(G)$, an (a,b)*-path* is a path with ends a and b. Similarly, for sets $A, B \subseteq V(G)$, an (A,B)*-path* is a path where one end is in A and the other end is in B.

Model theory. We work with first-order logic over a fixed signature Σ that consists of (possibly infinitely many) constant symbols and of relation symbols. A *model* is a Σ-structure, and is typically denoted $\mathbf{M}, \mathbf{N}$, etc. We usually do not distinguish between a model and its domain, when writing, for instance, $m \in \mathbf{M}$ or $f\colon \mathbf{M} \to X$, or $X \subseteq \mathbf{M}$. A graph G is viewed as a model over the signature consisting of one binary relation denoted E, indicating adjacency between vertices.

A *theory* T (over Σ) is a set of Σ-sentences. A *model of* a theory T is a model $\mathbf{M}$ such that $\mathbf{M} \models \varphi$ for all $\varphi \in T$. When a theory has a model, it is said to be *consistent*. The *theory of* a class of Σ-structures $\mathcal{C}$ is the set of all Σ-sentences φ such that $\mathbf{M} \models \varphi$ for all $\mathbf{M} \in \mathcal{C}$. The *elementary closure* $\overline{\mathcal{C}}$ of $\mathcal{C}$ is the set of all models $\mathbf{M}$ of the theory of $\mathcal{C}$. Thus $\mathcal{C} \subseteq \overline{\mathcal{C}}$, and $\mathcal{C}$ and $\overline{\mathcal{C}}$ have equal theories.

Let $\mathbf{M}$ and $\mathbf{N}$ be two structures with $\mathbf{M} \subseteq \mathbf{N}$, that is, the domain of $\mathbf{M}$ is contained in the domain of $\mathbf{N}$. Then $\mathbf{N}$ is an *elementary extension* of $\mathbf{M}$, written $\mathbf{M} \prec \mathbf{N}$, if for every formula $\varphi(\bar{x})$ (without parameters) and tuple $\bar{m} \in \mathbf{M}^{\bar{x}}$ we have $\mathbf{M} \models \varphi(\bar{m})$ if and only if $\mathbf{N} \models \varphi(\bar{m})$. We also say that $\mathbf{M}$ is an *elementary substructure* of $\mathbf{N}$.

Stability and dependence. A formula $\varphi(\bar{x}; \bar{y})$ is *stable* in a class $\mathcal{C}$ of structures if there exists $k \in \mathbb{N}$ such that for every $\mathbf{M} \in \mathcal{C}$, there are no sequences $\bar{a}_1, \ldots \bar{a}_k \in \mathbf{M}^{\bar{x}}$ and $\bar{b}_1, \ldots, \bar{b}_k \in \mathbf{M}^{\bar{y}}$ such that $\mathbf{M} \models \varphi(\bar{a}_i; \bar{b}_j)$ if and only if $i < j$ for $1 \leqslant i, j \leqslant k$. We say that a class $\mathcal{C}$ of graphs has a *stable edge relation* if the formula $E(x; y)$ is stable in $\mathcal{C}$. Equivalently, $\mathcal{C}$ excludes some ladder as a semi-induced subgraph, where a *ladder* (often called also *half-graph*) of order k is the graph with vertices $a_1, \ldots, a_k, b_1, \ldots, b_k$ and edges $a_i b_j$ for all $1 \leqslant i < j \leqslant k$. Note that replacing $<$ by $\leqslant$ in the above definitions does not change them.

A formula $\varphi(\bar{x}; \bar{y})$ is *dependent*, or *NIP* (standing for "not the independence property") in a class $\mathcal{C}$ if there exists $k \in \mathbb{N}$ such that for every $\mathbf{M} \in \mathcal{C}$, there are no tuples $\bar{a}_1, \ldots, \bar{a}_k \in \mathbf{M}^{\bar{x}}$ and $\bar{b}_J \in \mathbf{M}^{\bar{y}}$ for $J \subseteq \{1, \ldots, k\}$ such that $\mathbf{M} \models \varphi(\bar{a}_i; \bar{b}_J)$ if and only if $i \in J$ for $1 \leqslant i \leqslant k$ and $J \subseteq \{1, \ldots, k\}$. Observe that a formula which is stable is also dependent. A class $\mathcal{C}$ is *stable* (resp. *dependent*) if every formula $\varphi(\bar{x}; \bar{y})$ is stable (resp. dependent) in $\mathcal{C}$.

Let Σ be a signature and let $\widehat{\Sigma}$ be a signature extending Σ by (possibly infinitely many) unary relation symbols and constant symbols. A $\widehat{\Sigma}$-structure $\widehat{\mathbf{M}}$ is a *lift* of a Σ-structure $\mathbf{M}$ if $\mathbf{M}$ is obtained from $\widehat{\mathbf{M}}$ by forgetting the symbols from $\widehat{\Sigma} - \Sigma$. A class of $\widehat{\Sigma}$-structures $\widehat{\mathcal{C}}$ is a *unary expansion* of a class of Σ-structures $\mathcal{C}$ if every structure $\widehat{\mathbf{M}} \in \widehat{\mathcal{C}}$ is a lift of some structure $\mathbf{M} \in \mathcal{C}$. A class $\mathcal{C}$ of structures is *monadically stable* if every unary expansion $\widehat{\mathcal{C}}$ of $\mathcal{C}$ is stable. Similarly, $\mathcal{C}$ is *monadically dependent* (or *monadically NIP*) if every unary expansion $\widehat{\mathcal{C}}$ of $\mathcal{C}$ is dependent. A single structure $\mathbf{M}$ is monadically stable (resp. monadically dependent) if the class $\{\mathbf{M}\}$ is. Note that a class which is monadically stable (resp. monadically dependent) is stable (resp. dependent).

