Introduction to Descriptive Complexity

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 600

1 Ehrenfeucht-Fraïssé Games
- Definition of the Games
- Methodology for First-Order Expressibility
- First-Order Properties Are Local
- Bounded Variable Languages
- Zero-One Laws
- Ehrenfeucht-Fraïssé Games with Ordering

Subsection 1

Definition of the Games

The Game

- The game $\mathcal{G}_m^k$ is played by two players, Spoiler and Duplicator, on a pair of structures $\mathcal{A}$ and $\mathcal{B}$ of the same vocabulary τ.
- $\mathcal{G}_m^k$ is played for m rounds, using k pairs of pebbles.
- Spoiler tries to point out a difference between the two structures.
- Duplicator tries to match his moves so that the differences between them are hidden.
- At each move, Spoiler places one of the pebbles on an element of the universe of one of the two structures, i.e., he places pebble i on an element of $|\mathcal{A}|$ or $|\mathcal{B}|$.
- Duplicator then responds by placing the other pebble i on an element of the other structure.

Configurations

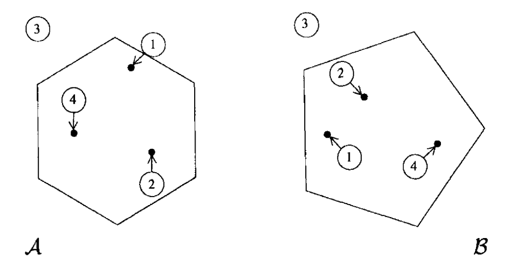
- The position of the game right after move r is denoted by (α_r, β_r).
- Such a k-**configuration** of $\mathcal{A}$, $\mathcal{B}$ is a pair of partial functions

$$\alpha : (\text{const}(\tau) \cup \{x_1, x_2, \ldots, x_k\}) \to |\mathcal{A}|$$
$$\beta : (\text{const}(\tau) \cup \{x_1, x_2, \ldots, x_k\}) \to |\mathcal{B}|,$$

  where we require that:
  - The domains of the functions α and β be equal, $\text{dom}(\alpha) = \text{dom}(\beta)$;
  - For all $c \in \text{const}(\tau)$, $\alpha(c) = c^{\mathcal{A}}$ and $\beta(c) = c^{\mathcal{B}}$.
- The meaning of $\alpha_r(x_i) = a$ and $\beta_r(x_i) = b$ is that just after move r, the i-th pebbles are sitting on $a \in |\mathcal{A}|$ and $b \in |\mathcal{B}|$.
- Some variable x_i is not in the domain of α_r iff, just after move r, the i-th pebbles are off the board.
- The valid positions of game $\mathcal{G}_m^k$ on $\mathcal{A}$, $\mathcal{B}$ consist of any possible k-configuration on $\mathcal{A}$ and $\mathcal{B}$.

Example

- Consider the following figure.



The current configuration has

$$\text{dom}(\alpha) = \text{dom}(\beta) = \{1, 2, 4\} \cup \text{const}(\tau).$$

So pebbles 1, 2 and 4 are currently placed on elements of $|\mathcal{A}|$, $|\mathcal{B}|$.

Both pebbles numbered 3 are off the board.

Treatment of Constants

- We denote by
$$\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$$
the k-pebble, m-move game on $\mathcal{A}, \mathcal{B}$, with initial configuration (α_0, β_0).
- $\mathcal{G}_m^k(\mathcal{A}, \mathcal{B})$ is the game in which all the pebbles start off the board.
- That is, in $\mathcal{G}_m^k(\mathcal{A}, \mathcal{B})$, $\mathrm{dom}(\alpha_0) = \mathrm{dom}(\beta_0) = \mathrm{const}(\tau)$.
- The reason we include the constants in the domain of every configuration is to make our treatment simpler.
- As will be seen, in Ehrenfeucht-Fraïssé games, constants behave exactly like pebbles that are fixed at the beginning of the game.

The Next Configuration

- At each move r, $1 \leq r \leq m$, Spoiler picks up a pair of pebbles and places one of them on an element of one of the two structures.
- Duplicator must then answer by placing the other pebble of the pair on an element of the other structure.
- Thus, for some $i \in \{1, 2, \ldots, k\}$, pair i of pebbles is placed on $a \in |\mathcal{A}|$ and $b \in |\mathcal{B}|$.
- Define the **next configuration** $(\alpha_r, \beta_r) = (\alpha_{r-1}[a/x_i], \beta_{r-1}[b/x_i])$,

$$\alpha_r(x_j) = \begin{cases} \alpha_{r-1}(x_j), & \text{if } j \neq i \\ a, & \text{if } j = i \end{cases}, \quad \beta_r(x_j) = \begin{cases} \beta_{r-1}(x_j), & \text{if } j \neq i \\ b, & \text{if } j = i \end{cases}$$

- Just after move r, the configuration α_r, β_r determines a relation

$$\beta_r \circ \alpha_r^{-1} \subseteq |\mathcal{A}| \times |\mathcal{B}|.$$

Induced Substructures

- The **induced substructure** $\langle \text{rng}(\alpha)\rangle^{\mathcal{A}}$ has universe the closure of $\text{rng}(\alpha)$ under all the functions of $\mathcal{A}$.
- When τ has no function symbols, this simply means that we add all the constants to $\text{rng}(\alpha)$.
- For $\tau = \langle R_1^{a_1}, \ldots, R_r^{a_r}, c_1, \ldots, c_s\rangle$,

$$|\langle S\rangle| = S \cup \{c_1^{\mathcal{A}}, \ldots, c_s^{\mathcal{A}}\}.$$

- The relations of $\langle \text{rng}(\alpha)\rangle^{\mathcal{A}}$ are restrictions of the relations of $\mathcal{A}$ to the universe of $\langle \text{rng}(\alpha)\rangle^{\mathcal{A}}$.

Winning a Round of the Game

- We say that **Duplicator wins round r of the game** iff the map

$$\alpha_r(x_j) \mapsto \beta_r(x_j), \quad x_j \in \mathrm{dom}(\alpha_r),$$

determines an isomorphism of the induced substructures,

$$\beta_r \circ \alpha_r^{-1} : \langle \mathrm{rng}(\alpha) \rangle^{\mathcal{A}} \cong \langle \mathrm{rng}(\beta) \rangle^{\mathcal{B}}.$$

- In particular, $\beta_r \circ \alpha_r^{-1}$ must be a one-to-one function.
- So we have

$$\alpha_r(x_i) = \alpha_r(x_j) \quad \text{iff} \quad \beta_r(x_i) = \beta_r(x_j).$$

- Also, all constants and relations of the structures must be preserved.

Example

- Suppose vocabulary τ includes:
  - The binary relation symbol E;
  - The constant symbol c.
- Suppose Duplicator wins round r of the game.
- Then
$$\langle c^{\mathcal{A}}, \alpha_r(x_i)\rangle \in E^{\mathcal{A}} \quad \text{iff} \quad \langle c^{\mathcal{B}}, \beta_r(x_i)\rangle \in E^{\mathcal{B}}.$$
- This can also be written as
$$(\mathcal{A}, \alpha_r) \vDash E(c, x_i) \quad \text{iff} \quad (\mathcal{B}, \beta_r) \vDash E(c, x_i).$$

Winning the Game

- **Duplicator wins the game** iff she wins every single round.
- Duplicator must preserve an isomorphism at all times.
- If a difference between the two structures is ever exposed, then Spoiler wins.
- $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ is a finite game of perfect information (these are terms formally defined in the field of Game Theory).
- By game theoretic results, one of the two players must have a winning strategy.
- We use the notation $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$ to mean that Duplicator has a winning strategy for $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$.
- $(\mathcal{A}, \alpha_0) \sim^k (\mathcal{B}, \beta_0)$ means, for all m, $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$.
- Similarly $(\mathcal{A}, \alpha_0) \sim_m (\mathcal{B}, \beta_0)$ means, for all k, $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$.

