Ehrenfeucht-Fraïssé Games: Applications and Complexity (Introductory Tutorial)

Angelo Montanari   Nicola Vitacolonna

Department of Mathematics and Computer Science
University of Udine

Udine, September 14, 2009 / GAMES 2009

Outline

Background on finite model theory

Books

- H.-D. Ebbinghaus and J. Flum
  Finite Model Theory
  Springer, 2nd edition, 2005
- L. Libkin
  Elements of Finite Model Theory
  Springer, 2004

Why finite model theory?

- Connections with computation
  - **Computational complexity**
    logical description of complexity classes (e.g., the problem $\mathrm{P} = \mathrm{NP}$ amounts to the question whether two fixed-point logics have the same expressive power in finite structures)
  - **Verification**
    finite structures can be used to describe finite runs of machines
  - **Database theory**
    the relational model identifies a database with a finite relational structure (formulas of a formal language can be viewed as programs to evaluate their meaning in a structure and, vice versa, one can express queries of a certain computational complexity in a given formal language)
- Genuinely finite queries
  - Has the domain even cardinality?

Most theorems fail, one method survives

We focus our attention on first-order (FO) logic

- Results of model theory often do not apply to the finite
  - Gödel's completeness theorem
  - Compactness theorem
  - Löwenheim-Skolem theorem
  - Definability and interpolation results
  - etc. . .
- Ehrenfeucht-Fraïssé games are an exception

Compactness fails in the finite

- γ_n: "there are at least n distinct elements"
  - $\gamma_n \stackrel{\text{def}}{=} \exists x_1 \cdots \exists x_n \bigwedge_{1 \leqslant i < j \leqslant n} (x_i \neq x_j)$
- $\Gamma = \{ \gamma_n \mid n > 0 \}$

- **General case**: every finite subset of Γ is satisfiable and thus (compactness theorem) Γ is satisfiable, that is, it has an (infinite) model

- **Finite structures**: every finite subset of Γ is satisfiable (it has a finite model), but Γ has no finite model

An application

- Connectivity is not FO-definable over the class of all graphs $\mathcal{G} = (G, E)$
  - The proof is via compactness
  - Assume ϕ defines connectivity
  - ψ_n: "there is no path of length $n + 1$ from c_1 to c_2"
  - Let $T = \{\, \psi_n \mid n > 0 \,\} \cup \{c_1 \neq c_2, \neg E(c_1, c_2), \phi\}$
  - Every finite subset of T is satisfiable, but T is not

- Is connectivity definable over all finite graphs?

Isomorphic and elementarily equivalent structures

Definition (Isomorphic structures)

Two relational structures $\mathcal{A}$, $\mathcal{B}$, over the same finite vocabulary τ, are isomorphic ($\mathcal{A} \cong \mathcal{B}$) if there is an isomorphism from $\mathcal{A}$ to $\mathcal{B}$, that is, a bijection $\pi : \mathcal{A} \mapsto \mathcal{B}$ preserving relations and constants.

Theorem

Every finite structure can be characterized in first-order logic (FO) up to isomorphism, that is, for every finite structure $\mathcal{A}$ there exists a FO sentence $\varphi_{\mathcal{A}}$ such that, for every $\mathcal{B}$, we have

$$\mathcal{B} \models \varphi_{\mathcal{A}} \textit{ iff } \mathcal{A} \cong \mathcal{B}.$$

Definition (Elementarily equivalent structures)

Two structures $\mathcal{A}$, $\mathcal{B}$ are elementarily equivalent ($\mathcal{A} \equiv \mathcal{B}$) if they satisfy the same FO sentences.

m-equivalent structures

Quantifier rank $\mathrm{qr}(\phi)$ of a FO-formula ϕ: maximum number of nested quantifiers in ϕ:

- if ϕ is atomic then $\mathrm{qr}(\phi) = 0$;
- $\mathrm{qr}(\neg\phi_1) = \mathrm{qr}(\phi_1)$;
- $\mathrm{qr}(\phi_1 \vee \phi_2) = \max(\mathrm{qr}(\phi_1), \mathrm{qr}(\phi_2))$;
- $\mathrm{qr}(\exists x\, \phi_1) = \mathrm{qr}(\phi_1) + 1$.

Example

$\phi = \forall x\, (P(x) \to \exists y\, Q(x, y) \vee \exists y\, R(y))$ has $\mathrm{qr}(\phi) = 2$.

Definition (m-equivalent structures)

Two structures $\mathcal{A}$ and $\mathcal{B}$ are m-equivalent, denoted $\mathcal{A} \equiv_m \mathcal{B}$, with $m \geqslant 0$, if they satisfy the same FO sentences of quantifier rank up to m.

Combinatorial Games

Ehrenfeucht-Fraïssé games are (logical) combinatorial games.

- Combinatorial games:
  - Two opponents
  - Alternate moves
  - No chance
  - No hidden information
  - No loops
  - The player who cannot move loses[1]

E. R. Berlekamp, J. H. Conway, and R. K. Guy
Winning Ways for your mathematical plays
A K Peters LTD, 2nd edition, 2001

[1] In Combinatorial Game Theory (CGT), this is called *normal play* (the opposite rule: "the player who cannot move wins" is called *misère play*, and it gives rise to quite different a theory)

Ehrenfeucht-Fraïssé games (EF-games)

- (Logical) combinatorial games
- The playground: two relational structures $\mathcal{A}$ and $\mathcal{B}$ (over the same finite vocabulary)
- Two players: **I** (Spoiler) and **II** (Duplicator)
- Perfect information
- Move by **I** : select a structure and pick an element in it
- Move by **II** : pick an element in the opposite structure
- Round: a move by **I** followed by a move by **II**
- Game: sequence of rounds
- **II** tries to imitate **I**
- A player who cannot move loses

Basics

- Vocabulary: finite set of relation symbols
- $\mathcal{A}$ and $\mathcal{B}$ structures on the same vocabulary
- $\vec{a} = a_1, \ldots, a_k \in \mathrm{dom}(\mathcal{A})$
- $\vec{b} = b_1, \ldots, b_k \in \mathrm{dom}(\mathcal{B})$
- Configuration: $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$, with $|\vec{a}| = |\vec{b}|$
  - It represents the relation $\{ (a_i, b_i) \mid 1 \leqslant i \leqslant |\vec{a}| \}$

Definition

$(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is a partial isomorphism if it is an isomorphism of the substructures induced by $\vec{a}$ and $\vec{b}$, respectively.

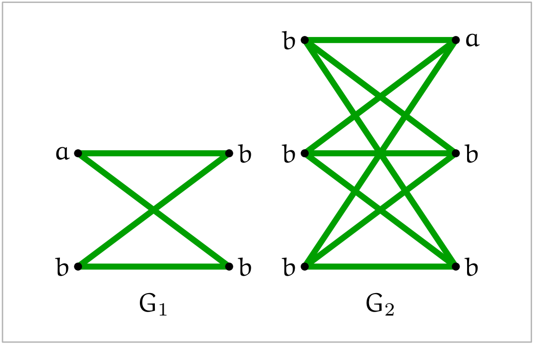
Winning strategies

- A play from $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ proceeds by extending the initial configuration with the pair of elements chosen by the two players, e.g.,
  - if **I** picks c in $\mathcal{A}$
  - and **II** replies with d in $\mathcal{B}$
  - then the new configuration is $(\mathcal{A}, \vec{a}, c, \mathcal{B}, \vec{b}, d)$
- Ending condition: a player repeats a move or the configuration is not a partial isomorphism

Definition

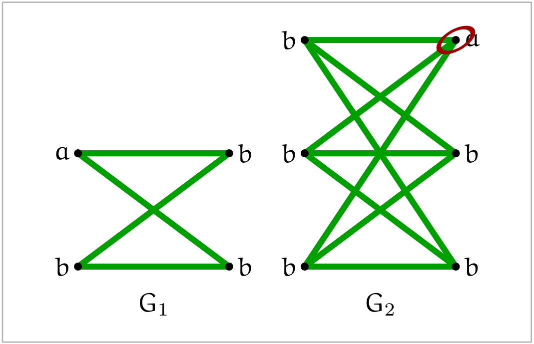
II has a winning strategy from $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ if every configuration of the game until an ending configuration is reached is a partial isomorphism, no matter how **I** plays.

An example on graphs



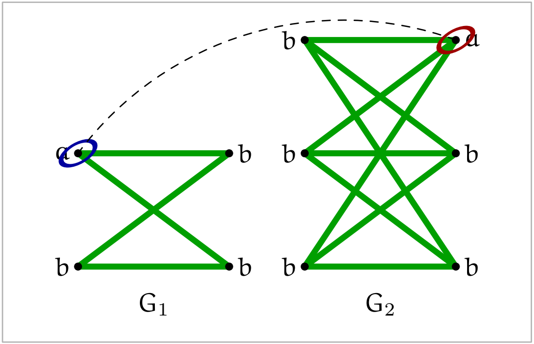
- **II** must respect the adjacency relation…
- …and pick nodes with the same label as **I** does

An example on graphs



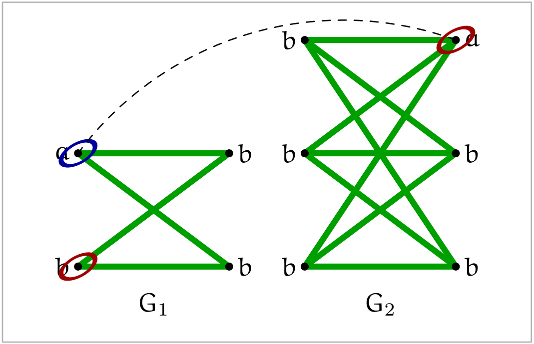
- **II** must respect the adjacency relation...
- ...and pick nodes with the same label as **I** does

An example on graphs



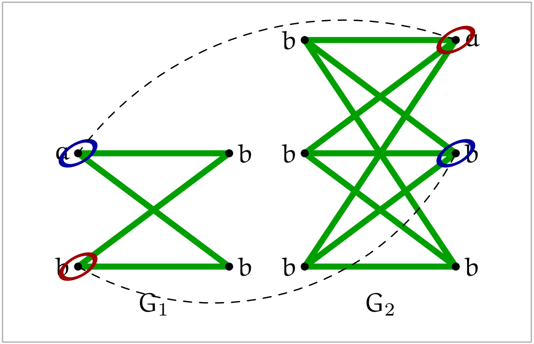
- **II** must respect the adjacency relation. . .
- . . . and pick nodes with the same label as **I** does

An example on graphs



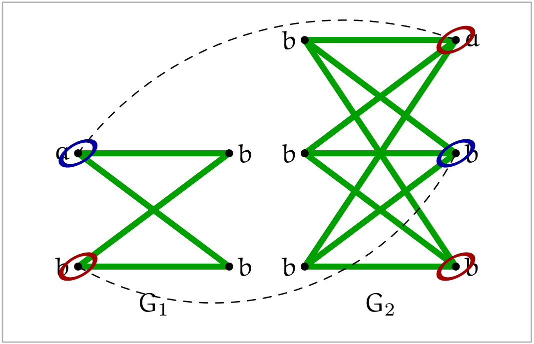
- **II** must respect the adjacency relation...
- ...and pick nodes with the same label as **I** does

An example on graphs



- **II** must respect the adjacency relation. . .
- . . . and pick nodes with the same label as **I** does

An example on graphs



- **II** must respect the adjacency relation. . .
- . . . and pick nodes with the same label as **I** does

Bounded and unbounded games

- Bounded game: $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$
  - the number of rounds is fixed: the game ends after m rounds have been played
- Unbounded game: $\mathcal{G}(\mathcal{A}, \mathcal{B})$
  - the game goes on as long as either a player repeats a move or the current configuration in not partial isomorphism
- **II** wins if and only if the ending configuration is a partial isomorphism

Unbounded games turn out to be useful to compare (finite) structures (comparison games): the remoteness of an unbounded game as a measure of structure similarity.