Flips. An *atomic flip* is an operation F specified by a pair (A, B) of (possibly intersecting) vertex sets, which complements the adjacency relation between the sets A and B in a given graph G. Formally, for a graph G, the graph obtained from G by applying the atomic flip F is the graph denoted $G \oplus \mathsf{F}$ with vertex set $V(G)$, where, for distinct vertices u, v in $V(G)$,

$$uv \in E(G \oplus \mathsf{F}) \qquad \iff \qquad \begin{cases} uv \notin E(G), & \text{if } (u,v) \in (A \times B) \cup (B \times A); \\ uv \in E(G), & \text{otherwise.} \end{cases}$$

A set of flips $\{\mathsf{F}_1, \ldots, \mathsf{F}_k\}$ defines an operation F that, given a graph G, results in the graph $G \oplus F := G \oplus \mathsf{F}_1 \oplus \cdots \oplus \mathsf{F}_k$. One can easily show that the order in which we carry out the atomic flips does not matter and that it would be useless to consider multisets. Abusing terminology, we will often just say that the operation F is a set of flips, and write $F = \{\mathsf{F}_1, \ldots, \mathsf{F}_k\}$.

Let $\mathcal{F}$ be a family of vertex sets. Then an $\mathcal{F}$-flip is a set of flips of the form $\{\mathsf{F}_1, \ldots, \mathsf{F}_k\}$, where each flip F_i is a pair (A, B) with $A, B \in \mathcal{F}$. Note that there are at most $2^{|\mathcal{F}|^2}$ different $\mathcal{F}$-flips. In our context, the family $\mathcal{F}$ will usually be a partition of the vertex set of some graph G. An *$\mathcal{F}$-flip* of a graph G, where $\mathcal{F}$ is a family of subsets of $V(G)$, is a graph G' obtained from G after applying an $\mathcal{F}$-flip. Whenever we speak about an $\mathcal{F}$-flip, it will be always clear from the context whether we mean a graph or the family of flips used to obtain it.

Flipper game. Fix a radius r. The Flipper game (or the Flipper/Localizer game) of radius r is played by two players, Flipper and Localizer, on a graph G as follows. At the beginning, set $G_0 := G$. In the ith round, for $i > 0$, the game proceeds as follows.

- If $|G_{i-1}| = 1$ then Flipper wins.
- Localizer chooses a vertex v in G_{i-1} and we set G^{loc}_{i-1} to be the subgraph of G_{i-1} induced by the ball $B^r(v)$ of radius r around v in G_{i-1}.
- Flipper chooses an atomic flip F and applies it to produce G_i, i.e. $G_i = G^{\mathrm{loc}}_{i-1} \oplus \mathsf{F}$.

Variants. It will be convenient to work with different variants of the Flipper game.

- *Batched flipping:* One can consider a variant of the Flipper game where Flipper in the ith move applies a set F of flips to G^{loc}_{i-1} to obtain G_i, where $|F| \leqslant g(i)$ for some function $g \colon \mathbb{N} \to \mathbb{N}$. This does not change the game significantly – if Flipper wins this extended game in m rounds, then Flipper wins the standard Flipper game in $\sum_{i=1}^{m} g(i)$ rounds.
- *Localization modes:* In the definition above, the graph shrinks at each step, and when localizing, distances are computed by only taking into account vertices of the current (shrinked) graph. For this reason, we sometimes call it the *shrinking* variant, or we say that the localization is shrinking. In the *confining* variant, Localizer still remains confined within short distance to his past moves, however Flipper always produces flips of the original graph and distances are measured with respect to the full vertex set.
- *Definability of flips:* We also consider the variant where Flipper is restricted to choosing flips defined using quantifier-free formulas with parameters from the original graph, which we call qf-definable flips.
- *Separation:* Finally, we will also use a variant where distances are measured according to the later defined *separation metric* capturing all possible flips that can be performed over a given (definable) partition of the vertex set.

It will turn out that all these variants are equivalent; for more discussion and formal proofs, see the full version [11]. In particular we will work with the confining Flipper game with qf-definable separation, whose formal definition is deferred to Section 3.

Flip-flatness. The following notion of flip-flatness was introduced in [9] and characterizes monadic stability for graph classes. Given a graph G, a set of vertices $A \subseteq V(G)$ is called *distance-r independent* if all vertices in A are pairwise at distance greater than r in G.

▶ **Definition 7** (Flip-flatness). *A class of graphs $\mathcal{C}$ is* flip-flat *if for every $r \in \mathbb{N}$ there exists a function $N_r \colon \mathbb{N} \to \mathbb{N}$ and a constant $s_r \in \mathbb{N}$ such that for all $m \in \mathbb{N}$, $G \in \mathcal{C}$, and $A \subseteq V(G)$ with $|A| \geqslant N_r(m)$, there exists a set F of flips with $|F| \leqslant s_r$ and $B \subseteq A$ with $|B| \geqslant m$ such that B is distance-r independent in $G \oplus F$.*

▶ **Theorem 8** ([9]). *A class of graphs is monadically stable if and only if it is flip-flat.*

3 Outline of the model-theoretic proof

We prove the implications between the conditions of Theorem 5 as depicted in Figure 2.