Ordering and BIT

- It is important to decide whether to include the numeric predicates $\leq$ and BIT in τ.
- Suppose these relations are available in the language.
- Then they form part of the definition of isomorphism.
- In this case, the game becomes much easier for Spoiler and much harder for Duplicator.
- For this reason, in this part, we assume, unless otherwise noted, that **ordering and BIT are not present**.

Equivalence Relation

Proposition

The relation $\sim_m^k$ is an equivalence relation.

- Consider a pair $(\mathcal{A}, \alpha_0)$.

  We clearly have

  $$I = \alpha_0 \circ \alpha_0^{-1} : \langle \text{rng}(\alpha_0) \rangle^{\mathcal{A}} \cong \langle \text{rng}(\alpha_0) \rangle^{\mathcal{A}}.$$

  Inductively, suppose in move r the Spoiler chooses $\alpha_r(x_i)$.

  Using the identity isomorphism, the Duplicator chooses the same element in the other copy of $\mathcal{A}$.

  Thus, we have $I = \alpha_r \circ \alpha_r^{-1} : \langle \text{rng}(\alpha_r) \rangle^{\mathcal{A}} \cong \langle \text{rng}(\alpha_r) \rangle^{\mathcal{A}}$.

  We conclude that $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{A}, \alpha_0)$.

  This proves that $\sim_m^k$ is reflexive.

Equivalence Relation

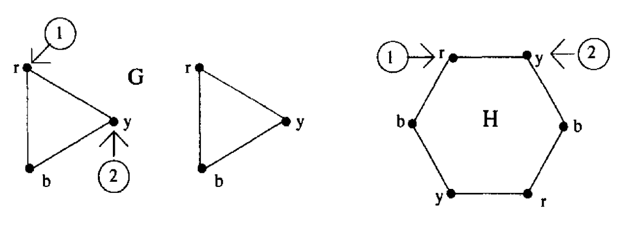
- Suppose, next, that $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$.
  First, $\beta_0 \circ \alpha_0^{-1} : \langle \text{rng}(\alpha_0) \rangle^{\mathcal{A}} \cong \langle \text{rng}(\beta_0) \rangle^{\mathcal{B}}$.
  Hence, $\alpha_0 \circ \beta_0^{-1} = (\beta_0 \circ \alpha_0^{-1})^{-1} : \langle \text{rng}(\beta_0) \rangle^{\mathcal{B}} \cong \langle \text{rng}(\alpha_0) \rangle^{\mathcal{A}}$.
  Next, assume $\alpha_{r-1} \circ \beta_{r-1}^{-1} : \langle \text{rng}(\beta_{r-1}) \rangle^{\mathcal{B}} \cong \langle \text{rng}(\alpha_{r-1}) \rangle^{\mathcal{A}}$.
  Suppose the Spoiler chooses $\alpha_r(x_i)$ or $\beta_r(x_i)$.
  By hypothesis, the Duplicator can respond with either some $\beta_r(x_i)$ or some $\alpha_r(x_i)$, respectively, such that

$$\beta_r \circ \alpha_r^{-1} : \langle \text{rng}(\alpha_r) \rangle^{\mathcal{A}} \cong \langle \text{rng}(\beta_r) \rangle^{\mathcal{B}}.$$

  Thus, $\alpha_r \circ \beta_r^{-1} = (\beta_r \circ \alpha_r^{-1})^{-1} : \langle \text{rng}(\beta_r) \rangle^{\mathcal{B}} \cong \langle \text{rng}(\alpha_r) \rangle^{\mathcal{A}}$.
  We conclude that $(\mathcal{B}, \beta_0) \sim_m^k (\mathcal{A}, \alpha_0)$.
  Transitivity can be proven similarly.

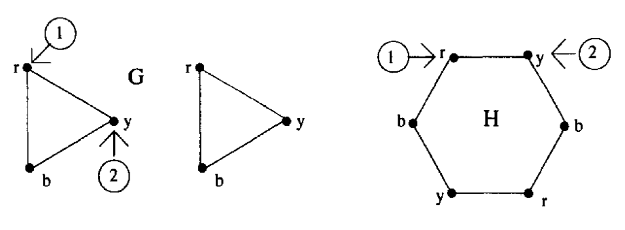
Example

- As an example, consider the two-pebble game on the colored graphs G and H shown in the figure.



- Here the vocabulary $\tau = \langle E^2, R^1, Y^1, B^1 \rangle$ consists of:
  - A binary edge relation;
  - Three unary relations, the "colorings" of the vertices.
- Assume that, initially, all pebbles are off the board, $\alpha_0 = \beta_0 = \varnothing$.

Example (Cont'd)



- Suppose that:
  - Spoiler's first move is to place Pebble 1 on a red vertex in G.
  - Duplicator answers by putting Pebble 1 on a red vertex in H.
  - Spoiler puts Pebble 2 on an adjacent yellow vertex in H.
  - Duplicator has a response because in G, $\alpha_1(x_1)$ also has an adjacent yellow vertex.
  - Spoiler puts Pebble 1 on the blue vertex in H, not adjacent to $\beta_2(x_2)$.
  - Duplicator answers with the blue vertex in G, not adjacent to $\alpha_2(x_2)$.

Example (Cont'd)

Proposition

Let G and H be the graphs shown in the preceding figure. Then, for all m, $G \sim^2_m H$, i.e.,

$$G \sim^2 H.$$

- The previous slide provided the gist of the argument.
- This needs to be generalized to cover all possible cases.
- One starts with $m = 0$ and proceeds by induction on m.

Example (Cont'd)

- Spoiler has an easy win for the game $\mathcal{G}_3^3(G,H)$.
- He can simply choose three points in the same triangle in G on three consecutive moves.
- Duplicator has no response because there is no triangle in H.
- Thus Spoiler wins.
- Observe that Spoiler's winning strategy in this three-pebble game is to "play the sentence" Δ which says that a triangle exists,

$$\Delta \equiv (\exists x_1)(\exists x_2)(\exists x_3)(E(x_1,x_2) \wedge E(x_2,x_3) \wedge E(x_3,x_1))$$

- Note that $G \vDash \Delta$ while $H \vDash \neg\Delta$.

Example (Cont'd)

- Consider, again, the sentence

$$\Delta \equiv (\exists x_1)(\exists x_2)(\exists x_3)(E(x_1,x_2) \wedge E(x_2,x_3) \wedge E(x_3,x_1)).$$

- It has three variables.
- This corresponds the number of pebble pairs in the game.
- Define the **quantifier rank** $\mathrm{qr}(\varphi)$ of a formula φ to be the depth of nesting of quantifiers in φ.
- Note that sentence Δ has quantifier rank 3.
- This corresponds to the number of moves in the game.

Example

- Consider the vocabulary τ_s, with ordering, but not successor.
- We look at a game on the strings $w_1 = 1101$ and $w_2 = 1011$
- Spoiler can win the two-move game on these two strings.
- He can place the x_1 pebble on the second 1 in w_1.
- Duplicator must answer by placing x_1 on some 1 in w_2.
  - Suppose she answers with the first 1.
    Spoiler can reply by placing x_2 on the first 1 in w_1.
    Duplicator has no reply.
  - Suppose Duplicator instead answers with the second or third 1 in w_2.
    Spoiler replies by placing x_2 on the 0 in w_1.
    Duplicator loses because w_2 has no 0 to the right of x_1.
- In this case, Spoiler's winning strategy is to play formula

$$\varphi \equiv (\exists x_1)(S(x_1) \wedge (\exists x_2)(S(x_2) \wedge x_2 < x_1) \wedge (\exists x_2)(\neg S(x_2) \wedge x_1 < x_2)).$$

- We have $w_1 \vDash \varphi$, but $w_2 \vDash \neg\varphi$.