Main result

First-order EF-games capture m-equivalence

Theorem (Ehrenfeucht, 1961)

***II** has a winning strategy in $\mathcal{G}_m((\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b}))$ if and only if $(\mathcal{A}, \vec{a})$ and $(\mathcal{B}, \vec{b})$ satisfy the same FO formulas of quantifier rank m and at most $|\vec{a}|$ free variables, written $(\mathcal{A}, \vec{a}) \equiv_m (\mathcal{B}, \vec{b})$.*

Some simple consequences.

- If two structures A and B are m-equivalent for every natural number m, then they are elementarily equivalent
- In finite structures, A and B are elementarily equivalent if and only if they are isomorphic (in general, this is not the case: consider, for instance, $\mathbb{N}$ and the ordered sum $\mathbb{N} \triangleleft \mathbb{Z}$)

Correspondence between games and formulas

EF-games have a natural logical counterpart which is based on the following simple properties of **II** winning strategies.

Given $m \geqslant 0$, two structures $\mathcal{A}$ and $\mathcal{B}$, a tuple $\vec{a}$ of elements of A, and a tuple $\vec{b}$ of elements of B, we have that:

- **II** wins $\mathcal{G}_0((\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b}))$ iff $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is a partial isomorphism
- for every $m > 0$, **II** wins $\mathcal{G}_m((\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b}))$ iff
  - for all $a \in A$, there exists $b \in B$ such that **II** wins $\mathcal{G}_{m-1}(\mathcal{A}, \vec{a}, a, \mathcal{B}, \vec{b}, b)$
  - for all $b \in B$, there exists $a \in A$ such that **II** win $\mathcal{G}_{m-1}(\mathcal{A}, \vec{a}, a, \mathcal{B}, \vec{b}, b)$

From games to formulas: Hintikka formulas

Definition (Hintikka formulas)

Given a structure $\mathcal{A}$ and a tuple $\vec{a}$ of elements of A, let

$$\varphi^0_{(\mathcal{A},\vec{a})}(\vec{x}) \stackrel{\text{def}}{=} \bigwedge_{\substack{\varphi(\vec{x}) \text{ atomic} \\ (\mathcal{A},\vec{a}) \models \varphi(\vec{x})}} \varphi(\vec{x}) \;\wedge \bigwedge_{\substack{\varphi(\vec{x}) \text{ atomic} \\ (\mathcal{A},\vec{a}) \models \neg\varphi(\vec{x})}} \neg\varphi(\vec{x})$$

and, for $m \geqslant 0$,

$$\begin{aligned}\varphi^{m+1}_{(\mathcal{A},\vec{a})}(\vec{x}) \stackrel{\text{def}}{=} & \bigwedge_{a \in A} \exists x_{n+1}\, \varphi^m_{(\mathcal{A},\vec{a},a)}(\vec{x}, x_{n+1}) \wedge \\ & \forall x_{n+1} \bigvee_{a \in A} \varphi^m_{(\mathcal{A},\vec{a},a)}(\vec{x}, x_{n+1}).\end{aligned}$$

For each m, $\varphi^m_{(\mathcal{A},\vec{a})}(\vec{x})$ is the m-*Hintikka formula*.

From games to formulas: Hintikka formulas (cont.)

The Hintikka formula $\varphi^{0}_{(\mathcal{A},\vec{a})}(\vec{x})$ describes the isomorphism type of the substructure of $\mathcal{A}$ induced by $\vec{a}$.

In general, $\varphi^{m}_{(\mathcal{A},\vec{a})}(\vec{x})$ describes to which isomorphism types the tuple $\vec{a}$ can be extended in m steps by adding one element in each step. Since the vocabulary is finite, the above conjunctions and disjunctions are finite even if the structure is infinite.

Theorem

For any given $(\mathcal{A}, \vec{a})$ *and* $(\mathcal{B}, \vec{b})$, *we have*

$$(\mathcal{B}, \vec{b}) \models \varphi^{m}_{(\mathcal{A},\vec{a})}(\vec{x}) \iff (\mathcal{A}, \vec{a}) \equiv_m (\mathcal{B}, \vec{b}) \iff$$

II *has a winning strategy in* $\mathcal{G}_m((\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b}))$.

From games to formulas: Hintikka formulas (cont.)

A winning strategy for **I** in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$ can be converted into a first-order sentence of quantifier rank at most m that is true in exactly one of A and B (the Hintikka formula $\varphi^m_{(\mathcal{A},\vec{a})}(\vec{x})$ or the Hintikka formula $\varphi^m_{(\mathcal{B},\vec{b})}(\vec{x})$).

A class $\mathcal{K}$ of structures (on the same finite vocabulary) is FO-definable if and only if there is $m \in \mathbb{N}$ such that **I** has a winning strategy whenever $\mathcal{A} \in \mathcal{K}$ and $\mathcal{B} \notin \mathcal{K}$.

From differentiating formulas to games

- Let $\mathcal{A}$ and $\mathcal{B}$ be fixed
- Let ϕ be a formula with quantifier rank m
- Let $\mathcal{A} \models \phi$ but $\mathcal{B} \not\models \phi$
- Repeat m times:

❶ If $\phi = \forall x_1\, \psi$, let $\phi \leftarrow \neg\phi$ and swap $\mathcal{A}$ and $\mathcal{B}$

- So, ϕ holds in $\mathcal{A}$ but not in $\mathcal{B}$ and its first quantifier is $\exists$

❷ Let $\psi \leftarrow \psi\{x_1/\bar{c}_1\}$, with $\bar{c}_1$ a fresh constant symbol

❸ Let **I** pick a_1 in $\mathcal{A}$ such that $(\mathcal{A}, a_1) \models \psi[\bar{c}_1/a_1]$ (since $\mathcal{A} \models \phi$, such an a_1 must exist)

❹ Whatever b_1 **II** chooses in $\mathcal{B}$, $(\mathcal{B}, b_1) \not\models \psi[\bar{c}_1/b_1]$

❺ Let $\mathcal{A} \leftarrow (\mathcal{A}, a_1)$, $\mathcal{B} \leftarrow (\mathcal{B}, b_1)$ and $\phi \leftarrow \psi$

- Switching between models is encoded in ϕ as quantifier alternations

Example

Consider the formula for density:
$\phi = \forall x_1 \forall x_2 \exists x_3\,(x_1 < x_2 \rightarrow x_1 < x_3 < x_2)$,
which holds in $(\mathbb{Q}, <)$ but not in $(\mathbb{Z}, <)$.

1. $\exists x_1 \exists x_2 \forall x_3\,(x_1 < x_2 \wedge \neg(x_1 < x_3 < x_2))$
2. **I** chooses z in $(\mathbb{Z}, <)$
3. $(\mathbb{Z}, <, z) \models \exists x_2 \forall x_3\,(\bar{c}_1 < x_2 \wedge \neg(\bar{c}_1 < x_3 < x_2))\,[\bar{c}_1/z]$
4. **II** replies q in $(\mathbb{Q}, <)$
5. **I** chooses $z + 1$ in $(\mathbb{Z}, <)$
6. $(\mathbb{Z}, <, z, z{+}1) \models \forall x_3\,(\bar{c}_1 < \bar{c}_2 \wedge \neg(\bar{c}_1 < x_3 < \bar{c}_2))\,[\bar{c}_1/z, \bar{c}_2/z{+}1]$
7. **II** replies with $q' > q$ in $(\mathbb{Q}, <)$ (otherwise it loses immediately)
8. **I** chooses $\frac{q'-q}{2}$ in $(\mathbb{Q}, <)$
9. $(\mathbb{Q}, <, q, q', \frac{q'-q}{2}) \models \bar{c}_1 < \bar{c}_2 \rightarrow \bar{c}_1 < \bar{c}_2 < c_3\,[\bar{c}_1/q, \bar{c}_2/q', \bar{c}_3/(\frac{q'-q}{2})]$

Winning vs. optimal strategies

Winning strategy $\neq$ Optimal strategy

The distinction between winning and optimal strategies is essential in unbounded games:

- In unbounded EF-games on finite structures, **I** wins unless $\mathcal{A} \cong \mathcal{B}$
- "Play randomly" is a winning strategy for **I**
- But, how far actually is the end of a game?
- What are the *best* moves for **I** (and **II**)?

Remoteness

Optimal strategies (in combinatorial games) can be characterized in terms of remoteness:

- Current player has no legal moves from (the current configuration of) $\mathcal{G} \Rightarrow rem(\mathcal{G}) = 0$
- Current player can move to a configuration with even remoteness $\Rightarrow rem(\mathcal{G}) = 1 +$ least even remoteness

  Win Quickly!

- Current player can only move to configurations with odd remoteness $\Rightarrow rem(\mathcal{G}) = 1 +$ greatest odd remoteness

  Lose Slowly!

- The parity of the remoteness tells the winner

Win quickly, lose slowly!

Remoteness in EF-games:

- For EF-games, remoteness in terms of rounds, not moves
- Remoteness of $\mathcal{G}$: the minimum m such that **I** wins $\mathcal{G}_m$ (simplified definition under the hypothesis $\mathcal{A} \not\equiv \mathcal{B}$)
- Optimal **I** 's move: given a configuration $\mathcal{G}$, a move by **I** is *optimal* if and only if, whatever **II** replies, the remoteness of the resulting configuration is less than or equal to $rem(\mathcal{G}) - 1$.
- Optimal **II** 's move: given a configuration $\mathcal{G}$ and a move by **I** , a reply by **II** is *optimal* if and only if the remoteness of the resulting position is
  - $rem(\mathcal{G}) - 1$, if **I** 's move is optimal
  - $rem(\mathcal{G})$, otherwise

Applications of EF-games

EF-games have been exploited to prove some basic results about first-order logic:

- Hanf's theorem
- Sphere lemma
- Gaifman's theorem

EF-games have been extensively used to prove negative expressivity results (sufficient conditions suffice)

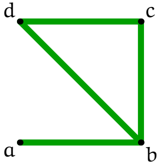
- Gaifman's theorem and normal forms for first-order logic

Gaifman graph

- Gaifman graph $G(\mathcal{A})$ of a structure $\mathcal{A}$: undirected graph $(\text{dom}(\mathcal{A}), E)$ where $(a, b) \in E$ iff a and b occur in the same tuple of some relation of $\mathcal{A}$
- the degree of a node a is the number of nodes $b (\neq a)$ such that $(a, b) \in E$ (the degree of G is the maximum of the degrees of its nodes)
- $\delta(a, b)$: length of the shortest path between a and b in $G(\mathcal{A})$

Example

$\mathcal{A} = (\{a, b, c, d\}, R, S)$, $R = \{(a, b)\}$, $S = \{(b, c, d)\}$
$\delta(a, c) = \delta(a, d) = 2$



$\mathfrak{r}$-sphere and $\mathfrak{r}$-neighborhood

Definition ($\mathfrak{r}$-sphere)

Let $\mathcal{A}$ be a structure with domain A, $a \in A$, and $r \in \mathbb{N}$. The r-sphere of a (in $\mathcal{A}$), denoted $S_r^{\mathcal{A}}(a)$, is defined as follows:

$$S_r^{\mathcal{A}}(a) \stackrel{\text{def}}{=} \{ b \in A \mid \delta(a, b) \leqslant r \}.$$

The notion of r-sphere can be extended to a vector $\vec{a} = a_1 \dots a_s$ (r-sphere $S_r^{\mathcal{A}}(\vec{a})$):

$$S_r^{\mathcal{A}}(\vec{a}) \stackrel{\text{def}}{=} \{ b \in A \mid \delta(\vec{a}, b) \leqslant r \} = S_r^{\mathcal{A}}(a_1) \cup \dots S_r^{\mathcal{A}}(a_s).$$

Definition ($\mathfrak{r}$-neighborhood)

The r-neighborhood $\mathcal{N}_r^{\mathcal{A}}(\vec{a})$ is the substructure of $\mathcal{A}$ induced by $S_r^{\mathcal{A}}(\vec{a})$.