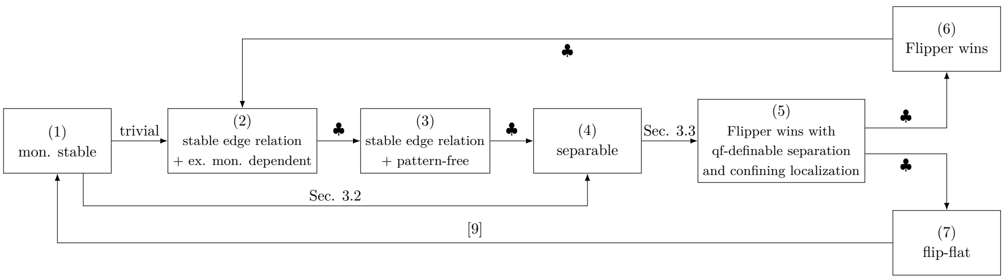

Figure 2 The implications that constitute Theorem 5. Implications marked with ♣ are proved in the full version of the paper.

The implication (1)→(2) is trivial. We prove (2)→(3) by contraposition: using the forbidden patterns, we derive the independence property for some existential formula. The implication (3)→(4) is the core part of our proof; due to space restrictions we will provide a sketch of the implication (1)→(4) (which requires fewer definitions) in Section 3.2. The implication (4)→(5) is proved by proposing a strategy with qf-definable separation in the confining game for Flipper and using compactness combined with (4) to argue that it leads to a victory within a bounded number of rounds. The proof is sketched in Section 3.3. We prove the implication (5)→(7) by (essentially) providing a strategy for Localizer in the confining game with qf-definable separation when the class is not flip-flat. Then we rely on the implication (7)→(1) from [9] to close the circle of implications; this proves the equivalence of (1)-(7) with the exception of (6). We remark that (7)→(1) is the easy implication of [9], hence our reasoning can also serve as an alternative proof of the flip-flatness characterization given in [9].

To put the Flipper game into the picture, we separately prove the implications (5)→(6) and (6)→(2). The implication (5)→(6) relies on a conceptually easy, but technically not-so-trivial translation of the strategies. In the implication (6)→(2) we use obstructions to existential monadic stability to give a strategy for Localizer in the Flipper game that enables her to endure for arbitrarily long.

3.1 Separation

The crucial new ingredient in our model-theoretic proof is the notion of *r-separation.* The definitions provided here are streamlined compared to the full version due to space restrictions.

Let G be a graph, and let $S \subseteq V(G)$ be a finite set of vertices. Consider the equivalence relation $\sim_S$ on $V(G)$, in which two vertices a, b are equivalent if either $a, b \in S$ and $a = b$, or $a, b \notin S$ and $N(a) \cap S = N(b) \cap S$. An *$S$-class* is an equivalence class of $\sim_S$. In other words, it is a set of vertices either of the form $\{s\}$ for some $s \in S$, or of the form $\{v \in V(G) - S \mid N(v) \cap S = T\}$ for some $T \subseteq S$. The *S-class* of a vertex $v \in V(G)$ is the unique S-class which contains v. Hence, $V(G)$ is partitioned into S-classes, and the number of S-classes is at most $|S| + 2^{|S|}$. An *S-flip* of a graph G is an $\mathcal{F}$-flip G' of G, where $\mathcal{F}$ is the partition of $V(G)$ into S-classes.

▶ **Definition 9** (r-separation)**.** *Let G be a graph, S a finite subset of vertices of G. We say that vertices a and b of G are* r-separated over S*, denoted by*[4] $a \mathrel{\raise0.2ex\hbox{$\bigcup$}\!\!\!\!\!|}^{r}_{S} b$*, if there exists a S-flip H of G such that* $\mathrm{dist}_H(a, b) > r$.

For any $r \in \mathbb{N}$ and any graph G, finite subset S of $V(G)$, and sets $A, B \in V(G)$, we write $A \downfootline^{r}_{S} B$ if there exists an S-flip H of G such that H has no (A, B)-path of length at most r. Note that $A \downfootline^{r}_{S} B$ is a stronger condition than $a \downfootline^{r}_{S} b$ for all $a \in A$ and $b \in B$, since we require that the same S-flip H is used for all $a \in A$ and $b \in B$. We write $\not\downfootline^{r}_{S}$ to denote the negation of the relation $\downfootline^{r}_{S}$. If $A \not\downfootline^{r}_{S} B$ we say that A and B are *r-connected over S*. If A consists of a single vertex a then we write $a \downfootline^{r}_{S} B$ for $A \downfootline^{r}_{S} B$.

We now formally introduce the confining Flipper game with qf-definable separation; the most important difference is that we evaluate distances in the original graph G: the localizing ball is always defined with respect to the distance induced by $\downfootline^{r}_{S}$ in G, where S is the set of vertices played by Flipper.

Fix a radius $r \in \mathbb{N}$. The game of radius r is played on a graph G as follows. Let $A_0 = V(G)$ and $S_0 = \emptyset$. For $k = 1, 2, \ldots$, the kth round proceeds as follows.

- If $|A_{k-1}| = 1$, then Flipper wins.
- Otherwise, Localizer picks $c_k \in A_{k-1}$ and we set

  $$A_k := A_{k-1} - \left\{ w \;\middle|\; w \underset{S_{k-1}}{\overset{r}{\downfootline}} c_k \right\}$$

  (where separation is evaluated in the graph G).
- Then Flipper picks $s_k \in V(G)$ and we set $S_k := S_{k-1} \cup \{s_k\}$, and proceed to the next round.