Languages Related to Games

Definition

Define language $\mathcal{L}^k$ to be the restriction of language $\mathcal{L}$ in which only variables $x_1, \ldots, x_k$ occur.
Define language $\mathcal{L}^k_m$ to be the restriction of $\mathcal{L}^k$ to formulas of quantifier rank at most m.
Define $\mathcal{L}_m$ to be the set of formulas of quantifier-rank at most m.

Definition

Let $\mathcal{A}$ and $\mathcal{B}$ be two structures of some vocabulary τ. We say that $\mathcal{A}$ and $\mathcal{B}$ are $\mathcal{L}$ **equivalent** ($\mathcal{A} \equiv_{\mathcal{L}} \mathcal{B}$) iff they agree on all formulas from $\mathcal{L}$,

$$\mathcal{A} \equiv \mathcal{B} \quad \text{iff} \quad \text{for all } \varphi \in \mathcal{L}(\tau),\ \mathcal{A} \vDash \varphi \ \Leftrightarrow\ \mathcal{B} \vDash \varphi,$$

$$\mathcal{A} \equiv^k_m \mathcal{B} \quad \text{iff} \quad \text{for all } \varphi \in \mathcal{L}^k_m(\tau),\ \mathcal{A} \vDash \varphi \ \Leftrightarrow\ \mathcal{B} \vDash \varphi.$$

Inequivalent Formulas of a Fixed Rank

Lemma

For all natural numbers k and m, there are only finitely many inequivalent first-order formulas in $\mathcal{L}_m^k(\tau)$.

- It suffices to show that, for every k and m, there exist only finitely many inequivalent formulas of rank $\leq m$ that have $x_1, \ldots, x_k$ free.

  We use induction on m, keeping k variable.

  Note that there are finitely many atoms in $x_1, \ldots, x_k$.

  Consider a formula of quantifier rank 0.

  It has an equivalent in disjunctive normal form.

  Clearly, there are finitely many such forms using finitely may atoms.

Inequivalent Formulas of a Fixed Rank (Cont'd)

- Consider, next, a formula with quantifier rank $\leq n+1$, with $x_1, \ldots, x_k$ free.

  It has an equivalent disjunctive normal form.

  Its ingredients are:

  - Rank $\leq n$ formulas;
  - Formulas $\exists x_{k+1}\varphi$, where φ has rank $\leq n$ and $x_1, \ldots, x_k$ free.

  By the induction hypothesis, there are only finitely many inequivalent formulas of each of those types.

Fundamental Theorem of Ehrenfeucht-Fraïssé Games

Theorem

Let $\mathcal{A}$ and $\mathcal{B}$ be structures of the same finite, relational vocabulary and let α_0, β_0 be a k-configuration of $\mathcal{A}$, $\mathcal{B}$. Then the following are equivalent:

1. $(\mathcal{A}, \alpha_0) \sim_m^k (\mathcal{B}, \beta_0)$;
2. $(\mathcal{A}, \alpha_0) \equiv_m^k (\mathcal{B}, \beta_0)$.

- We prove the equivalence of 1 and 2 by induction on m.

  For $m = 0$, Duplicator wins the zero move game

  iff $\beta_0 \circ \alpha_0^{-1}$ is an isomorphism of the induced substructures

  iff, for every quantifier free $\gamma \in \mathcal{L}(\tau)$, $(\mathcal{A}, \alpha_0) \vDash \gamma$ iff $(\mathcal{B}, \beta_0) \vDash \gamma$.

  Note that γ may have as free variables only those variables that occur in $\text{dom}(\alpha_0) = \text{dom}(\beta_0)$.

  Thus, 1 and 2 are equivalent for $m = 0$.

Fundamental Theorem (Induction Step)

- Assume the theorem is true for m.
  Suppose that $\mathcal{A}$ and $\mathcal{B}$ disagree on the formula $\varphi \in \mathcal{L}^k_{m+1}$.
  Note that, if φ is $\alpha \wedge \beta$, then $\mathcal{A}$ and $\mathcal{B}$ disagree on one of α and β.
  Similarly, if φ is $\neg\alpha$, then they disagree on α.
  So we may assume that φ is $(\exists x_i)\psi$.
  Suppose that $(\mathcal{A}, \alpha_0) \vDash \varphi$ and $(\mathcal{B}, \beta_0) \vDash \neg\varphi$.
  Spoiler's first move in $\mathcal{G}^k_{m+1}(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ is to place pebble i on a witness for ψ in $\mathcal{A}$.
  Wherever Duplicator responds, it will not be a witness for ψ, because there is none in $\mathcal{B}$.
  Thus, after the first move, $(\mathcal{A}, \alpha_1)$ and $(\mathcal{B}, \beta_1)$ disagree on the quantifier depth m formula ψ.
  By the inductive hypothesis, Spoiler has a winning strategy for the remaining m-move game.
  Thus, we have shown that 1 implies 2.

Fundamental Theorem (Induction Step Converse)

- Conversely, suppose that $(\mathcal{A}, \alpha_0) \equiv^k_{m+1} (\mathcal{B}, \beta_0)$.

  Let Spoiler make his first move in the game $\mathcal{G}^k_{m+1}(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$.

  Suppose he places pebble i on an element of $\mathcal{A}$, thus defining α_1.

  By the preceding lemma, there are only finitely many inequivalent formulas in $\mathcal{L}^k_m$.

  Let Φ be the conjunction of all these formulas satisfied by $(\mathcal{A}, \alpha_1)$.

  Thus, we know that $(\mathcal{A}, \alpha_0) \vDash (\exists x_i)\Phi$.

  By hypothesis, $(\mathcal{B}, \beta_0) \vDash (\exists x_i)\Phi$.

  Duplicator places the other pebble i on a witness in $\mathcal{B}$ of Φ.

  Thus, (A, α_1) and $(\mathcal{B}, \beta_1)$ both satisfy Φ, a complete description of every formula from $\mathcal{L}^k_m$ that $(\mathcal{A}, \alpha_1)$ satisfies.

  Therefore, $(\mathcal{A}, \alpha_1) \equiv^k_m (\mathcal{B}, \beta_1)$.

  It follows, by induction, that Duplicator has a winning strategy for the remaining m moves of the game.

Observation on Spoiler's Strategies

- It turns out that in any game $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$, we never have to consider a move in which Spoiler pebbles an element that is already pebbled by another pebble or constant.

Lemma

Consider a game $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$. If Spoiler has a winning strategy, then he still has a winning strategy if he is never allowed to place a pebble on a constant or an element that already has another pebble sitting on it.

- We refer to the version of $\mathcal{G}_m^k(\mathcal{A}, \alpha_0, \mathcal{B}, \beta_0)$ in which Spoiler is never allowed to place a pebble on a constant or an element that already has another pebble sitting on it as the *restricted version*.

  We use contraposition.

  Assume Duplicator has a winning strategy in the restricted version.

  We show that she also has a winning strategy in the ordinary version.

Observation on Spoiler's Strategies (Cont'd)

- If Spoiler does not repeat, then Duplicator uses her winning strategy in $\mathcal{G}_m^k$, ignoring preceding repetitions.

  Consider a move $r \leq m$ in the unrestricted game, in which Spoiler places a pebble j on a constant or an element that already has another pebble i sitting on it.

  Duplicator's strategy involves placing the second pebble j on the element of the other structure in which the other pebble i is sitting.

  By induction, Duplicator was winning the game up to this round.