Hanf's theorem

- $\mathcal{A} \leftrightarrows_r \mathcal{B}$: there is a bijection $f\colon A \to B$ such that $\mathcal{N}_r^{\mathcal{A}}(a) \cong \mathcal{N}_r^{\mathcal{B}}(f(a))$ for every $a \in A$.

Theorem (Hanf, 1965)

Let $\mathcal{A}$ and $\mathcal{B}$ be two structures such that for any $r \in \mathbb{N}$, each r-sphere in $\mathcal{A}$ or $\mathcal{B}$ contains finitely many elements. Then, $\mathcal{A}$ and $\mathcal{B}$ are elementarily equivalent if $\mathcal{A} \leftrightarrows_r \mathcal{B}$ for every $r \in \mathbb{N}$.

- Hanf's result does not hold if the Gaifman graph of (at least) one structure has infinite degree, e.g., the usual ordering relation on natural numbers
- Hanf's theorem is of interest only for infinite structures: two finite structures are elementarily equivalent if and only if they are isomorphic

Sphere theorem

- $\mathcal{A} \leftrightarrows_r^t \mathcal{B}$: isomorphic r-neighborhoods occur the same number of times in both structures or they occur more than t times in both structures

Theorem (Sphere theorem)

Given $\mathcal{A}$ and $\mathcal{B}$ with degree at most d and $m \in \mathbb{N}$, if $\mathcal{A} \leftrightarrows_r^t \mathcal{B}$ for $r = 3^m$ and $t = m \cdot d^{3^{m+1}}$, then $\mathcal{A} \equiv_m \mathcal{B}$.

- For all m there are r and t such that $\leftrightarrows_r^t$ is finer than $\equiv_m$ with respect to the class of structures with degree $\leqslant d$
- Strong hypotheses (it is a sufficient condition)
  - isomorphic neighborhoods
  - uniform threshold for all neighborhood sizes
  - scattering of neighborhoods is not taken into account
- Hanf's and Sphere Theorems proofs use EF-games

References for Hanf's and Sphere theorems

- W. Hanf
  Model-Theoretic Methods in the Study of Elementary Logic
  The Theory of Model, 1965

- W. Thomas
  On logics, tilings, and automata
  ICALP'91, 1991

- W. Thomas
  On the Ehrenfeucht-Fraïssé game in Theoretical Computer Science
  LNCS 668, 1993

- R. Fagin, L. J. Stockmeyer, and M. Y. Vardi
  On monadic NP vs monadic co-NP
  Information and Computation, 1995

Gaifman's theorem

- r-local formula: has "bounded" quantifiers:

$$\exists y\,(\delta(\vec{x}, y) \leqslant r \wedge \phi)$$
$$\forall y\,(\delta(\vec{x}, y) \leqslant r \rightarrow \phi)$$

- $\delta(\vec{x}, y) \leqslant r$ is FO-definable
- existentially r-local sentence:

$$\exists x_1 \cdots \exists x_s \left(\bigwedge_{1 \leqslant i < j \leqslant s} \delta(x_i, x_j) > 2r \wedge \bigwedge_{1 \leqslant i \leqslant s} \phi_r^{(i)}(x_i) \right)$$

where $\phi_r^{(i)}$ are r-local

Theorem (Gaifman's theorem)

Every first-order sentence is logically equivalent to a boolean combination of existentially local sentences.

Remarks on Gaifman's theorem

- Gaifman's normal form is effective
- Gaifman's proof uses EF-games to prove the invariant

$$\mathcal{N}^{\mathcal{A}}_{7^{m-i}-1}(a_1 \cdots a_i) \equiv_{f(i)} \mathcal{N}^{\mathcal{B}}_{7^{m-i}-1}(b_1 \cdots b_i)$$

- r-local formulas with $r \leqslant 7^{\mathsf{qr}(\phi)}$
- $f(i)$-equivalence instead of isomorphism
- first-order logic can only talk of scattered small substructures
- first-order logic can only express local properties

Expressive Power of First-Order Logic

First-order logic is at the same time

- too strong
  - any finite structure can be defined (up to isomorphism)
- too weak
  - natural properties cannot be expressed (such as, for instance, "the domain has even cardinality")

Weak does not necessarily mean bad
"weak expressive power can also be a good thing, as it implies transfer of properties across different situations. In non-standard arithmetic, one computes in the structure $\mathbb{N} \triangleleft \mathbb{Z}$ using the infinite numbers to simplify calculations, and then transfers the outcome back to $\mathbb{N}$, provided it is a first-order statement about $<$." (van Benthem's course on logical games, Chapter 2,"Model Comparison Games")

The EF-problem

Definition

The EF-problem is the problem of determining whether **II** has a winning strategy in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$, given $\mathcal{A}$, $\mathcal{B}$ and an integer m.

Sufficient conditions

Corollary (of Ehrenfeucht-Fraïssé's theorem)

*A class $\mathcal{K}$ of structures is not FO-definable if and only if, for all $m \in \mathbb{N}$, there are $\mathcal{A} \in \mathcal{K}$ and $\mathcal{B} \notin \mathcal{K}$ such that **II** has a winning strategy in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$.*

- Sufficient conditions allow us to prove negative expressivity results

Example

Let $\mathcal{L}_k \stackrel{\text{def}}{=} (\{1, \ldots, k\}, <)$. It is known that

$$n = p \text{ or } n, p \geqslant 2^m - 1 \Rightarrow \textbf{II} \text{ wins } \mathcal{G}_m(\mathcal{L}_n, \mathcal{L}_p)$$

"The class of linear orderings of even cardinality is not FO-definable": given m, choose $\tilde{n} = 2^m$ and $\tilde{p} = 2^m + 1$; **II** wins $\mathcal{G}_m(\mathcal{L}_{\tilde{n}}, \mathcal{L}_{\tilde{p}})$ (i.e., $\mathcal{L}_{\tilde{n}} \equiv_m \mathcal{L}_{\tilde{p}}$).

A library of sufficient conditions

- R. Fagin and L. J. Stockmeyer and M. Y. Vardi
  On monadic NP vs monadic co-NP
  Information and Computation, 1995
- T. Schwentick
  On winning Ehrenfeucht games and monadic NP
  Annals of Pure and Applied Logic, 1996
- S. Arora and R. Fagin
  On winning strategies in Ehrenfeucht-Fraïssé games
  Theoretical Computer Science, 1997
- H. J. Keisler and W. B. Lotfallah
  Shrinking games and local formulas
  Annals of Pure and Applied Logic, 2004

Arora and Fagin's condition

- "Approximately" isomorphic neighborhoods
- Still based on a multiplicity argument
- Neighborhoods must be tree-like structures

Definition (simplified for directed graphs)

- The $(m, 0)$-color of an element a is its label plus a description of whether it is a constant and whether it has a self-loop
- the $(m, r+1)$-color of a is its (m, r)-color plus a list of triples, one for each possible (m, r)-color τ:
  1. the number of elements b with (m, r)-color τ such that $E(a, b)$ but not $E(b, a)$, counted up to m
  2. the number of elements b with (m, r)-color τ such that $E(b, a)$ but not $E(a, b)$, counted up to m
  3. the number of elements b with (m, r)-color τ such that $E(a, b)$ and $E(b, a)$, counted up to m

Arora and Fagin's condition (cont.)

Let the color of a directed edge be the ordered pair of colors of its nodes.

Theorem

Let $\mathcal{A} = (A, E)$ and $\mathcal{B} = (B, E)$ be two structures of degree at most d, and let $m \in \mathbb{N}$. If

- *there is a bijection $f \colon A \to B$ such that a and $f(a)$ have the same (m, r)-color, with $r = 3^{2m}$, for all $a \in A$,*
- *$\mathcal{A}$ and $\mathcal{B}$ do not have (undirected) cycles of length less than r,*
- *whenever $E^{\mathcal{A}}(a, b)$ holds but $E^{\mathcal{B}}(f(a), f(b))$ does not hold, or vice versa, then there are at least d^r edges in both structures having the same (m, r)-color as (a, b), (resp., $(f(a), f(b))$),*

*then **II** has a winning strategy in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$.*

Applications of Arora and Fagin's condition

- Directed reachability is not in monadic Σ_1^1 (a simpler proof of Ajtai and Fagin's result)
- Graph connectivity is not in monadic Σ_1^1
- Both results can be shown to hold even if the vocabulary is expanded with particular built-in relations of degree $n^{o(1)}$, where n is the size of the structure
- The requirement of the absence of small cycles can be relaxed at the expense of adding further hypotheses

Schwentick's extension theorem

Schwentick's work moves from the following question: Under which conditions can a "local" strategy be extended?

He develops a method that allows, under certain conditions, the extension of a winning strategy for **II** on some small parts of two finite structures to a global winning strategy.

- The structures must be isomorphic except for some small parts, for which local winning strategies exist by hypothesis
- The advantage is that there are no further constraints, either on the degree or on the internal characteristics of the substructures.

Schwentick's extension theorem (cont.)