As previously, we may allow Flipper to add $g(i)$ vertices to S_{i-1} in the ith round, where $g \colon \mathbb{N} \to \mathbb{N}$ is some fixed function. Again, if Flipper can win this new game in m rounds, then Flipper can also win the original game in $\sum_{i=1}^{m} g(i)$ rounds.

In the full paper [11, Lemma 3.2], we prove that a winning strategy for Flipper in the above game can be adapted by to win in the original Flipper game. The idea is that Flipper carries out all possible S-flips and then flips back to (an induced subgraph of) the original graph.

[4] The symbol $\downfootline$ denotes forking independence in stable theories. Its use here is justified by the relationship of r-separation and forking independence in monadically stable theories, which is briefly explained in the full version of the paper.

▶ **Lemma 10.** *There exists a function $f : \mathbb{N} \to \mathbb{N}$ such that for every radius r and every graph G the following holds. If Flipper wins the confining game with qf-definable separation of radius $2r$ on G in at most k rounds, then Flipper wins the shrinking game with arbitrary flips of radius r on G in at most $f(k)$ rounds.*

Note that the converse direction, allowing to translate a winning strategy of Flipper from the shrinking to the confining variant, is not immediately clear. However, the equivalence of the two games ultimately follows from Theorem 5.

3.2 Finite separators in monadically stable models

In this section, we provide our key model-theoretic characterization of monadically stable graphs. We will use $\mathbf{M}$ to denote a graph that is typically infinite.

▶ **Definition 11.** *A graph $\mathbf{M}$ is* r-separable *if for every elementary extension $\mathbf{N}$ of $\mathbf{M}$, and every $v \in \mathbf{N} - \mathbf{M}$, there is a finite set $S \subseteq \mathbf{M}$ such that $v \mathop{\downarrow}^{r}_{S} \mathbf{M}$ in $\mathbf{N}$.*

The main result of this section is the following theorem.

▶ **Theorem 12.** *Every monadically stable graph $\mathbf{M}$ is r-separable, for every $r \in \mathbb{N}$.*

The proof will rely on Lemma 13 and Lemma 14 below. To state them, we will need one more definition. Let $\mathbf{M}$ be a graph and $A, B \subseteq \mathbf{M}$. We say that $a, a' \in A$ have the same *E-type* over B if $N(a) \cap B = N(a') \cap B$; this is clearly an equivalence relation. We denote the set of E-types of A over B by $\mathrm{Types}^E(A/B)$.

▶ **Lemma 13.** *Fix $r \in \mathbb{N}$. Let $\mathbf{M}$ be a monadically stable graph, let $\mathbf{N}$ be an elementary extension of $\mathbf{M}$, and let $v \in \mathbf{N}$ be such that the r-ball $B^r(v)$ around v in $\mathbf{N}$ is disjoint from $\mathbf{M}$. Then $\mathrm{Types}^E(B^r(v)/\mathbf{M})$ is finite.*

We now briefly outline the idea behind the proof of Lemma 13. It is a folklore result that in an infinite bipartite graph with sides L and R there is an infinite induced matching, or an infinite induced co-matching, or an infinite induced ladder, or $\mathrm{Types}^E(L/R)$ is finite. Assume towards a contradiction that the last option (with $L = B^r(v)$ and $R = \mathbf{M}$) does not hold. Since we work with a monadically stable graph, we cannot have an infinite ladder. Therefore, there is an infinite induced matching or co-matching between $B^r(v)$ and $\mathbf{M}$. By symmetry, we can assume the former. From this we obtain, for any k, vertices $a_1, \ldots, a_k$ in $\mathbf{M}$ and $b_1, \ldots, b_k$ in $B^r(v) \subseteq \mathbf{N} - \mathbf{M}$ such that the corresponding pairs a_i, b_i form a semi-induced matching and any b_i, b_j are connected by a path of length at most $2r$ that passes through v. Let H denote the subgraph of N induced by $b_1, \ldots, b_k$ together with the paths connecting them (we pick one such path for each pair). Using the fact that $\mathbf{M}$ is an elementary substructure of $\mathbf{N}$, we can then show that there exist as many disjoint copies of H in $\mathbf{M}$ as we want, and all these copies behave in the same way towards $a_1, \ldots a_k$ as the original H. Consequently, we can for each pair a_i, a_j use one copy of H to create a short path between a_i and a_j, and all these paths can be defined using a single first-order formula, which in turn defines a subdivided clique with k principal vertices. Since k is arbitrary, this means that $\mathbf{M}$ is not monadically dependent, as desired.

▶ **Lemma 14.** *For any graphs $\mathbf{M}$ and $\mathbf{N}$ with $\mathbf{M} \prec \mathbf{N}$ and such that $\mathbf{N}$ is monadically stable and for any set $U \subseteq \mathbf{N} - \mathbf{M}$ such that $\mathrm{Types}^E(U/\mathbf{M})$ is finite, there exists a finite set $S \subseteq \mathbf{M}$ and an S-flip which:*
- *1-separates U from $\mathbf{M}$; and*
- *does not flip the S-class $T := \{v \in \mathbf{N} : \forall s \in S. \neg E(v, s)\}$ with any other S-class (including itself), as long as $T \cap U$ is nonempty.*