  By the preceding theorem, the two induced substructures were isomorphic.

  Duplicator's last play ensures that the new induced substructures are identical to the previous ones.

  Therefore, the new induced substructures are still isomorphic.

  Therefore, Duplicator also wins the current round of the game.

Clique

Proposition

Let CLIQUE(k) be the set of undirected graphs that contain a clique, i.e., a complete subgraph, of size k. In the language without ordering, CLIQUE(k) is expressible with k variables but not $k-1$ variables,

$$\text{CLIQUE}(k) \in \mathcal{L}^k(\tau_g)(\text{wo}\leq) - \mathcal{L}^{k-1}(\tau_g)(\text{wo}\leq).$$

- We may write CLIQUE(k) in $\mathcal{L}^k$ using the formula

$$(\exists x_1 x_2 \ldots x_k)(\text{distinct}(x_1, \ldots, x_k) \\ \wedge E(x_1, x_2) \wedge \cdots \wedge E(x_1, x_k) \wedge \cdots \wedge E(x_{k-1}, x_k)).$$

We must now show that k variables are necessary.

We show $K_k \sim^{k-1} K_{k-1}$, where K_r is the complete graph on r vertices.

Clique (Cont'd)

- Duplicator has a simple winning strategy for $\mathcal{G}_{k-1}(K_k, K_{k-1})$.

  Suppose Spoiler places the a pebble on an unpebbled vertex in one of the two graphs.

  Duplicator places the corresponding pebble on any unpebbled vertex in the other graph.

  By hypothesis, there are only $k-1$ pebble pairs.

  So such an unpebbled vertex is always available.

  Now edges exist between all points in each graph.

  So this is clearly a winning strategy

  Thus, any one-to-one correspondence is an isomorphism.

Necessity of Ordering to Express Parity

Proposition

In the absence of ordering, the boolean query on graphs that is true iff there are an odd number of vertices requires $n+1$ variables, for graphs with n or more vertices. The same is true for the query that there are an odd number of edges.

- Let G_n be the graph on n vertices that has a loop at each vertex but no other edges.

  We claim that $G_n \sim^n G_{n+1}$.

  Suppose Spoiler places a pebble on a vertex not already pebbled.

  Duplicator matches the move by pebbling a vertex not already pebbled in the other graph.

Necessity of Ordering to Express Parity (Cont'd)

- Each graph has at least n vertices.

  Moreover, there are no edges between different vertices.

  So this is a winning strategy for Duplicator.

  It follows that $G_n \equiv^n G_{n+1}$.

  So the parity of the number of vertices or the number of edges in not expressible in $\mathcal{L}^n$.

Paths

Proposition

Let the formula $\mathrm{PATH}_k(x,y) \in \mathcal{L}(\tau_g)$ mean that there is a path of length at most 2^k from x to y. With or without ordering, quantifier rank k is necessary and sufficient to express PATH_k. Furthermore, only three variables are necessary to express PATH_k. In symbols,

$$\mathrm{PATH}_k \in \mathcal{L}^3_k(\tau_g)(\mathrm{wo}\leq) - \mathcal{L}_{k-1}(\tau_g).$$

- For the upper bound, we express PATH_k inductively as follows.

$$\begin{array}{rcl} \mathrm{PATH}_0(x,y) & \equiv & x = y \vee E(x,y); \\ \mathrm{PATH}_{k+1}(x,y) & \equiv & (\exists z)(\mathrm{PATH}_k(x,z) \wedge \mathrm{PATH}_k(z,y)). \end{array}$$

So PATH_k is expressible using three variables and quantifier rank k.

Paths (Cont'd)

- We claimed only three variables are needed.

  This is because the right hand side of the inductive definition of $\text{PATH}_{k+1}(x,y)$ may be written in a way that reuses variables.

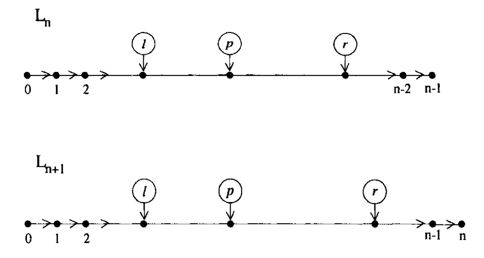
$$\begin{array}{rcl} \text{PATH}_k(x,z) & \equiv & (\exists y)(\text{PATH}_{k-1}(x,y) \wedge \text{PATH}_{k-1}(y,z)); \\ \text{PATH}_k(z,y) & \equiv & (\exists x)(\text{PATH}_{k-1}(z,x) \wedge \text{PATH}_{k-1}(x,y)). \end{array}$$

Paths (Cont'd)

- We turn to proving the lower bound.
  Let $L_n \in \mathrm{STRUC}[\tau_g]$ be a directed line segment of length $n-1$.
  So $\|L_n\| = n$.
  Let the ordering on L_n be from left to right.

  Suppose $n = 2^{k+1} + 1$.
  Then

  $L_n \vDash \mathrm{PATH}_{k+1}(0, \max)$;
  $L_{n+1} \vDash \neg\mathrm{PATH}_{k+1}(0, \max)$.



  We show next that $L_n \sim_k L_{n+1}$ and the lower bound follows.

Paths (Cont'd)

- The idea behind Duplicator's winning strategy is that in quantifier rank s, or, equivalently, with s moves remaining in the game, distances greater than 2^s are indistinguishable from infinite distances.

  With this idea in mind, let us define the notion $i =_d j$ to mean that i and j are equal or are both greater than d.

  Duplicator's winning strategy in $\mathcal{G}_k(L_n, L_{n+1})$ is to maintain the following invariant.

  After the move m of $\mathcal{G}_k(L_n, L_{n+1})$, and for all $p, q \in \text{dom}(\alpha_m)$,

$$\begin{array}{rcl} \text{DIST}(\alpha_m(p), \alpha_m(q)) & =_{2^{k-m}} & \text{DIST}(\beta_m(p), \beta_m(q)); \\ \alpha_m(p) \le \alpha_m(q) & \Leftrightarrow & \beta_m(p) \le \beta_m(q). \end{array}$$

  Note that this implies that Duplicator wins the game, because a map that preserves distances of length at most one is an isomorphism.

Paths (Cont'd)

- The equation holds after move 0, since

$$\text{DIST}(0^{L_n}, \max^{L_n}) = 2^{k+1} =_{2^k} 2^{k+1} + 1 = \text{DIST}(0^{L_{n+1}}, \max^{L_{n+1}}).$$

  Assume inductively that the invariant holds just after move m.

  Let Spoiler begin move $m+1$ by placing pebble p on some vertex.

  Let ℓ and r be the closest pebbles to the left and right of p.

  The inductive assumption tells us that

$$\text{DIST}(\alpha_m(\ell), \alpha_m(r)) =_{2^{k-m}} \text{DIST}(\beta_m(\ell), \beta_m(r)).$$

  Assume without loss of generality that ℓ is the closer of the two pebbles to p or that they are equidistant.

Paths (Cont'd)

- Duplicator's response is to place the other pebble p on the point to the right of the other ℓ so as to have

$$\mathrm{DIST}(\alpha_{m+1}(\ell), \alpha_{m+1}(p)) = \mathrm{DIST}(\beta_{m+1}(\ell), \beta_{m+1}(p)).$$

It follows, that

$$\mathrm{DIST}(\alpha_{m+1}(p), \alpha_{m+1}(r)) =_{2^{k-(m+1)}} \mathrm{DIST}(\beta_{m+1}(p), \beta_{m+1}(r)).$$

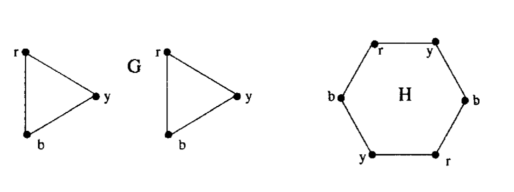
So the invariant holds after move $m+1$.