- Let $\mathcal{C}$ and $\mathcal{D}$ be subtructures of $\mathcal{A}$ and $\mathcal{B}$, respectively
- Suppose that **II** has a winning strategy in $\mathcal{G}_m(\mathcal{C}, \mathcal{D})$ for some m
- **II** has a winning strategy in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$ if
  1. **II**'s strategy for $\mathcal{G}_m(\mathcal{C}, \mathcal{D})$ can be extended to a winning strategy in $\mathcal{G}_m(\mathcal{N}^{\mathcal{A}}_{2^m}(C), \mathcal{N}^{\mathcal{B}}_{2^m}(D))$, so that, at every round the two chosen elements have the same distance from C and D, respectively
  2. there is an isomorphism α: $(A \setminus C) \to (B \setminus D)$ such that $\delta(x, C) = \delta(\alpha(x), D)$ for all $x \in \mathcal{N}^{\mathcal{A}}_{2^m}(C) \setminus C$

Proof's idea

- Divide the domains of the structures into three regions:
- inner area: $I = C \cup D$
- outer area: $O = (A \setminus \mathcal{N}^{\mathcal{A}}_{2^m}(C)) \cup (B \setminus \mathcal{N}^{\mathcal{B}}_{2^m}(D))$
- the area in between
- At each round, the inner or outer areas may grow, according to the played moves
- Separation invariant: after round i the distance from every element in the inner area and every element in the outer area is greater than 2^{m-i}
- So, the winning strategy for **II** is guaranteed by the isomorphism α in the outer area, and by the extended winning strategy in the inner area and the area in between

Extensions

- different distance functions can be used
- winning strategies for several pairs of substructures can be combined
- The separation invariant may be required for some relations, but not for others (e.g., linear ordering), by adding a kind of homogeneity condition that guarantees that elements in the inner and outer areas behave in the same way with respect to the relations that do not satisfy the separation invariant

Applications of Schwentick's extension theorem

- Connectivity of finite graphs is not expressible in monadic Σ_1^1 in the presence of built-in relations of degree $n^{o(1)}$ (the same result as Arora and Fagin's) or even in the presence of a built-in linear ordering
- Monadic Σ_1^1 with a built-in linear ordering is more expressive than monadic Σ_1^1 with a built-in successor relation

Shrinking games

- Similar to Schwentick's extension theorem, but it works in the opposite direction, by shrinking the playground according to a sequence of "scattering parameters"
- The authors use Ehrenfeucht–Fraïssé type games with a shrinking horizon between structures to obtain a spectrum of normal form theorems of the Gaifman type
- They improve the bound in the proof of Gaifman's theorem from $7^{qr(\phi)}$ to $4^{qr(\phi)}$ and they provide bounds for other normal form theorems

Shrinking games

- Let $\vec{s} = s_0, s_1, \ldots$ a possibly infinite sequence of natural numbers, called scattering parameters
- The sequence of local radii associated to $\vec{s}$ is defined as follows:

$$r_0 = 1$$
$$r_{n+1} = 2r_n + s_n$$

- A set C is s-scattered if $\delta(a, b) > s$ for all distinct $a, b \in C$
- A sequence $\vec{s}$ shrinks rapidly if $2r_j \leqslant s_j$ for all j
- Given $\vec{s} = s_0, s_1, \ldots$ that shrinks rapidly, if C is s_j-scattered then the r_j-neighborhood around any $c \in C$ does not contain any other element of C

Shrinking games: local rounds

Let $\vec{s} = s_0, s_1, \ldots$ be a sequence that shrinks rapidly

Definition ($\vec{s}$-shrinking game)

Given $\mathcal{A}$ and $\mathcal{B}$ and $m \in \mathbb{N}$, the m-round $\vec{s}$-shrinking game is as follows:

- **I** chooses $1 \leqslant i < m$ and plays either a local or a scattered round
- a local round is played as follows (assuming that **I** plays in $\mathcal{A}$):
  1. **I** chooses $a \in \mathcal{N}^{\mathcal{A}}_{r_i+s_i}(\vec{a})$
  2. **II** replies with $b \in \mathcal{N}^{\mathcal{B}}_{r_i+s_i}(\vec{b})$

Shrinking games: scattered rounds

- a scattered round is played as follows:
  1. **I** chooses a non-empty set of s_i-scattered elements $C \subseteq \mathcal{N}^{\mathcal{A}}_{r_i}(\vec{a})$ such that **II** has a winning strategy in each i round $(s_0, \ldots, s_{i-1})$-shrinking game from $(\mathcal{A}, c, \mathcal{A}, d)$ for $c, d \in C$ (if $|\vec{a}| = 0$ then **I** chooses $m - i$ elements in A)
  2. **II** replies with a non empty set of s_i-scattered elements $D \subseteq \mathcal{N}^{\mathcal{B}}_{r_i}(\vec{b})$ such that $|C| = |D|$
  3. **I** chooses $d \in D$
  4. **II** chooses $c \in C$
  5. the position is extended with (c, d) and i rounds are left
- The ending and winning conditions are as in standard EF-game

Theorem

*Let $m \in \mathbb{N}$ and let $\vec{s} = s_0, s_1, \ldots$ be a sequence that shrinks rapidly. If **II** has a winning strategy in the m-round $\vec{s}$-shrinking game for $\mathcal{A}$ and $\mathcal{B}$ then **II** has a winning strategy in $\mathcal{G}_m(\mathcal{A}, \mathcal{B})$.*

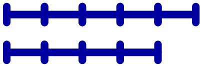
Sufficient vs "iff" conditions

$$\mathcal{L}_k \stackrel{\text{def}}{=} (\{1, \ldots, k\}, <)$$

It is known that

$$n = p \text{ or } n, p \geqslant 2^m - 1 \Rightarrow \textbf{II} \text{ wins } \mathcal{G}_m(\mathcal{L}_n, \mathcal{L}_p).$$

- Given $\mathcal{L}_5$ and $\mathcal{L}_6$, does **II** win $\mathcal{G}_3(\mathcal{L}_5, \mathcal{L}_6)$?



In fact,

$$n = p \text{ or } n, p \geqslant 2^m - 1 \Leftrightarrow \textbf{II} \text{ wins } \mathcal{G}_m(\mathcal{L}_n, \mathcal{L}_p)$$

- Given $\mathcal{L}_5$ and $\mathcal{L}_6$, does **II** win $\mathcal{G}_3(\mathcal{L}_5, \mathcal{L}_6)$? No!
- Complete characterizations are needed to exploit games algorithmically

Solving Games

Example

$$n = p \text{ or } n, p \leqslant 2^m - 1 \Leftrightarrow \textbf{II} \text{ wins } \mathcal{G}_m(\mathcal{L}_n, \mathcal{L}_p)$$

Assume $n < p$. Then:

1. The remoteness can be computed as:

$$\textit{rem}(\mathcal{G}(\mathcal{L}_n, \mathcal{L}_p)) = \lfloor \log(n+1) \rfloor + 1$$

2. A move of **I** from $\mathcal{G}(\mathcal{L}_n, \mathcal{L}_p)$ is optimal if and only if **I** picks
   - an element in $[\lfloor n/2 \rfloor + 1, p - \lfloor n/2 \rfloor - 1]$ in $\mathcal{L}_p$, or
   - $(n-1)/2$ in $\mathcal{L}_n$, if n is odd
3. Similarly, the set of **II**'s optimal replies can be computed

Complexity results

B. Khoussainov and J. Liu,
On Complexity of Ehrenfeucht-Fraïssé Games
LFCS, 2007, Annals of Pure and Applied Logic, in press

A. Montanari and A. Policriti and N. Vitacolonna,
An Algorithmic Account of Winning Strategies in Ehrenfeucht Games on Labeled Successor Structures
LPAR, 2005

E. De Maria, A. Montanari, N. Vitacolonna,
Games on Strings with a Limited Order Relation
LFCS, 2009

E. Pezzoli,
Computational Complexity of Ehrenfeucht-Fraïssé Games on Finite Structures
CSL, 1998

EF-games on specific classes

- Equivalence relations (with/without colors)
- Embedded equivalence relations
- Trees (with level predicates)
- Labelled successor structures
- Labelled linear structures with a bounded ordering

Equivalence relations: local strategy

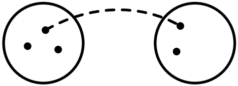
Definition

Structures $\mathcal{A} = (A, E)$, where E is an equivalence relation on A.

Definition

- For $m, n, t \in \mathbb{N}$, $m =_t n$ iff $m = n$ or both $m, n > t$
- $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is t-locally safe iff $\vec{a} \to \vec{b}$ is a partial isomorphism and $|[a_i]| =_t |[b_i]|$ for all $a_i \in \vec{a}$.

When a position is t-locally safe, there is not incentive for **I** to play in a class that has already been chosen, in a game with at most t rounds.



1-locally safe, but not 2-locally safe

Equivalence relations: "small disparity"

- $q_t^{(\mathcal{A},\vec{a})}$: number of classes of size t in $\mathcal{A}$ not containing any a_i (free classes)
- Let $\Delta^{(\mathcal{A},\vec{a})}_{(\mathcal{B},\vec{b})} = \{\, t \mid q_t^{(\mathcal{A},\vec{a})} \neq q_t^{(\mathcal{B},\vec{b})} \,\}$
- Let $q_t = \min\{q_t^{(\mathcal{A},\vec{a})}, q_t^{(\mathcal{B},\vec{b})}\}$

Lemma

*Given $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ and $t \in \Delta^{(\mathcal{A},\vec{a})}_{(\mathcal{B},\vec{b})}$, **I** can reach a position that is not t-locally safe after $q_t + 1$ rounds.*

Corollary

***I** has a winning strategy in at most $q_t + 1 + t$ rounds, with $t \in \Delta^{(\mathcal{A},\vec{a})}_{(\mathcal{B},\vec{b})}$.*

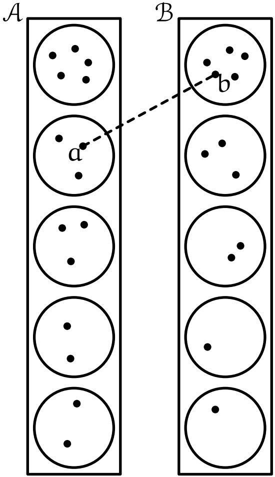
- **I** selects q_t distinct classes of size t ("global" moves)
- Then, he plays one more "global" move in a class of size t to which **II** cannot reply "appropriately"
- Then, he plays t rounds in the same class ("local" moves)

Example

- 2-locally safe, but not 3-locally safe

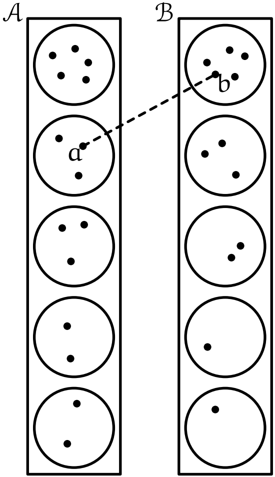
t	$q_t^{(\mathcal{A},a)}$	$q_t^{(\mathcal{B},b)}$
1	0	2
2	2	1
3	1	1
4	0	0
5	1	0

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$



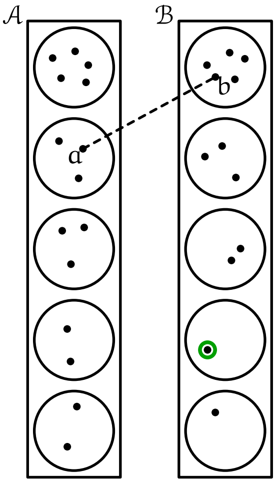
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0$, $q_2 = 1$, $q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



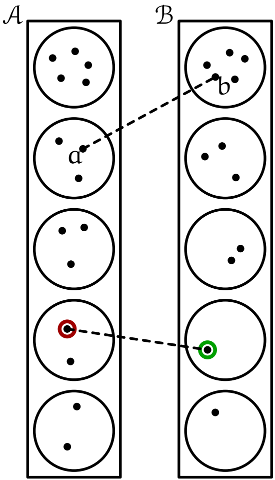
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1,2,5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ **I** can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ **I** can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ **I** can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



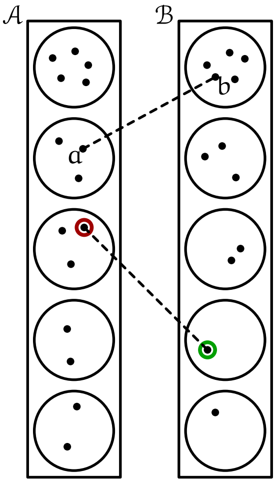
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



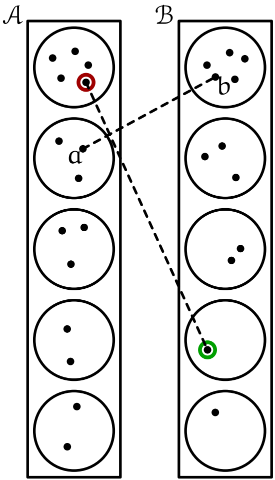
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