The idea behind the proof of Lemma 14 can be briefly summarized as follows. Roughly speaking, we aim to find a finite $S \subseteq \mathbf{M}$ such that if $a, a' \in U$ are in the same S-class, then $N(a) \cap \mathbf{M} = N(a') \cap \mathbf{M}$, and analogously, if $b, b' \in \mathbf{M}$ are in the same S-class, then $N(a) \cap U = N(a') \cap U$. Then we can use these properties to suitably flip between S-classes to obtain the result. First we note that since $\mathrm{Types}^E(U/\mathbf{M})$ is finite, we have that $\mathrm{Types}^E(\mathbf{M}/U)$ is also finite, and so by taking one representative from each class of $\mathrm{Types}^E(\mathbf{M}/U)$ we find a finite subset $S_{\mathbf{M}} \subseteq \mathbf{M}$ such that any two elements in the same $S_{\mathbf{M}}$-class have the same neighborhood in $\mathbf{M}$. Clearly, by the same idea we could find a finite subset set S_U of U such that vertices in the same S_U-class have the same neighborhood in U. However, we need our set S_U to be contained in $\mathbf{M}$. To achieve this, we rely on a fundamental fact about stable formulas known as *definability of types*, which allows us to show that sets $N(u) \cap \mathbf{M}$ (where $u \in U$) can be defined from within $\mathbf{M}$ by looking at S_u-classes, where S_u is a finite subset of $\mathbf{M}$. Since there are only finitely many types of vertices in U with respect to the adjacency towards $\mathbf{M}$, we can list them as $u_1, \dots, u_k$ and set $S_U := \cup_i S_u$. We can then take $S = S_{\mathbf{M}} \cup S_U$. We remark that the set S defined in the actual proof of Lemma 14 contains more vertices; we refer to the full version of the paper for details.

An inductive proof of Theorem 12 now follows by putting together Lemmas 13 and 14.

Proof sketch of Theorem 12. We proceed by induction on r. Let $\mathbf{M}$ be a monadically stable graph and let $\mathbf{N}$ be an elementary extension of $\mathbf{M}$. For every $v \in \mathbf{N} - \mathbf{M}$, we have to find a finite set $S \subseteq \mathbf{M}$ such that $v \mathop{\smile\!\!\!\!\!\!\mid}^{r}_{S} \mathbf{M}$ in $\mathbf{N}$. The base case $r = 0$ is immediate as we may take S to be $\varnothing$ since $v \notin \mathbf{M}$. In the inductive step, assume that the result is proved for the distance $r \in \mathbb{N}$. Stated differently, assume there is a finite $S \subseteq \mathbf{M}$ and an S-flip $\mathbf{N}'$ of $\mathbf{N}$ in which the r-ball around v is disjoint from $\mathbf{M}$. It is easily checked that an S-flip of a monadically stable graph is monadically stable, and so $\mathbf{N}'$ is monadically stable. Moreover, one can also show that $\mathbf{N}'$ is an elementary extension of the subgraph of $\mathbf{N}'$ induced by the domain of $\mathbf{M}$. We can therefore apply Lemma 13 to $\mathbf{N}', \mathbf{M}$ and $B^r_{\mathbf{N}'}(v)$. By Lemma 13, $\mathrm{Types}^E(B^r(v)/\mathbf{M})$ is finite. Now Lemma 14 applied to $B^r(v)$ finishes the inductive step and the proof (we are using the fact that we obtain a set S and an S-flip, which doesn't flip the S-class that contains $B^{r-1}(v)$). ◀

Since monadic stability is preserved in the elementary closure[5], we get the following corollary, proving the implication (1)→(4) in Theorem 5.

▶ **Corollary 15.** *If $\mathcal{C}$ is a monadically stable class of graphs and $r \in \mathbb{N}$, then every $\mathbf{M} \in \overline{\mathcal{C}}$ is r-separable.*

3.3 From separability to winning the confining Flipper game

For brevity, in this section we use "Flipper game" to refer to the confining Flipper game with qf-definable separation.

▶ **Theorem 16.** *Fix $r \in \mathbb{N}$, and let $\mathcal{C}$ be a class of graphs such that every $G \in \overline{\mathcal{C}}$ is r-separable. Then there exists $k \in \mathbb{N}$ such that Flipper wins the Flipper game with radius r in k rounds on every $G \in \mathcal{C}$.*

In the proof we will use the Tarski-Vaught test, which we now recall.

[5] The preservation of monadic stability in the elementary closure is true but not obvious (follows from [6]). However, in the full version of the paper, we prove the implication (3)→(4) of Theorem 5 and only require the preservation of edge-stability and pattern-freeness, which we prove easily.

▶ **Theorem 17** (Tarski-Vaught Test). *The following conditions are equivalent for any structures* $\mathbf{M}$ *and* $\mathbf{N}$ *with* $\mathbf{M} \subseteq \mathbf{N}$*.*

- *The structure* $\mathbf{N}$ *is an elementary extension of* $\mathbf{M}$*.*
- *For every formula* $\varphi(y; \bar{x})$ *and tuple* $\bar{m} \in \mathbf{M}^{\bar{x}}$*, if* $\mathbf{N} \models \varphi(n; \bar{m})$ *holds for some* $n \in \mathbf{N}$*, then* $\mathbf{N} \models \varphi(n'; \bar{m})$ *holds for some* $n' \in \mathbf{M}$*.*

In the rest of this section we sketch the proof of Theorem 16. Fix an enumeration $\varphi_1, \varphi_2, \ldots$ of all formulas (in the signature of graphs) of the form $\varphi(y, x_1, \ldots, x_\ell)$, with $\ell \geqslant 0$. We define a strategy of Flipper in any graph G. In the kth round, after Localizer picks $c_k \in A_{k-1}$, Flipper first sets $S := S_{k-1} \cup \{c_k\}$ and marks c_k. Then, for every $i = 1, \ldots, k$, for the formula $\varphi_i(y, \bar{x})$, Flipper does the following.