Thus, Duplicator wins the game.

We have proved that PATH_k is not expressible with quantifier rank less than k, even for ordered structures.

Necessity of the Three Variables

- Three variables used to express paths are necessary.
- In a previous proposition, we saw a connected graph H of diameter three and a disconnected graph G such that $G \sim^2 H$.



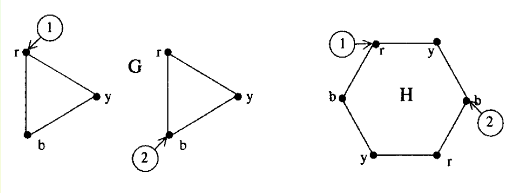
- Thus, by a previous theorem, CONNECTED is not expressible using 2 variables, no matter what the quantifier rank.

Necessity of the Three Variables (Cont'd)

- Suppose, towards a contradiction, that PATH_k is expressible in $\mathcal{L}^2$, for some $k \geq 2$.

  Then G and H differ on the $\mathcal{L}^2$ formula $(\forall x_1 x_2)\text{PATH}_k(x_1, x_2)$.

  In fact, let (α_0, β_0) be the 2-configuration of graphs G, H shown.



  Observe that $(G, \alpha_0) \sim^2 (H, \beta_0)$.

  However, the two structures disagree on the formula $\text{PATH}_1(x_1, x_2)$.

  Hence, for $k \geq 1$, PATH_k is not expressible in $\mathcal{L}^2$.

Subsection 2

Methodology for First-Order Expressibility

Methodology Theorem

Theorem (Methodology Theorem)

Let $\mathcal{C}$ be any class of finite or infinite structures of some finite, relational vocabulary. Let $S \subseteq \mathcal{C}$ be a boolean query on $\mathcal{C}$. To prove that S is not first-order describable on $\mathcal{C}$ it is necessary and sufficient to show that, for all $r \in \mathbb{N}$, there exist structures $\mathcal{A}_r, \mathcal{B}_r \in \mathcal{C}$ such that:

1. $\mathcal{A}_r \in S$ and $\mathcal{B}_r \notin S$;
2. $\mathcal{A}_r \sim_r \mathcal{B}_r$.

- Suppose the given condition holds.

  Statements 1 and 2 imply that $\mathcal{A}_r$ and $\mathcal{B}_r$ agree on all formulas in $\mathcal{L}_r$, but disagree on S.

  Thus, S is not expressible in $\mathcal{L}_r$ for any r.

Methodology Theorem (Cont'd)

- We say that $\varphi \in \mathcal{L}_r$ is a **complete quantifier rank** r **sentence** if, for every other quantifier rank r sentence ψ of the same vocabulary,

$$\varphi \vdash \psi \quad \text{or} \quad \varphi \vdash \neg\psi.$$

  Let $\varphi_1, \ldots, \varphi_B$ be a list of all inequivalent, complete quantifier rank r sentences.

  For every quantifier rank r sentence ψ, each φ_i must assert either ψ or $\neg\psi$.

  Observe that each structure from $\mathcal{C}$ satisfies a unique φ_i.

Methodology Theorem (Cont'd)

- Suppose there exist $\mathcal{A}_r \in S$ and $\mathcal{B}_r \in \mathcal{C} - S$ satisfying the same φ_i.

  Then $\mathcal{A}_r$ and $\mathcal{B}_r$ satisfy the above conditions.

  Suppose there is no such pair.

  Then the φ_i's are partitioned by S.

  In this case, let

  $$Y = \{i : (\exists \mathcal{A} \in S)(\mathcal{A} \vDash \varphi_i)\}.$$

  Define

  $$\varphi \equiv \bigvee_{i \in Y} \varphi_i.$$

  Then φ is a first-order formula of quantifier rank r that expresses S.

Gaifman Graph

- Let $\tau = \langle R_1^{a_1}, \ldots, R_r^{a_r}, c_1, \ldots, c_s \rangle$ be a vocabulary.
- Let $\mathcal{A}$ be any τ-structure.
- Define the **Gaifman graph** $G_{\mathcal{A}} = (|\mathcal{A}|, E_{\mathcal{A}})$ by

$$E_{\mathcal{A}} = \{(a,b) : (\exists i)(\exists \langle d_1, \ldots, d_{a_i} \rangle \in R_i^{\mathcal{A}})(a, b \in \{d_1, \ldots, d_{a_i}\})\}.$$

- There is an edge between a and b in the Gaifman graph iff a and b occur in the same tuple of some relation of $\mathcal{A}$.
- Example: Let $\mathcal{A} \in \text{STRUC}[\tau_g]$ be a graph.
  Then $G_{\mathcal{A}} = \mathcal{A}$.

Universe of a Neighborhood of an Element

- Let $(\mathcal{A}, \alpha_r)$ be the configuration of structure $\mathcal{A}$ after move r of a game.
- Define the **universe of the neighborhood of element** a **at distance** d to be the set of elements of distance at most d from a in the Gaifman graph,

$$|N(a,d)| = \{b \in |\mathcal{A}| : \mathrm{DIST}(a,b) \le d\},$$

  where $\mathrm{DIST}(a,b)$ stands for $\mathrm{DIST}_{G_{(\mathcal{A},\alpha_r)}}(a,b)$.
- $N(a,d)$ is almost an induced substructure of $(\mathcal{A}, \alpha_r)$.
  - It does inherit the relations from $\mathcal{A}$.
  - However, it contains only those constants and pebbled points that are within distance d of a.

d-Type

- Define the d-**type of** a to be the isomorphism type of $N(a, d)$.
- Isomorphisms must send:
  - Each constant $c_j^{\mathcal{A}}$ to $c_j^{\mathcal{B}}$;
  - Each pebbled point $\alpha_r(x_i)$ to $\beta_r(x_i)$.
- Neighborhood $N(a, d)$ and, thus, the d-type of a depend on the current configuration $(\mathcal{A}, \alpha_r)$.
- If the configuration is not clear from the context, then we say the d-**type of** a **with respect to configuration** $(\mathcal{A}, \alpha_r)$.

Hanf's Theorem

Theorem (Hanf's Theorem)

Let $\mathcal{A},\mathcal{B} \in \text{STRUC}[\tau]$ and let $r \in \mathbb{N}$. Suppose that, for each possible 2^r-type t, $\mathcal{A}$ and $\mathcal{B}$ have exactly the same number of elements of type t. Then

$$\mathcal{A} \equiv_r \mathcal{B}.$$

- We must show that Duplicator wins the game $\mathcal{G}_r(\mathcal{A},\mathcal{B})$.

  Duplicator's winning strategy is to maintain the invariant that, after move m, $0 \le m \le r$,

  $$(\mathcal{A},\alpha_m),(\mathcal{B},\beta_m) \text{ have same number of each } 2^{r-m}\text{-type.}$$

  In $\mathcal{G}_r(\mathcal{A},\mathcal{B})$ there is no bound on the number of pebbles.

  So we may assume that Spoiler uses a new pebble at each step.

  Thus, Duplicator wins iff she wins at the last round.

Hanf's Theorem (Cont'd)

- Suppose Duplicator preserves the invariant.

  Then after the last move, the neighborhoods of distance one around each constant or pebbled point are isomorphic to the corresponding neighborhoods in the other structure.

  It follows that Duplicator wins the game.

  The invariant holds for $m = 0$ by assumption.

  Inductively, assume that it holds after move m.

  On move $m + 1$, let Spoiler choose some vertex v.