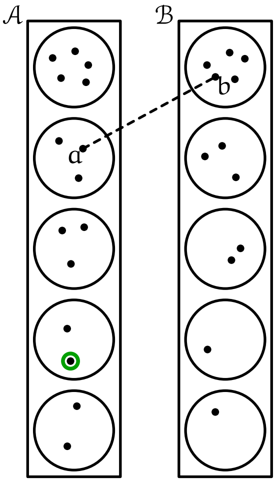
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



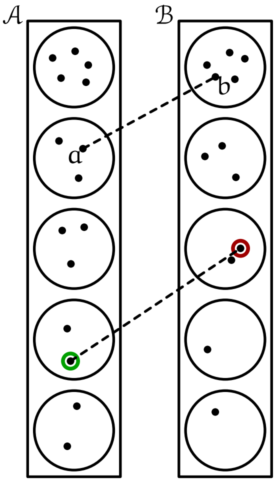
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



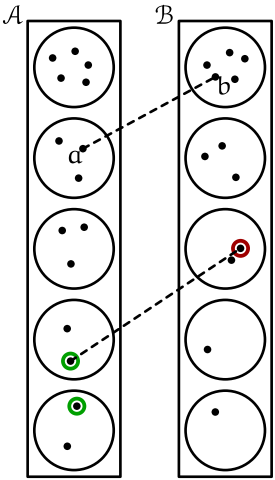
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



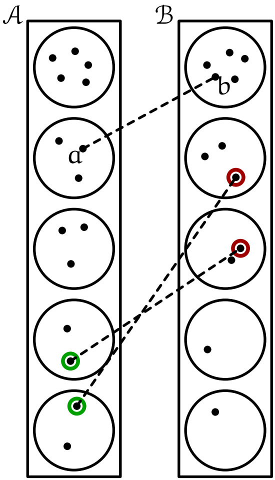
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



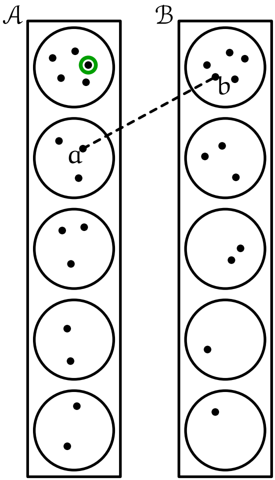
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



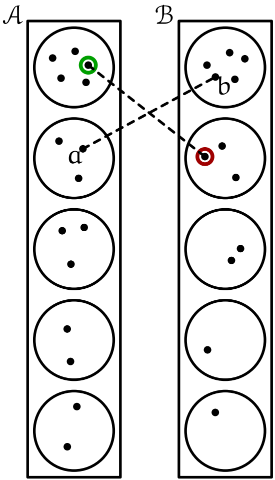
Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



Example (cont.)

- $\Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} = \{1, 2, 5\}$
- $q_1 = 0,\ q_2 = 1,\ q_5 = 0$
- $1 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 1-locally safe configuration in $q_1 + 1 = 1$ round
- $2 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 2-locally safe configuration in $q_2 + 1 = 2$ rounds
- $5 \in \Delta^{(\mathcal{A},a)}_{(\mathcal{B},b)} \Rightarrow$ I can reach a not 5-locally safe configuration in $q_5 + 1 = 1$ round



Equivalence relations: "large disparity"

- $q_{\geqslant t}^{(\mathcal{A},\vec{a})}$: number of free classes of size $\geqslant t$
- Let $\Gamma_{(\mathcal{B},\vec{b})}^{(\mathcal{A},\vec{a})} = \{\, t \mid q_{\geqslant t}^{(\mathcal{A},\vec{a})} \neq q_{\geqslant t}^{(\mathcal{B},\vec{b})} \,\}$
- Let $q_{\geqslant t} = \min\{q_{\geqslant t}^{(\mathcal{A},\vec{a})}, q_{\geqslant t}^{(\mathcal{B},\vec{b})}\}$

Lemma

Given $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ *and* $t \in \Gamma_{(\mathcal{B},\vec{b})}^{(\mathcal{A},\vec{a})}$, ***I*** *can reach a position that is not* t*-locally safe after* $q_{\geqslant t}$ *rounds.*

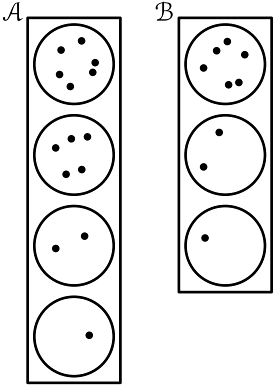
Corollary

I *has a winning strategy in at most* $q_{\geqslant t} + t$ *rounds, with* $t \in \Gamma_{(\mathcal{B},\vec{b})}^{(\mathcal{A},\vec{a})}$.

- **I** selects $q_{\geqslant t}$ distinct free classes of size $\geqslant t$ ("global" moves)
- Then, only one structure remains with a free class of size $\geqslant t$
- **I** plays t rounds in that class ("local" moves)

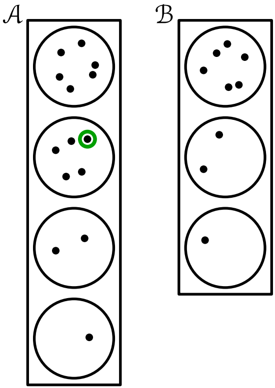
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let I pick a free class with $\geqslant t$ elements
- II replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- II replies with a "small" class
- I starts to play locally
- II must reply locally
- I wins
- $q_{\geqslant t} + t$ rounds needed



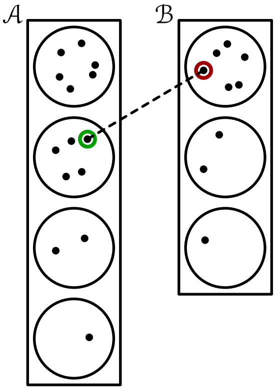
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a “small” class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



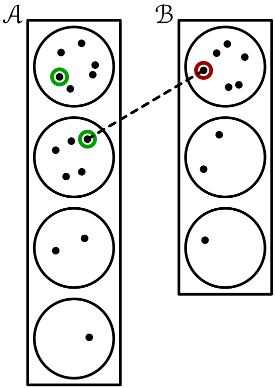
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a “small” class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



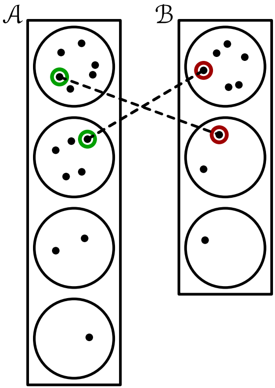
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a "small" class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



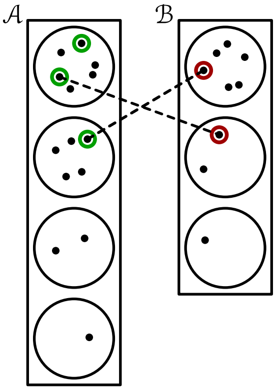
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a "small" class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



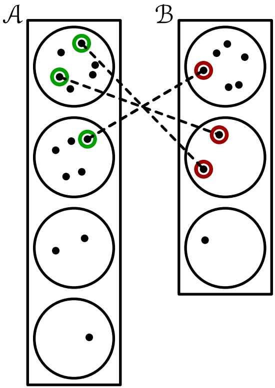
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a "small" class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



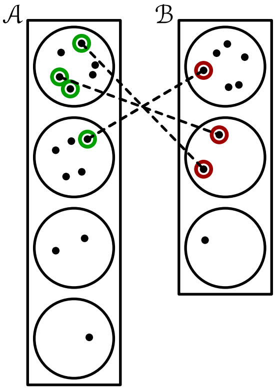
Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a "small" class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



Example

- Initially, empty configuration
- Let $t = 3$
- Then $q_{\geqslant t} = 1$
- let **I** pick a free class with $\geqslant t$ elements
- **II** replies accordingly
- Now there is a free class of size $\geqslant t$ only in $\mathcal{A}$
- **II** replies with a "small" class
- **I** starts to play locally
- **II** must reply locally
- **I** wins
- $q_{\geqslant t} + t$ rounds needed



Equivalence relations: characterization

Definition

Given $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ and $m \in \mathbb{N}$, $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is m-globally safe iff

- $q_t > m - t - 1$ for all $t \in \Delta^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})}$
- $q_{\geqslant t} > m - t$ for all $t \in \Gamma^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})}$

Theorem

II wins $\mathcal{G}_m(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ iff $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is m-locally safe and m-globally safe.

Corollary

The remoteness of $\mathcal{G}(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is

$$\min \left\{ \min\{\, t + q_{\geqslant t} \mid t \in \Gamma^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})} \,\}, 1 + \min\{\, t + q_t \mid t \in \Delta^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})} \,\} \right\}.$$

The remoteness can be computed in $O(|A| + |B|)$ time and space.

Sketch of the proof

Theorem

II *wins* $\mathcal{G}_m(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ *iff* $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ *is* m*-locally safe and* m*-globally safe.*

- If a position is m-locally safe and **I** play a local move, then **II** can reach a position $(m-1)$-locally safe
- If a position is m-globally safe, then **II** can reach a position $(m-1)$-globally safe
  - The only tricky case is when **I** chooses an element in a free class of size $t \in \Delta^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})}$ or $t \in \Gamma^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})}$
  - But, m-global safety allows **II** to reply properly
- The result easily extends to structures colored homogenously, i.e., if $E(x, y)$ then $x \in P \Leftrightarrow y \in P$, for all $x, y \in A$ and unary predicate P

Equivalence structures with one color

Definition

Structures $\mathcal{A} = (A, E, P)$, where E is an equivalence relation on A and P is a unary predicate.

Definition

- Let $P[a_i]$ be the set of elements $a_j \in [a_i]$, with $a_j \neq a_i$, such that $P(a_j)$ holds ("a_j is colored")
- Define $\neg P[a_i]$ similarly
- $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is t-locally safe iff $\vec{a} \to \vec{b}$ is a partial isomorphism and $|P[a_i]| =_t |P[b_i]|$ and $|\neg P[a_i]| =_t |\neg P[b_i]|$ for all $a_i \in \vec{a}$
- $q_\tau^{(\mathcal{A}, \vec{a})}$: number of "equivalent" free classes of $(\mathcal{A}, \vec{a})$ of "type τ"
- Let $\Delta^{(\mathcal{A}, \vec{a})}_{(\mathcal{B}, \vec{b})} = \{ \tau \mid q_\tau^{(\mathcal{A}, \vec{a})} \neq q_\tau^{(\mathcal{B}, \vec{b})} \}$

Embedded equivalence structures: local strategy

Definition
Structures $\mathcal{A} = (A, E_1, \ldots, E_h)$ where each E_i is an equivalence relation on A and $E_i \subseteq E_j$ for $i < j$.

- We consider the case $h = 2$
- Let $\mathcal{A} = (A, E_1, E_2)$ and $\mathcal{B} = (B, E_1, E_2)$

Definition
A local game on $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is a game played only within non-free equivalence classes, i.e., classes containing some $a_i \in \vec{a}$ or $b_i \in \vec{b}$.

Definition
$(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is t-locally safe iff **II** has a winning strategy in the t-round local game on $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$.