> For each $\bar{a} \in S^{\bar{x}}$ such that $G \models \exists y.\varphi_i(y, \bar{a})$, Flipper marks any vertex $b \in V(G)$ such that $G \models \varphi_i(b, \bar{a})$.

We say that any strategy of Flipper with this property is *Localizer-complete.* The marked vertices form Flipper response in the kth round, and we set S_k to be the union of S_{k-1} and all the marked vertices. Note that there is a function $f \colon \mathbb{N} \to \mathbb{N}$ such that $|S_k| \leqslant f(k)$ for all $k \in \mathbb{N}$, regardless of which vertices Localizer picks or which of the formulas $\exists y.\varphi_i(y, \bar{a})$ hold.

We prove that there is a number $k \in \mathbb{N}$ such that when Flipper plays according to any Localizer-complete strategy on a graph $G \in \mathcal{C}$, then he wins in at most k rounds. Assume that the conclusion of the theorem does not hold. Then, there exists a sequence of graphs $G_1, G_2, \ldots \in \mathcal{C}$, where in G_n Localizer has a strategy ensuring that Flipper does not win for at least n rounds. We shall now prove that there is some graph G in the elementary closure of $\mathcal{C}$ and a vertex in the graph that survives in the arena indefinitely, when Flipper plays according to a Localizer-complete strategy. We will then use the r-separability of G to derive a contradiction.

▷ Claim 18. There exists a graph $G \in \overline{\mathcal{C}}$, a strategy of Localizer, and a Localizer-complete strategy of Flipper for which the Flipper Game on G lasts indefinitely and the intersection of the arenas $\bigcap_{n<\omega} A_n$ is nonempty.

Proof sketch. For every graph $G_n \in \mathcal{C}$, choose any Localizer-complete strategy of Flipper, and any strategy of Localizer ensuring the game continues for more than n rounds.

In each G_i, use constants to mark moves of Localizer and Flipper in a play in which they play for i moves according to the chosen strategies, and moreover mark by c_ω an arbitrary vertex that remains the arena after i rounds. We then consider, for every $i \in \mathbb{N}$, a sentence ψ_i that is true in a graph if and only if the play encoded by the introduced constants is a valid i move play in the Flipper game and c_ω is in the arena after the ith move. We then have $G_i \models \psi_i$ for each i. By a compactness argument, we can argue that there exists a graph $G \in \overline{\mathcal{C}}$ such that $G \models \psi_i$ for every i. Then we have in G that $c_\omega \in A_i$ for each i, and so $c_\omega \in \bigcap_{n<\omega} A_n$, which means that $\bigcap_{n<\omega} A_n$ is nonempty, as desired. ◁

Let $G \in \overline{\mathcal{C}}$ be the graph produced by Claim 18, along with the strategies of Localizer and Flipper. By assumption, G is r-separable. Recall that $A_0 \supseteq A_1 \supseteq \ldots$ is the sequence of arenas in the play, $c_1, c_2, \ldots$ is the sequence of moves of Localizer, and $S_0 \subseteq S_1 \subseteq \ldots$ is the sequence of sets of vertices marked by Flipper. Denote $A_\omega := \bigcap_{n<\omega} A_n$, and $S_\omega := \bigcup_{n<\omega} S_n$. We will get a contradiction with the previous claim by proving the following claim:

▷ Claim 19. A_ω is empty.

Proof. Observe that for each $k \in \mathbb{N}$, we have $c_k \notin S_{k-1}$: as soon as Localizer plays c_k in S_{k-1}, the arena A_k shrinks to a single vertex and Flipper wins in the following round. Then, A_k is disjoint from S_{k-1}: since Localizer plays c_k outside of S_{k-1}, each vertex of S_{k-1} becomes separated from c_k and thus is removed from the arena. It follows that $A_\omega \cap S_\omega = \emptyset$.

Since Flipper follows a Localizer-complete strategy, S_ω induces an elementary substructure of G by the Tarski-Vaught test (Theorem 17). We also have that $c_1, c_2, \ldots \in S_\omega$ by construction. Now suppose for a contradiction that there exists some $c_\omega \in A_\omega$. We remark that $c_\omega \notin S_\omega$.

By Theorem 12, there exists a finite set $S \subseteq S_\omega$ such that $c_\omega \mathop{\downarrow}^r_S S_\omega$. As S is finite, there is some $n < \omega$ such that $S \subseteq S_n$, so in particular, $c_\omega \mathop{\downarrow}^r_{S_n} S_\omega$. On the other hand, $c_\omega \not\mathop{\downarrow}^r_{S_n} c_{n+1}$, as $c_\omega \in A_{n+1}$. This is a contradiction since $c_{n+1} \in S_\omega$. ◁

However, this means that there exists a graph $G \in \overline{\mathcal{C}}$ and strategies of Localizer and Flipper, for which A_ω is simultaneously nonempty (Claim 18) and empty (Claim 19). This contradicts the existence of the graphs $G_1, G_2, \ldots \in \mathcal{C}$ and completes the proof of Theorem 16.

4 Outline of the algorithmic proof

In this part we outline the proof of Theorem 6 by sketching a winning Flipper strategy whose moves can be computed in time $\mathcal{O}_{\mathcal{C},r}(n^2)$.