  Duplicator responds with any vertex v' of the same 2^{r-m}-type as v.

  We have to show that the invariant still holds.

  The inductive assumption immediately implies that, $(\mathcal{A}, \alpha_m)$, $(\mathcal{B}, \beta_m)$ have same number of each $2^{r-(m+1)}$-type.

  Furthermore, the neighborhood $N(a, 2^{r-(m+1)})$ of $(\mathcal{A}, \alpha_m)$ is different from the same neighborhood of $(\mathcal{A}, \alpha_{m+1})$ iff $\mathrm{DIST}(a, v) \le 2^{r-(m+1)}$.

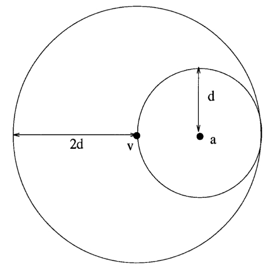
Hanf's Theorem (Cont'd)

- Consider the isomorphism $f : N(v, 2^{r-m}) \to N(v', 2^{r-m})$.

  It maps every vertex a in $N(v, 2^{r-(m+1)})$ to a corresponding $a' \in N(v', 2^{r-(m+1)})$.

  The key idea is that f maps $N(a, 2^{r-(m+1)})$ isomorphically onto $N(a', 2^{r-(m+1)})$ because these smaller neighborhoods lie inside

  $$\text{dom}(f) = N(v, 2^{r-m}).$$



  Thus, there is a one-to-one correspondence between the isomorphism types of these neighborhoods close to v and v'.

  So the one-to-one correspondence between the other neighborhoods is undisturbed.

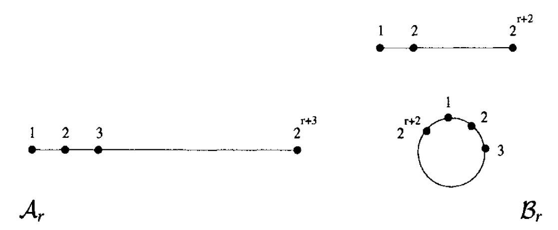
  Thus, Duplicator's strategy preserves the invariance.

Acyclicity

Proposition

Acyclicity is not first-order expressible.

- Let $\mathcal{A}_r$ be a line segment on 2^{r+3} vertices.
  Let $\mathcal{B}_r$ be the union of:
  - A line segment on 2^{r+2} vertices;
  - A cycle on 2^{r+2} vertices.



Observe that $\mathcal{A}_r$ and $\mathcal{B}_r$ both have the same number of each 2^r-type.
By the theorem, $\mathcal{A}_r \equiv_r \mathcal{B}_r$. So Acyclicity is not first-order expressible.

Subsection 3

First-Order Properties Are Local

The Degree of a Structure

- The **degree** of a graph is the maximum number of edges adjacent to any vertex.
- The **degree** of a structure $\mathcal{A}$ is the degree of its Gaifman graph.
- We prove a strengthening of Hanf's Theorem for graphs of bounded degree.
- It relaxes the requirement that the number of instances of a given r-type in the two structures be equal.
- It requires instead that both numbers are sufficiently large.

(n, σ)-Equivalence

- Let $\mathcal{A}$ and $\mathcal{B}$ be structures and let n, s be integers.
- We say that $\mathcal{A}$ and $\mathcal{B}$ are (n, s)-**equivalent** if, for each n-type σ, at least one of the following holds:
  - $\mathcal{A}$ and $\mathcal{B}$ have the same number of neighborhoods of type σ;
  - Both $\mathcal{A}$ and $\mathcal{B}$ have more than s neighborhoods of type σ.

Bounded-Degree Hanf Theorem

Theorem (Bounded-Degree Hanf Theorem)

Let r and d be fixed. There is an integer s, such that, for all structures $\mathcal{A}$ and $\mathcal{B}$ of degree at most d,

$$\text{if } \mathcal{A} \text{ and } \mathcal{B} \text{ are } (2^r, s)\text{-equivalent, then } \mathcal{A} \equiv_r \mathcal{B}.$$

- We must show that Duplicator wins the game $\mathcal{G}_r(\mathcal{A}, \mathcal{B})$.

  Set $s = rd^{2^r} + 1$.

  Duplicator's winning strategy is to maintain the invariant that, after move m, $0 \le m \le r$,

  > $(\mathcal{A}, \alpha_m)$, $(\mathcal{B}, \beta_m)$ have the same number of each 2^{r-m}-type, or both have over $(r-m)d^{2^r} + 1$ elements of this type.

  The invariant holds for $m = 0$ by assumption.

Bounded-Degree Hanf Theorem (Cont'd)

- Inductively, assume that it holds after move m.

  On move $m + 1$, let Spoiler choose some vertex v.

  Duplicator responds with any vertex v' of the same 2^{r-m}-type as v.

  We have to show that the invariant still holds.

  The inductive assumption immediately implies that

  > $(\mathcal{A}, \alpha_m)$, $(\mathcal{B}, \beta_m)$ have the same number of each $2^{r-(m+1)}$-type, or both have over $(r - m)d^{2^r} + 1$ elements of this type.

Bounded-Degree Hanf Theorem (Cont'd)

- Just as in the proof of a previous theorem, the only neighborhoods that change are those within distance $2^{r-(m+1)}$ of v.

  Furthermore, the same number of neighborhoods change in the same way in $\mathcal{A}$ as in $\mathcal{B}$.

  The only harm that can be done to the invariant is that the number of some types can be reduced by the same amount in $\mathcal{A}$ and in $\mathcal{B}$.

  The number of vertices within distance $\rho = 2^{r-(m+1)}$ of v is

$$\le \frac{d^{\rho+1}}{d-1} < d^{2^r}.$$

  Thus, we have that the invariant holds for $m+1$.

Linear Recognition of Bounded-Degree Structures

- The definition of linear time in the following is linear time on a unit-cost RAM with $O(\log n)$ bit word size.

Theorem

Let $\varphi \in \mathrm{FO}$. Then over bounded degree structures, φ is recognizable in linear time.

- For simplicity, assume that the structures in question are bounded degree graphs.

  Let them be represented via adjacency lists.

  Let r be the quantifier rank of φ.

  Let d be the degree of the graphs in question.

  The number of possible 2^r-types in degree d graphs is large but bounded.

Linear Recognition of Bounded-Degree Structures (Cont'd)

- The linear time algorithm is to:
  - Determine the 2^r type of each vertex;
  - Count, up to s, how many of each type occurs.

  This information is what we can call the $(2^r, s)$-**description** of G.

  By a previous theorem, the $(2^r, s)$-description of G determines whether G satisfies φ.

  We could in principle build, beforehand, a table that lists, for each of the finitely many possible $(2^r, s)$-descriptions, whether or not a graph with this description satisfies φ.

  From G's description, we can use the table to check in constant additional time whether G satisfies φ.

Subsection 4

Bounded Variable Languages

The k-Variable Property

- A theory Σ satisfies the k-**variable property** if every first-order formula is equivalent with respect to Σ to a first-order formula that has only k bound variables.
- Gabbay has shown that the set of models of Σ has the k-variable property for some k iff there exists a finite basis for the set of all temporal-logic connectives over these models.
- We will show, using Ehrenfeucht-Fraïssé games, that the set of linearly ordered structures has the 3-variable property.
- The set of bounded degree trees also has the k-variable property, for appropriate k.
- In this section we consider all structures, not just finite structures.