- t-round local games are characterized as in “flat” equivalence games

Embedded equivalence structures: global strategy

Definition

- Type of an E_2-class X of $\mathcal{A}$: $tp(X) = (q_1, \ldots, q_{|X|})$, where q_k is the number of E_1-classes of size k contained in X
- $tp(X) \equiv_t tp(Y)$ iff **II** wins $\mathcal{G}_t((X, E_1 \restriction X), (Y, E_1 \restriction Y))$
- (Free) t-multiplicity of type σ in $(\mathcal{A}, \vec{a})$:

$$q_{\sigma,t}^{(\mathcal{A},\vec{a})} \stackrel{\text{def}}{=} \left|\{\, Y \mid Y \text{ is a free } E_2\text{-class of } (\mathcal{A}, \vec{a}) \land tp(Y) \equiv_t \sigma \,\}\right|$$

- $\Delta_{(\mathcal{B},\vec{b})}^{(\mathcal{A},\vec{a})} = \{\, (\sigma, t) \mid q_{\sigma,t}^{(\mathcal{A},\vec{a})} \neq q_{\sigma,t}^{(\mathcal{B},\vec{b})} \}$

Lemma

Given $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ *and* $(\sigma, t) \in \Delta_{(\mathcal{B},\vec{b})}^{(\mathcal{A},\vec{a})}$, ***I*** *has a winning strategy in* $\min\{\, q_{\sigma,t}^{(\mathcal{A},\vec{a})}, q_{\sigma,t}^{(\mathcal{B},\vec{b})} \} + 1 + t$ *rounds.*

Trees with height h

Definition

A tree $\mathcal{T}$ is a pair $(T, \leqslant)$ where

1. $\leqslant$ is a partial ordering with a unique minimum
2. for all $x \in T$, $\{ y \mid y \leqslant x \}$ is finite and linearly ordered
3. maximal elements are leaves
4. Level of a node: distance from the root
5. Height of $\mathcal{T}$: number of levels -1

- $\mathcal{K}_h$: class of trees of height h
- $x \leqslant y$ iff x is an ancestor of y
- The idea of Khoussainov and Liu's paper is to map $\mathcal{K}_h$ into the class of embedded equivalence relations of height h
- Sounds nice!
- Unfortunately, it does not work (without a level predicate)

Mapping trees onto embedded equivalences

- $T' \stackrel{\text{def}}{=} T \cup \{ (x, a_x) \mid x \text{ is a leaf of } \mathcal{T} \}$
- E_1: minimal equivalence containing $\{ (x, a_x) \mid x \text{ is a leaf of } \mathcal{T} \}$
- E_{i+1}: minimal equivalence containing $E_i \cup (T_1 \times T_1) \cup \cdots \cup (T_k \times T_k)$, where $T_1, \ldots, T_k$ are the subtrees rooted at nodes of level $h - i + 1$
- $E_i \subseteq E_{i+1}$ (E_i is finer than E_{i+1})
- Embedded equivalence structure induced by $\mathcal{T}$:

$$\mathcal{A}(\mathcal{T}) \stackrel{\text{def}}{=} (T', E_1, \ldots, E_h)$$

Claim

❶ $\mathcal{T}_1 \cong \mathcal{T}_2$ *iff* $\mathcal{A}(\mathcal{T}_1) \cong \mathcal{A}(\mathcal{T}_2)$ *(ok!)*

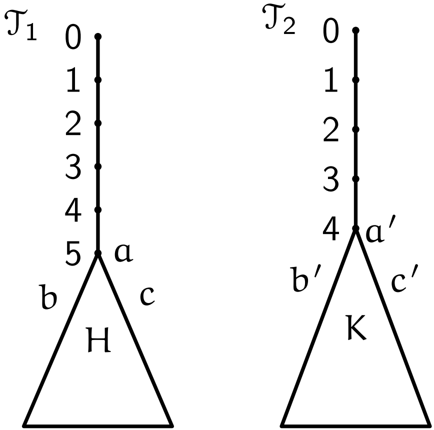
❷ ***II*** *wins* $\mathcal{G}_m(\mathcal{T}_1, \mathcal{T}_2)$ *iff* ***II*** *wins* $\mathcal{G}_m(\mathcal{A}(\mathcal{T}_1), \mathcal{A}(\mathcal{T}_2))$ *(wrong!)*

Why it does not work

Claim (wrong)

II *wins* $\mathcal{G}_m(\mathcal{T}_1, \mathcal{T}_2)$ *iff* ***II*** *wins* $\mathcal{G}_m(\mathcal{A}(\mathcal{T}_1), \mathcal{A}(\mathcal{T}_2))$.

- Observe that $x \leqslant y$ iff x has level t, y has level $s \leqslant t$ and $E_{h-t+1}(x, y)$



Binary trees

K. Doets
On n-Equivalence of Binary Trees
Notre Dame Journal of Formal Logic, 1987

This note presents a simple characterization of the class of all trees which are n-elementary equivalent with B_m: the binary tree with one root all of whose branches have length m (for each pair of positive integers n and m). [...] Section 2 introduces the class $Q(n)$ of binary trees and proves that every tree in it is n-equivalent with B_m whenever $m \geqslant 2^n - 1$. Section 3 shows that, conversely, each n-equivalent of a B_m with $m > 2^n - 1$ belongs to $Q(n)$. Finally, all n-equivalents of B_m for $m < 2^n - 1$ are isomorphic to B_m.

Labelled successor structures (LSS)

- Let Σ be a finite alphabet
- Let $u \in \Sigma^*$ be a word on Σ
- Let $u[i]$ be the ith letter of u

Definition

A (labelled) successor structure is a pair $(u, \vec{i})$, where the elements of $\vec{i}$ are distinguished indices of u.
Successor structures $(u, \vec{i})$ interpret FO-formulas $\phi(\vec{x})$ in the vocabulary $(=, s, (P_a)_{a \in \Sigma})$ according to the following rules:

$$
\begin{aligned}
(u, \vec{i}) &\models x_h = x_l && \text{if } i_h = i_l; \\
(u, \vec{i}) &\models s(x_h, x_l) && \text{if } i_l = i_h + 1; \\
(u, \vec{i}) &\models P_a(x_h) && \text{if } u[i_h] = a.
\end{aligned}
$$

Local conditions

$$\eta(i, j) = \begin{cases} j - i & \text{if } |i - j| \leqslant d; \\ \infty & \text{otherwise.} \end{cases}$$

Definition

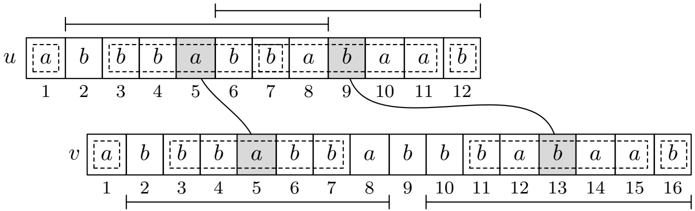
A configuration $(u, \vec{i}, v, \vec{j})$ is t-locally safe iff, for all $i_h, i_l \in \vec{i}$,

1. $\eta_{2^t}(i_h, i_l) = \eta_{2^t}(j_h, j_l)$
2. $\mathcal{N}^u_{2^t-1}(i_h) = \mathcal{N}^v_{2^t-1}(j_h)$

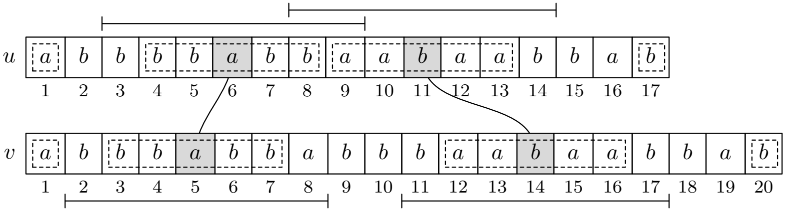
- If a configuration is not t-locally safe, **I** has a "local" winning strategy in t rounds
- **II** can turn a t-locally safe configuration into a $(t-1)$-locally safe configuration if **I** plays "locally"

Local safety: an example

Not 2-locally safe:



2-locally safe:

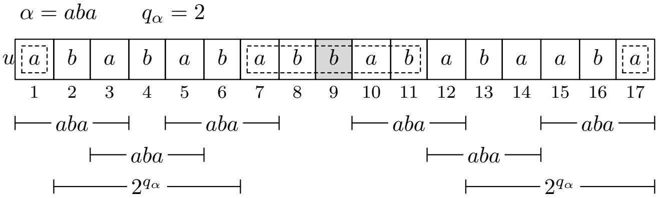


Free factors

Definition

- Let α be a word of length $2^t - 1$
- An occurrence of α centered at index k in $(u, \vec{i})$ is free iff $|k - \vec{i}| > 2^{t-1}$
- (Free) multiplicity of α in $(u, \vec{i})$: number of free occurrences of α in $(u, \vec{i})$
- Scattering of α in $(u, \vec{i})$: cardinality of a maximal 2^t-scattered subset of the free occurrences of α in $(u, \vec{i})$
- (A set $X \in \mathbb{N}$ is d-scattered iff $|x - y| > d$ for all $x, y \in X$)

Multiplicity and scattering: an example



- Let $\alpha = aba$ ($t = 2$)
- Centers of free occurrences of aba in $(u, 9)$: $\{2, 4, 6, 13, 16\}$
- Multiplicity: 5
- Scattering: 2 ($\{2, 4, 6\}$, $\{13, 16\}$)

LSS: Characterization

- Let $p_\alpha^{(u,\vec{i})}$ denote the free multiplicity
- Let $q_\alpha^{(u,\vec{i})}$ denote the scattering
- Let $\Delta_{(v,\vec{j})}^{(u,\vec{i})} = \{\, \alpha \mid p_\alpha^{(u,\vec{i})} \neq p_\alpha^{(v,\vec{j})} \vee q_\alpha^{(u,\vec{i})} \neq q_\alpha^{(v,\vec{j})} \,\}$
- $\Delta_{(v,\vec{j})}^{(u,\vec{i})}$ is the set of words that **I** can potentially exploit in order to win
- Let $q_\alpha = \min\{\, q_\alpha^{(u,\vec{i})}, q_\alpha^{(v,\vec{j})} \,\}$

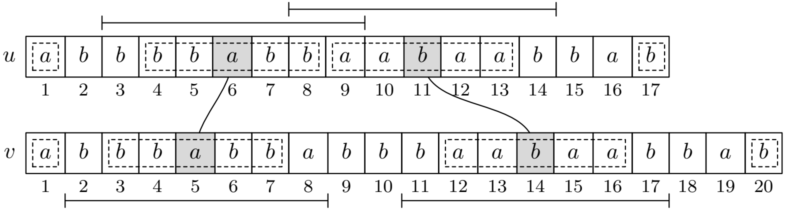
Definition

A configuration $(u, \vec{i}, v, \vec{j})$ is m-globally safe iff $q_\alpha > m - \log_2(|\alpha| + 1)$ for all words $\alpha \in \Delta_{(v,\vec{j})}^{(u,\vec{i})}$.

Theorem

II *has a winning strategy in* $\mathcal{G} = \mathcal{G}_m(u, \vec{i}, v, \vec{j})$ *iff* $\mathcal{G}$ *is* m*-locally safe and* m*-globally safe.*

Example



	α	$p_\alpha^{(u,6,11)}$	$p_\alpha^{(v,5,14)}$	$q_\alpha^{(u,6,11)}$	$q_\alpha^{(v,5,14)}$
$q=1$	a	4	5	4	5
	b	7	9	4	5
$q=2$	abb	2	3	2	3
	bab	1	2	1	2
	bba	1	2	1	2
	bbb	1	1	1	1

It is also 2-globally safe!