Let us first sketch a natural approach to use the flip-flatness characterization of monadic stability (see Definition 7) to derive a winning strategy for Flipper. Consider the radius-r Flipper game on a graph G from a monadically stable class $\mathcal{C}$. For convenience we may assume for now that we work with an extended version of the game where at each round Flipper can apply a bounded (in term of the round's index) number of flips, instead of just one (see the discussion in the preliminaries). As making a vertex isolated requires one flip – between the vertex in question and its neighborhood – we can always assume that the flips applied by Flipper in round i make all the i vertices previously played by Localizer isolated. Hence, Localizer needs to play a new vertex in each round, thus building a growing set X of her moves.

Fix some constant $m \in \mathbb{N}$. According to flip-flatness, there exists some number $N := N_{2r}(m)$ with the property that once X has grown to the size N, we find a set of flips F – whose size is bounded independently of m – and a set Y of m vertices in X that are pairwise at distance greater than $2r$ in $G \oplus F$. It now looks reasonable that Flipper applies the flips from F within his next move. Indeed, since after applying F the vertices of Y are at distance more than $2r$ from each other, the intuition is that F robustly "disconnects" the graph so that the subsequent move of the Localizer will necessarily localize the game to a simpler setting. This intuition is, however, difficult to capture: flip-flatness a priori does not provide any guarantees on the disconnectedness of $G \oplus F$ other than that the vertices of Y are far from each other.

The main idea for circumventing this issue is to revisit the notion of flip-flatness and strengthen it with an additional *predictability* property. Intuitively, predictability says that being given any set of 5 vertices in Y as above is sufficient to uniquely reconstruct the set of flips F. Formally, we prove the following strengthening of the results of [9]. Here and later on, $O(G)$ denotes the set of linear orders on the vertices of G.

▶ **Theorem 20** (Predictable flip-flatness). *Fix a radius $r \in \mathbb{N}$ and a monadically stable class of graphs $\mathcal{C}$. Then there exist the following:*

- *An unbounded non-decreasing function $\alpha_r \colon\ N \to \mathbb{N}$ and a bound $\lambda_r \in \mathbb{N}$.*

- *A function* FF_r *that maps each triple* $(G \in \mathcal{C}, \preccurlyeq \in O(G), X \subseteq V(G))$ *to a pair* (Y, F) *such that:*
  - F *is a set of at most* λ_r *flips in* G*, and*
  - Y *is a set of* $\alpha_r(|X|)$ *vertices of* X *that is distance-*r *independent in* $G \oplus F$*.*
- *A function* $\mathsf{Predict}_r$ *that maps each triple* $(G \in \mathcal{C}, \preccurlyeq \in O(G), Z \subseteq V(G))$ *with* $|Z| = 5$ *to a set* F *of flips in* G *such that the following holds:*
  - *For every* $X \subseteq V(G)$*, if* $(Y, F) = \mathsf{FF}_r(G, \preccurlyeq, X)$ *and* $Z \subseteq Y$*, then* $F = \mathsf{Predict}_r(G, \preccurlyeq, Z)$*.*

Moreover, given G*,* $\preccurlyeq$*, and* Z*,* $\mathsf{Predict}_r(G, \preccurlyeq, Z)$ *can be computed in time* $\mathcal{O}_{\mathcal{C},r}(|V(G)|^2)$*.*

Let us explain the intuition behind the mappings FF_r and $\mathsf{Predict}_r$ provided by Theorem 20. The existence of bounds α_r and λ_r and of the function FF_r with the properties as above is guaranteed by the standard flip-flatness, see Definition 7 and Theorem 8. However, in the proof we pick the function FF_r in a very specific way, so that the flip set F is defined in a somewhat minimal way with respect to a given vertex ordering $\preccurlyeq$. This enables us to predict what the flip set F should be given any set of 5 vertices from Y. This condition is captured by the function $\mathsf{Predict}_r$.

We remark that the predictability property implies the following condition, which we call *canonicity*, and which may be easier to think about. (We assume the notation from Theorem 20.)

- For every $G \in \mathcal{C}$, $\preccurlyeq \in O(G)$, and $X, X' \subseteq V(G)$, if we denote $(Y, F) = \mathsf{FF}_r(G, \preccurlyeq, X)$ and $(Y', F') = \mathsf{FF}_r(G, \preccurlyeq, X')$, then $|Y \cap Y'| \geqslant 5$ entails $F = F'$.

Indeed, to derive canonicity from predictability note that $F = \mathsf{Predict}_r(G, \preccurlyeq, Z) = F'$, where Z is any 5-element subset of $Y \cap Y'$. Predictability strengthens canonicity by requiring that the mapping from 5-element subsets to flip sets is governed by a single function $\mathsf{Predict}_r$, which is moreover efficiently computable. We prove Theorem 20 in the appended full version of the paper. The proof is based on the combinatorial tools from [9], which were developed to prove the standard flip-flatness. However, the generated sets of flips have to be chosen and analyzed with much greater care.

We now outline how Flipper can use predictable flip-flatness for radius $2r$ to win the radius-r Flipper game in a bounded number of rounds. Suppose the game is played on a graph G; we also fix an arbitrary ordering $\preccurlyeq$ of vertices of G. Flipper will keep track of a growing set X of vertices played by the Localizer. The game proceeds in a number of *eras*, where at the end of each era X will be augmented by one vertex. In an era, Flipper will spend $2 \cdot \binom{|X|}{5}$ rounds trying to robustly disconnect the current set X. To this end, for every 5-element subset Z of X Flipper performs a pair of rounds:

- In the first round, Flipper computes $F := \mathsf{Predict}_{2r}(G, \preccurlyeq, Z)$ and applies the flips from F. Subsequently, Localizer needs to localize the game to a ball of radius r in the F-flip of the current graph.
- In the second round, Flipper reverses the flips by applying F again, and Localizer again localizes.