Models and Formulas

Lemma

Let $\Sigma \subseteq \mathcal{L}$ be a first-order theory. Let $\mathcal{L}'$ and $\mathcal{L}''$ be subsets of $\mathcal{L}$, such that $\mathcal{L}'$ is closed under the boolean connectives. Let $k \in \mathbb{N}$. The following conditions are equivalent:

1. For all models $\mathcal{A}$ and $\mathcal{B}$ of Σ and all k-configurations α, β of $\mathcal{A}, \mathcal{B}$,
$$(\mathcal{A}, \alpha) \equiv_{\mathcal{L}'} (\mathcal{B}, \beta) \quad \text{implies} \quad (\mathcal{A}, \alpha) \equiv_{\mathcal{L}''} (\mathcal{B}, \beta);$$
2. For all $\varphi \in \mathcal{L}''$, with free variables among $x_1, \ldots, x_k$, there exists $\psi \in \mathcal{L}'$, such that
$$\Sigma \vDash \varphi \leftrightarrow \psi.$$

(2 → 1) Suppose every formula in $\mathcal{L}''$ is equivalent to a formula in $\mathcal{L}'$.
Let $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ be $\mathcal{L}'$-equivalent.
Then $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ are $\mathcal{L}''$-equivalent.

Models and Formulas (Cont'd)

(1 → 2) Suppose $\Sigma \cup \{\varphi\}$ is inconsistent.

Then we may take $\psi \equiv$ **false**.

Otherwise, let T be the set of all complete $\mathcal{L}'$-types over the variables $x_1, \ldots, x_k$ that is consistent with $\Sigma \cup \{\varphi\}$.

Let $\Gamma \in T$ be such a type.

Observe that $\Sigma \cup \Gamma \models \varphi$.

Otherwise, we could construct models $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ of $\Sigma \cup \Gamma$ that disagree on φ. This is impossible by 1.

By the Compactness Theorem, there exists a formula ψ_Γ, such that

$$\Sigma \models \psi_\Gamma \to \varphi.$$

Models and Formulas (Cont'd)

- There exists a formula ψ_Γ, such that

$$\Sigma \vDash \psi_\Gamma \to \varphi.$$

Define the set of formulas

$$D = \{\neg\psi_{\Gamma_i} : \Gamma_i \in T\}.$$

Then $\Sigma \cup D \cup \{\varphi\}$ is inconsistent.

By Compactness, there must be some finite $F \subseteq T$, such that

$$\Sigma \vDash \bigwedge_{\Gamma_i \in F} \neg\psi_{\Gamma_i} \to \neg\varphi.$$

We can take

$$\psi = \bigvee_{\Gamma_i \in F} \psi_{\Gamma_i}.$$

Proving the k-Variable Property

- Let $\Sigma \subseteq \mathcal{L}$ be a theory.
- Let $\mathcal{L}' = \mathcal{L}^k$, and let $\mathcal{L}'' = \mathcal{L}$.
- In this case, the lemma implies that Condition 1, which may be proved using Ehrenfeucht-Fraïssé games, is sufficient to show that every formula in $\mathcal{L}$ that has at most k free variables is equivalent to a formula in $\mathcal{L}^k$.
- To prove the k-variable property, we must also show that any formula with more than k free variables is equivalent to a formula with at most k bound variables.

Strategy

- Let $\mathcal{L}$ be a first-order relational language with no relation symbols of arity greater than k.
- Suppose that $\Sigma \subseteq \mathcal{L}$ is a theory.
- Suppose, also, that $R_1, R_2, \ldots$ are an infinite set of monadic relation symbols from $\mathcal{L}$, that do not occur in Σ.
- Even though we have infinitely many R_i's, we consider only structures in which only finitely many relations are non-empty.
- Suppose that, for every pair of such structures $\mathcal{A}, \mathcal{B}$ satisfying Σ and every pair of k-configurations α, β, we have

$$(\mathcal{A}, \alpha) \equiv^k (\mathcal{B}, \beta) \ \Rightarrow \ (\mathcal{A}, \alpha) \equiv (\mathcal{B}, \beta).$$

- Then Σ has the k-variable property.
  - This follows essentially from the lemma.
  - Additional free variables can be replaced by new monadic relation symbols.

Linear Ordered Structures

Theorem

The set of linear ordered structures satisfies the 3-variable property. The structures may also include any number of monadic relation symbols.

- By the preceding remarks, it suffices to show that, for any pair of linear orders $\mathcal{A}, \mathcal{B}$ and any pair of 3-configurations α, β,

  $$(\mathcal{A}, \alpha) \equiv^3 (\mathcal{B}, \beta) \quad \text{implies} \quad (\mathcal{A}, \alpha) \equiv (\mathcal{B}, \beta).$$

  We prove the slightly stronger result that for all m,

  $$(\mathcal{A}, \alpha) \sim_m^3 (\mathcal{B}, \beta) \quad \text{implies} \quad (\mathcal{A}, \alpha) \sim_m (\mathcal{B}, \beta).$$

  We prove this by induction on m.

  The base case, $m = 0$, is clear because extra pebbles cannot help Spoiler in the zero move game.

  Assume that the implication holds for m.

Linear Ordered Structures (First Case)

- Suppose that $(\mathcal{A}, \alpha) \sim^3_{m+1} (\mathcal{B}, \beta)$.

  We describe a winning strategy for Duplicator in $\mathcal{G}_{m+1}(\mathcal{A}, \alpha, \mathcal{B}, \beta)$.

  Suppose that in the initial configuration, $|\mathrm{dom}(\alpha)| < 3$.

  That is, suppose that fewer than three pebbles are on the board.

  In this case, wherever Spoiler plays, Duplicator can answer using her winning strategy for the game $\mathcal{G}^3_{m+1}(\mathcal{A}, \alpha, \mathcal{B}, \beta)$.

  Let α_1, β_1 be the resulting configuration.

  We know that $(\mathcal{A}, \alpha_1) \sim^3_m (\mathcal{B}, \beta_1)$.

  Thus, by the inductive assumption, $(\mathcal{A}, \alpha_1) \sim_m (\mathcal{B}, \beta_1)$.

  So Duplicator wins the remaining m moves of the game.

Linear Ordered Structures (Second Case)

- Suppose $|\alpha| = |\beta| = 3$.

  Renumber the variables, if necessary, so that $(\mathcal{A}, \alpha)$ and $(\mathcal{B}, \beta)$ both satisfy $x_1 < x_2 < x_3$.

  Let α_ℓ, β_ℓ and α_r, β_r be the restrictions of α, β to the domains $\{x_1, x_2\}$ and $\{x_2, x_3\}$, respectively.

  Since $(\mathcal{A}, \alpha) \sim^3_{m+1} (\mathcal{B}, \beta)$, Duplicator wins the three-variable, $(m+1)$-move games on these reduced configurations,

  $$(\mathcal{A}, \alpha_\ell) \sim^3_{m+1} (\mathcal{B}, \beta_\ell) \quad \text{and} \quad (\mathcal{A}, \alpha_r) \sim^3_{m+1} (\mathcal{B}, \beta_r).$$

  Since the domains of these configurations have size less than three, we know by the previous case that,

  $$(\mathcal{A}, \alpha_\ell) \sim_{m+1} (\mathcal{B}, \beta_\ell) \quad \text{and} \quad (\mathcal{A}, a_r) \sim_{m+1} (\mathcal{B}, \beta_r).$$

Linear Ordered Structures (Cont'd)

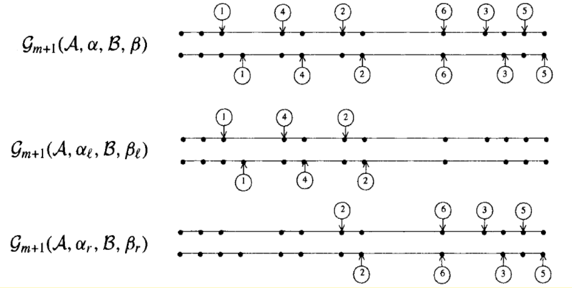
- We now combine Duplicator's winning strategies for the games

$$\mathcal{G}_{m+1}(\mathcal{A}, \alpha_\ell, \mathcal{B}, \beta_\ell) \quad \text{and} \quad \mathcal{G}_{m+1}(\mathcal{A}, a_r, \mathcal{B}, \beta_r)$$

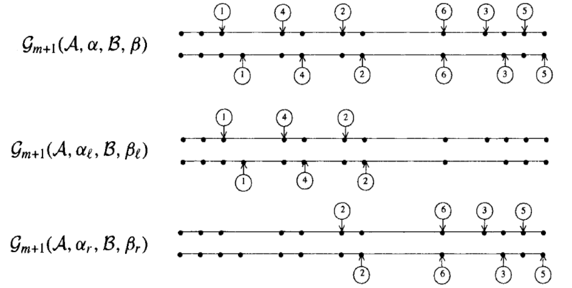
  to devise a winning strategy for the game $\mathcal{G}_{m+1}(\mathcal{A}, \alpha, \mathcal{B}, \beta)$.