Definability and $\mathfrak{m}$-equivalence

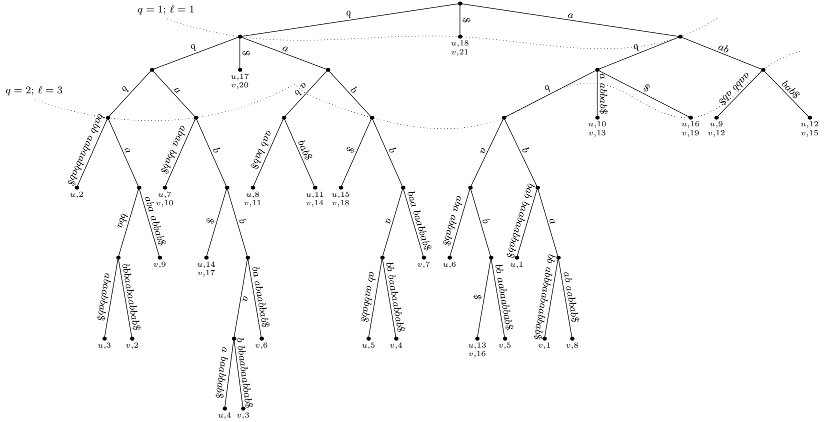
$\mathcal{L}$	Definable class	$\mathfrak{m}$-equivalence
FO(s)	threshold locally testable	Previous theorem

- From FO(s) to FO($<$):

$$\text{FO}(<_p),\ \text{where } x <_p y \Leftrightarrow 0 < y - x \leqslant p.$$

$\mathcal{L}$	Definable class	$\mathfrak{m}$-equivalence
FO($<$)	$*$-free	$<_p$, with $p \to \infty$

Testing $\equiv_m$ with generalized suffix trees



- Let $n = |u| + |v|$
- Remoteness of $\mathcal{G}(u, v)$: $O(n \log n)$ time and space
- **I**'s optimal moves: $O(n^2 \log n)$ time, $O(n \log n)$ space
- **II**'s optimal moves: $O(n)$ time and space (if the remoteness is known)

An emerging pattern

Let $\mathcal{A}$ and $\mathcal{B}$ arbitrary structures.

Definition

A t-round local game on $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is a game played on $\mathcal{N}^{\mathcal{A}}_{2^t-1}(\vec{a})$ and $\mathcal{N}^{\mathcal{B}}_{2^t-1}(\vec{b})$ such that, at round $t-k+1$, with $1 \leqslant k \leqslant t$, **I** must choose an element at distance at most 2^{k-1} from $\vec{a}$ or from $\vec{b}$.

Definition

A configuration $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$ is t-locally safe if **II** has a winning strategy in the t-round local game on $(\mathcal{A}, \vec{a}, \mathcal{B}, \vec{b})$.

- We write $(\mathcal{A}, \vec{a}) \equiv^{\mathsf{loc}}_t (\mathcal{B}, \vec{b})$
- **II** can play t rounds provided that **I** plays "near" distinguished elements (nearer and nearer after each round)

How to count neighborhoods?

- The analysis of equivalence structures shows that we need to count up to isomorphism and up to $\equiv_t^{\mathsf{loc}}$-equivalence (in equivalence structures, neighborhoods coincide with equivalence classes; two equivalence classes are isomorphic iff they have the same number of elements and they are $\equiv_t^{\mathsf{loc}}$-equivalent iff they both have at least t elements)
- The analysis of labelled successor structures shows that we need to count both the (free) multiplicity and the scattering of neighborhoods (for equivalence structures, the two notions collapse into one)

Conjecture

Counting the multiplicity and scattering of "small" neighborhoods up to isomorphism and up to $\equiv_t^{loc}$-equivalence is enough for characterizing the "global" winning strategy for arbitrary structures.

Complexity of the EF-Problem

- It is easy to prove that the problem is in PSPACE
- The difficult part is proving hardness for PSPACE
- The problem is in fact PSPACE-complete
- It is proved by reducing QBF (Quantified Boolean Formula) to the problem of determining whether **II** has a winning strategy
- QBF formulas have the form

$$\exists x_1 \forall x_2 \exists x_3 \cdots Q x_n \, (C_1 \wedge \cdots \wedge C_k)$$

where each C_i is a disjunction of literals

The EF-problem is PSPACE-complete

Theorem (Pezzoli)

The EF-problem for finite structures over any fixed signature that contains at least one binary and one ternary relation is PSPACE-complete.

- The proof for hardness goes by reducing QBF to the EF-problem
- Given a QBF formula ϕ of the form

$$\exists x_1 \forall x_2 \cdots \exists x_{2r-1} \forall x_{2r} \, (C_1 \wedge \cdots \wedge C_n),$$

  we build two structures $\mathcal{A}$ and $\mathcal{B}$ over $\Sigma = \{E, H\}$, where E is binary and H is ternary, such that **I** wins $\mathcal{G}_{2r+1}(\mathcal{A}, \mathcal{B})$ iff ϕ is satisfiable

Sketch of the proof

- **I**'s moves correspond to existential quantifiers
- **II**'s moves correspond to universal quantifiers
- Structures $\mathcal{A}$ and $\mathcal{B}$ consist of r blocks
- Each block is made of a certain number of subgraphs, called "gadgets", which are of three types: J, L, and I
- Some elements of the domains are labelled by truth values or pairs of truth values
- Some elements in the last block (block r) are labelled by clauses of ϕ
- A pair of consecutive rounds i, $i+1$ is played within block $\lceil i/2 \rceil$ and corresponds to instantiate a pair of consecutive variables $\exists x_i \forall x_{i+1}$

Sketch of the proof (cont.)

- At round i, **I** assigns the truth value T (resp., F) to variable x_i by choosing an element in block $\lceil i/2 \rceil$ of one of the structures (say, $\mathcal{A}$) "labelled" by T (resp., F)
- **II** is forced to reply by choosing an element "labelled" by a pair of truth values TT or TF (resp., FT or FF) in $\mathcal{B}$, which corresponds to recording **I**'s assignment (the first truth value) and to assign a truth value to variable x_{i+1} (the second truth value)
- At round $i + 1$, **I** chooses an "unlabelled" element in $\mathcal{B}$
- **II** is forced to reply by recording the truth value of x_{i+1} in $\mathcal{A}$ by choosing an element "labelled" the same as the second truth value chosen at round i

Sketch of the proof (cont.)

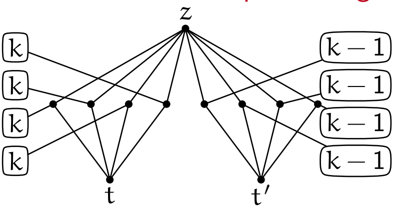
E.g., the pair of rounds may go like this:

round i	round $i+1$	
$s : T(x_i)$	$d : F(x_i)$	$\mathcal{A}$
$d : TF(x_i x_{i+1})$	$s : r$	$\mathcal{B}$

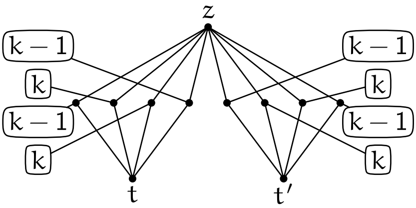
- The "labelling" is encoded by a ternary relation H such that $H(u, v, w)$ holds iff
  - u and v are adjacent in the same block
  - w is in the last block and is labelled by clause C_k
  - Clause C_k is made true by the truth values that label u and/or v

Gadgets J_k, L_k

Circled node are special neighbours



Gadget J_k

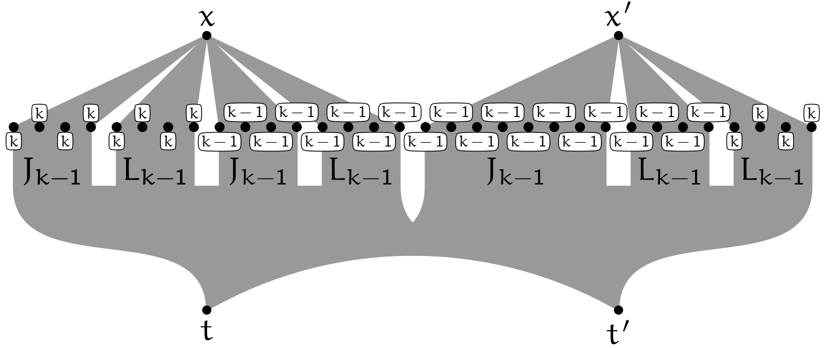


Gadget L_k

- four nodes in the middle have k special neighbours and target t
- four nodes in the middle have $k-1$ special neighbours and target t'

- four nodes in the middle have k special neighbours (two with target t and two with target t')
- four nodes in the middle have $k-1$ special neighbours (two with target t and two with target t')

Gadget I_k



- x is linked to 16 nodes
- x' is linked to other 16 nodes
- Each node in the middle is the source of a gadget J_{k-1} or L_{k-1}
- All gadgets share the same two targets t and t'
- Each node in the middle has either k or $k-1$ special neighbors
- I_k is symmetric if I_k's special neighbors are removed

Forcing pairs

Lemma

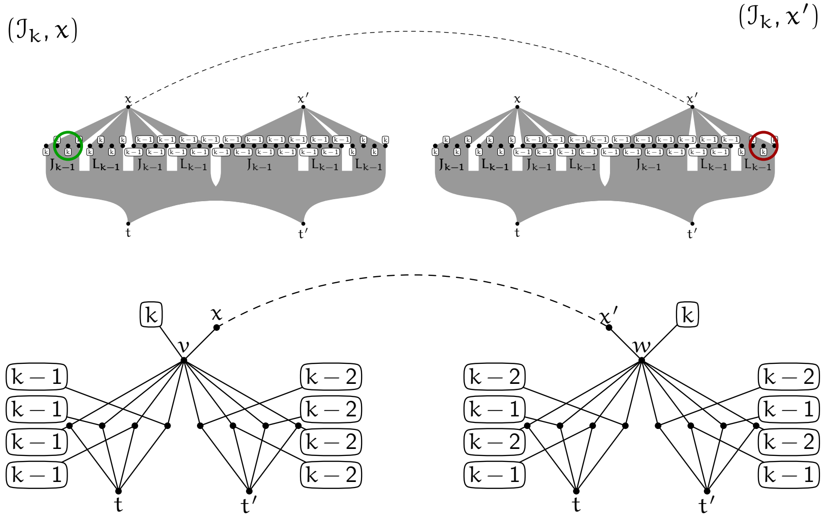
*In the $(k+1)$-moves EF-game on (I_k, x, I_k, x'), **I** can force the pair (t, t'), but **II** has a winning strategy in the k-moves EF-game that allows him to answer t with t and t' with t'.*

- In the $(k+1)$-moves game **I** starts by playing $v = kxJ$ (i.e., v has k special neighbors, it is adjacent to x and it is the source of a gadget J_{k-1})
- **II** must answer with $w = kx'L$
  - otherwise, **I** wins by moving into the special neighbors
- **I** chooses $w(k-1)t'$ in L_{k-1}
- **II** must answer $v(k-1)t$ in J_{k-1}

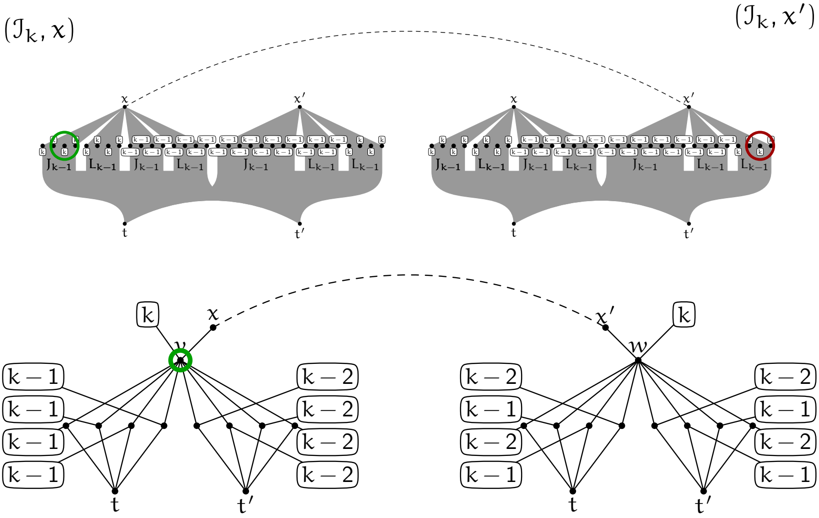
Remark

The above lemma says nothing about who has a winning strategy.