Thus, after performing a pair of rounds as above, we end with an induced subgraph of the original graph, which moreover is contained in a ball of radius r in the F-flip. Having performed all the $\binom{|X|}{5}$ pairs of rounds as above, Flipper makes the last round of this era: he applies flips that isolates all vertices of X, thus forcing Localizer to play any vertex outside of X that is still available. This adds a new vertex to X and a new era begins.

Let us sketch why this strategy leads to a victory of Flipper within a bounded number of rounds. Suppose the game proceeds for N eras, where N is such that $\alpha_{2r}(N) \geqslant 7$. Then we can apply predictable flip-flatness to the set X built within those eras, thus obtaining a pair $(Y, F) := \mathsf{FF}_{2r}(G, \preccurlyeq, X)$ such that $|Y| = 7$ and F is a set of flips such that Y is distance-$2r$ independent in $G \oplus F$. Enumerate Y as $\{v_1, \ldots, v_7\}$, according to the order in which they were added to X during the game. Let $Z := \{v_1, \ldots, v_5\}$ and note that $F = \mathsf{Predict}_{2r}(G, \preccurlyeq, Z)$. Observe that in the era following the addition of v_5 to X, Flipper considered Z as one of the 5-element subsets of the (current) set X. Consequently, within one of the pairs of rounds in this era, he applied flips from F and forced Localizer to localize the game subsequently. Since v_6 and v_7 are at distance larger than $2r$ in $G \oplus F$, this necessarily resulted in removing v_6 or v_7 from the graph. This is a contradiction with the assumption that both v_6 and v_7 were played later in the game.

References

1 Algorithms, Logic and Structure Workshop in Warwick – Open Problem Session. URL: https://warwick.ac.uk/fac/sci/maths/people/staff/daniel_kral/alglogstr/openproblems.pdf, 2016. [Online; accessed 23-Jan-2023].

2 Hans Adler and Isolde Adler. Interpreting nowhere dense graph classes as a classical notion of model theory. *Eur. J. Comb.*, 36:322–330, 2014. doi:10.1016/j.ejc.2013.06.048.

3 J.T. Baldwin and S. Shelah. Second-order quantifiers and the complexity of theories. *Notre Dame Journal of Formal Logic*, 26(3):229–303, 1985.

4 Édouard Bonnet, Ugo Giocanti, Patrice Ossona de Mendez, Pierre Simon, Stéphan Thomassé, and Szymon Toruńczyk. Twin-width iv: Ordered graphs and matrices. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, pages 924–937, New York, NY, USA, 2022. Association for Computing Machinery. doi:10.1145/3519935.3520037.

5 Édouard Bonnet, Eun Jung Kim, Stéphan Thomassé, and Rémi Watrigant. Twin-width I: tractable FO model checking. In Sandy Irani, editor, *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, pages 601–612. IEEE, 2020. doi:10.1109/FOCS46700.2020.00062.

6 Samuel Braunfeld and Michael C Laskowski. Existential characterizations of monadic NIP. *arXiv preprint*, 2022. arXiv:2209.05120.

7 Anuj Dawar. Homomorphism preservation on quasi-wide classes. *J. Comput. Syst. Sci.*, 76(5):324–332, 2010. doi:10.1016/j.jcss.2009.10.005.

8 Jan Dreier, Nikolas Mählmann, and Sebastian Siebertz. First-order model checking on structurally sparse graph classes. *arXiv preprint*, 2023. arXiv:2302.03527.

9 Jan Dreier, Nikolas Mählmann, Sebastian Siebertz, and Szymon Toruńczyk. Indiscernibles and flatness in monadically stable and monadically NIP classes. *arXiv preprint*, 2022. arXiv:2206.13765.

10 Zdenek Dvořák. Induced subdivisions and bounded expansion. *Eur. J. Comb.*, 69:143–148, 2018. doi:10.1016/j.ejc.2017.10.004.

11 Jakub Gajarský, Nikolas Mählmann, Rose McCarty, Pierre Ohlmann, Michał Pilipczuk, Wojciech Przybyszewski, Sebastian Siebertz, Marek Sokołowski, and Szymon Toruńczyk. Flipper games for monadically stable graph classes, 2023. doi:10.48550/ARXIV.2301.13735.

12 Robert Ganian, Petr Hliněný, Jaroslav Nešetřil, Jan Obdržálek, and Patrice Ossona de Mendez. Shrub-depth: Capturing height of dense graphs. *Log. Methods Comput. Sci.*, 15(1), 2019. doi:10.23638/LMCS-15(1:7)2019.

13 Robert Ganian, Petr Hliněný, Jaroslav Nešetřil, Jan Obdržálek, Patrice Ossona de Mendez, and Reshma Ramadurai. When trees grow low: Shrubs and fast MSO_1. In *37th International Symposium on Mathematical Foundations of Computer Science 2012, MFCS 2012*, volume 7464 of *Lecture Notes in Computer Science*, pages 419–430. Springer, 2012.

14 Martin Grohe, Stephan Kreutzer, and Sebastian Siebertz. Deciding first-order properties of nowhere dense graphs. *J. ACM*, 64(3):17:1–17:32, 2017. doi:10.1145/3051095.

15 Jaroslav Nešetřil and Patrice Ossona de Mendez. On nowhere dense graphs. *Eur. J. Comb.*, 32(4):600–617, 2011. doi:10.1016/j.ejc.2011.01.006.

16 Klaus-Peter Podewski and Martin Ziegler. Stable graphs. *Fundamenta Mathematicae*, 100(2):101–107, 1978.