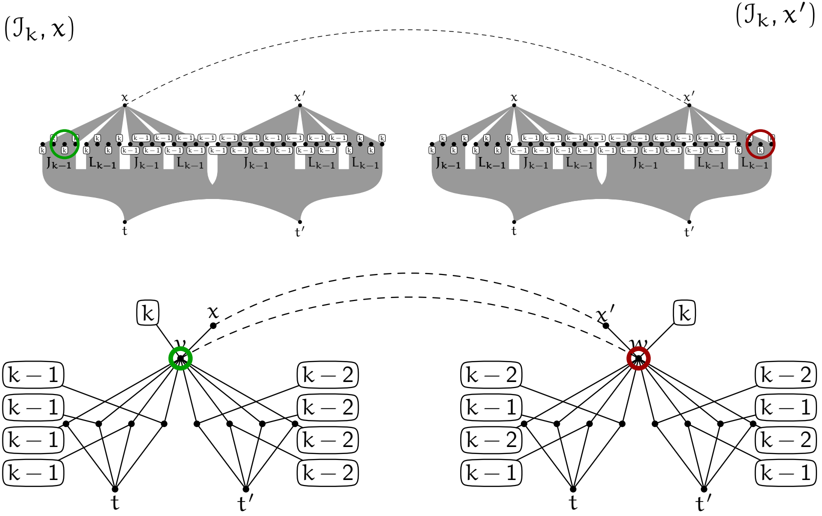
Forcing pairs (cont.)



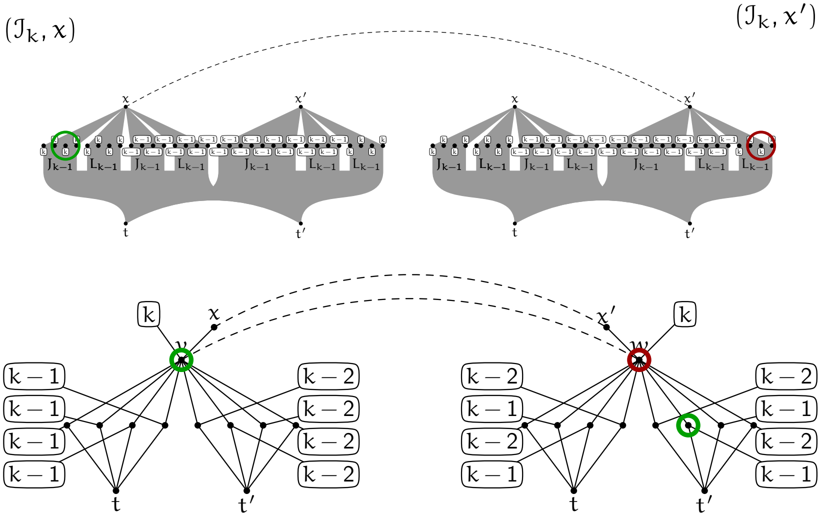
Forcing pairs (cont.)



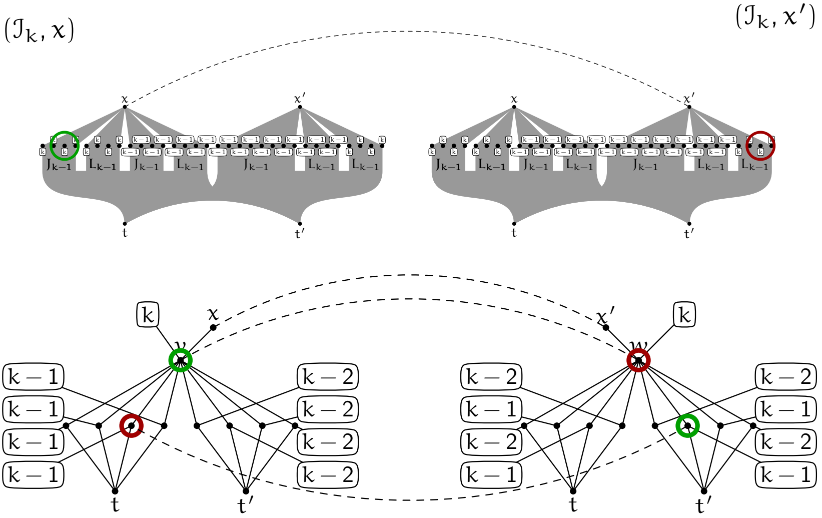
Forcing pairs (cont.)



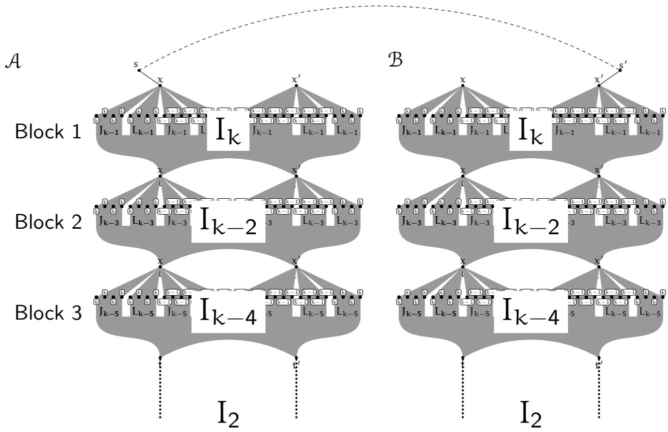
Forcing pairs (cont.)



Forcing pairs (cont.)



The whole structure



- Up to now, $\mathcal{A}$ and $\mathcal{B}$ are $(2r+1)$-equivalent
- A small modification of the structure is made
- A ternary relation is introduced, which establishes a correspondence between a winning strategy for **I** and the satisfiability of a given formula

The ternary relation H

- t and t' in the last block are replaced by two sets of elements labelled by clauses of ϕ
- $H(u, v, w)$ holds iff u and v are consecutive in the same block i, w is in the last block and it is labelled by a clause C_k and one of the following holds:
  - u is labelled $a \in \{T, F\}$, v is labelled $b \in \{T, F\}$, or
  - u is labelled ab, with $a, b \in \{T, F\}$, v is not labelled, or
  - u is labelled ac, v is labelled b, with $a, b, c \in \{T, F\}$,
- and assigning a to x_i and b to x_{i+1} makes C_k true

Lawful strategies for **I**

- **I** starts playing in $\mathcal{A}$
- Then, **I** will play in $\mathcal{A}$ at every odd round and in $\mathcal{B}$ at every even round
- Besides, **I** plays on the "left" of $\mathcal{A}$ in odd rounds and on the "right" of $\mathcal{B}$ in even rounds
- At each odd round, **II** is forced to record **I**'s choice in $\mathcal{B}$, i.e., if **I** picks an element labelled T in $\mathcal{A}$ then **II** must reply with TT or TF, but not with FF or FT (otherwise, she is bound to lose in less than $2r + 1$ rounds)
- Similarly, **II** is forced to record its choice in $\mathcal{A}$ at the next round, i.e., if she has chosen TF in $\mathcal{B}$ then she will pick an element labelled by F in $\mathcal{A}$
- If **II** fails to play like that, at some following round **I** may pick an element labelled by a clause C that appears in some triple of H, but **II** would not be able to do so in the opposite structure

Example

Example:
let $\phi = \exists x_1 \exists x_2 \left((x_1 \vee \bar{x}_2) \wedge \bar{x}_1 \right)$.
Suppose that during a game the following labelling is determined:

round 1	round 2	
$s : \mathsf{F}(x_1)$	$d : \mathsf{F}(x_2)$	$\mathcal{A}$
$d : \mathsf{TF}(x_1 x_2)$	$s : r$	$\mathcal{B}$

Note that **II** has not recorded the correct move made by **I**. At last round (round 3), **I**, instead of playing an unlabelled element, chooses clause $\bar{x}_1$ in $\mathcal{A}$, which determines a triple in H. **II**, however, cannot put any tuple in H in $\mathcal{B}$.

Complexity results for pebble games

- Pebble games are a variant of EF-games in which each player has a limited number of pebbles and re-use them
- They correspond to formulas with a bounded number of variables

Theorem

*Given a positive integer k and structures $\mathcal{A}$ and $\mathcal{B}$ the problem of determining whether **II** has a winning strategy in the existential k-pebble game on $\mathcal{A}$ and $\mathcal{B}$ is EXPTIME-complete.*

Corollary

All algorithms for determining whether k-strong consistency can be established are inherently exponential.

P. G. Kolaitis, J. Panttaja
On the Complexity of Existential Pebble Games
CSL 2003

The proof of EXPTIME-completeness is not that easy. . .

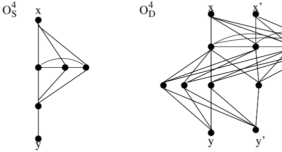


Fig. 3. Single Input One-Way Switch O^4

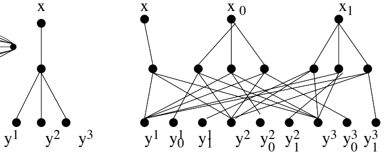
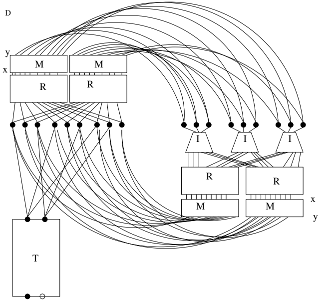


Fig. 6. I^3 Gadget



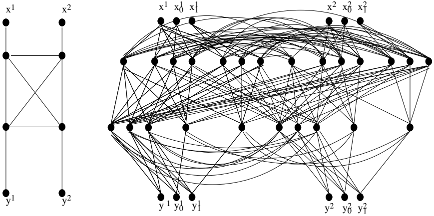


Fig. 7. A subgraph of M^4

Fig. 10. This is component decomposition of the Duplicator's graph for the reduction

Conclusions

- EF-games not explored much algorithmically
  - What is the complexity of the EF-problem for (labelled) arbitrary trees?
  - What is complexity of the EF-problem for signature containing only a binary relations E (i.e., graphs)?
  - The question for the complexity of first-order equivalence for finite structures, that is, isomorphism, is open (strictly related to the graph isomorphism problem)
- Simpler proofs?
- May notions from Combinatorial Game Theory help?
  - Berlekamp's et al. *Winning Ways*