  We are playing a game with an unlimited number of pebbles.

  So Spoiler need never reuse a pebble.



Linear Ordered Structures (Cont'd)



- Duplicator's strategy is as follows.
  - If Spoiler places a pebble to the left of pebble two, then Duplicator answers according to her winning strategy in $\mathcal{G}_{m+1}(\mathcal{A}, \alpha_\ell, \mathcal{B}, \beta_\ell)$.
  - If Spoiler places a pebble to the right of pebble two, then Duplicator answers according to her winning strategy in $\mathcal{G}_{m+1}(\mathcal{A}, \alpha_r, \mathcal{B}, \beta_r)$.

After the $m+1$ moves, Duplicator has won both of the subgames.

Linear Ordered Structures (Cont'd)

- Consider the maps:
  - From the chosen points of $\mathcal{A}$ to the chosen points of $\mathcal{B}$ in the left subgame;
  - From the chosen points of $\mathcal{A}$ to the chosen points of $\mathcal{B}$ in the right subgame.

  These maps are both isomorphisms.

  Furthermore:
  - All the chosen points in the left subgame are less than x_2;
  - All the chosen points in the right subgame are greater than x_2.

  Thus, the map from all the pebbled points in $\mathcal{A}$ to the pebbled points in $\mathcal{B}$ is an isomorphism.

  So Duplicator wins $\mathcal{G}_{m+1}(\mathcal{A}, \alpha, \mathcal{B}, \beta)$.

Subsection 5

Zero-One Laws

Extension Axioms

- Extension axioms can be written for any finite relational vocabulary.
- We first write them for undirected graphs.
- Consider the following sentence γ_k, whose meaning is that "there are at least $k-1$ distinct vertices and any $k-1$ tuple of distinct vertices may be extended to a k tuple in any conceivable way".

$$\begin{aligned}
&\gamma_k \equiv (\exists x_1 \ldots x_{k-1}.\mathrm{distinct}(x_1,\ldots,x_{k-1})) \wedge \\
&(\forall x_1 \ldots x_{k-1}.\mathrm{distinct}(x_1,\ldots,x_{k-1})) \\
&((\exists x_k.\mathrm{distinct}(x_1,\ldots,x_k))(E(x_1,x_k) \wedge E(x_2,x_k) \wedge \cdots \wedge E(x_{k-1},x_k)) \\
&\wedge (\exists x_k.\mathrm{distinct}(x_1,\ldots,x_k))(E(x_1,x_k) \wedge E(x_2,x_k) \wedge \cdots \wedge \neg E(x_{k-1},x_k)) \\
&\wedge \cdots \\
&\wedge (\exists x_k.\mathrm{distinct}(x_1 \ldots x_k))(E(x_1,x_k) \wedge E(x_2,x_k) \wedge \cdots \\
&\qquad\qquad \wedge E(x_{i-1},x_k) \wedge \neg E(x_i,x_k) \wedge \cdots \wedge \neg E(x_{k-1},x_k)) \\
&\wedge \cdots \\
&(\exists x_k.\mathrm{distinct}(x_1 \ldots x_k))(\neg E(x_1,x_k) \wedge \neg E(x_2,x_k) \wedge \cdots \wedge \neg E(x_{k-1},x_k))).
\end{aligned}$$

Truth in Almost All Structures

- Define $\mu_n(\varphi)$ to be the percentage of (ordered) structures of size n that satisfy φ,

$$\mu_n(\varphi) = \frac{|\{G : \|G\| = n; G \vDash \varphi\}|}{|\{G : \|G\| = n\}|}.$$

Lemma

For any fixed $k > 0$,

$$\lim_{n \to \infty} \mu_n(\gamma_k) = 1.$$

- Consider a random graph G of size n.

  Let $v_1, \ldots, v_{k-1}$ be a $(k-1)$-tuple of distinct vertices from G.

  Let x be any of the remaining $n + 1 - k$ vertices.

  Let c be any of the k conjuncts of the sentence γ_k.

Truth in Almost All Structures (Cont'd)

- Conjunct c asserts $k-1$ independent conditions on the existence of edges from x, each of which has probability $\frac{1}{2}$.

  For this reason, the probability that x does not meet condition c for $v_1, \ldots, v_{k-1}$ is

  $$\alpha = 1 - \frac{1}{2^{k-1}}.$$

  Thus, the probability that none of the $(n+1-k)$ x's satisfies condition c is $\alpha^{n+1-k} = \left(1 - \frac{1}{2^{k-1}}\right)^{n+1-k}$.

  It follows that the probability that G does not satisfy γ_k is

  $$< k \cdot n^{k-1} \alpha^{n+1-k} = k \cdot n^{k-1} \left(1 - \frac{1}{2^{k-1}}\right)^{n+1-k}.$$

  This expression goes quickly to 0 as n goes to infinity.

Structures Satisfying γ_k

- The sentence γ_k says that any next move in the game $\mathcal{G}^k$ can be matched by Duplicator.

Lemma

Let G and H be undirected graphs satisfying γ_k. Then

$$G \sim^k H.$$

- We show, by induction on m, that $G \sim_m^k H$.
  Consider the base case, when $m = 0$.
  Then there are no chosen points. So $G \sim_0^k H$ holds vacuously.
  Suppose that $G \sim_m^k H$.
  Let Duplicator play the $m + 1$ move game as follows.
  For the first m moves she follows her winning strategy for $\mathcal{G}_m^k(G, H)$.
  Thus, she has not lost yet.

Structures Satisfying γ_k (Cont'd)

- Consider the last move.

  Suppose that Spoiler picks up pair k of pebbles and places one of them on some vertex v of G.

  By a previous observation, we may assume that the previously pebbled points are all distinct.

  By hypothesis, $H \vDash \gamma_k$.

  So there exists a vertex v' of H, such that for all $j < k$,

  there is an edge from v' to $\beta_m(x_j)$ in H
  iff there is an edge from v to $\alpha_m(x_j)$ in G.

  Thus, Duplicator answers by putting her pebble k on v'.

  In this way, she wins the game.

Generalizing γ_k

- We can generalize γ_k as follows.
- Let $\tau = \langle R_1^{a_1}, \ldots, R_r^{a_r} \rangle$ be a vocabulary with no constant symbols.
- Let A be the set of all atomic formulas of the form $R_i(y_1, \ldots, y_{a_i})$, such that

$$x_k \in \{y_1, \ldots, y_{a_i}\} \subseteq \{x_1, \ldots, x_k\}.$$

- Define $\gamma_k(\tau)$ to be the following conjunction, which says that "every $(k-1)$-tuple may be extended to a k-tuple in any conceivable way",

$$\begin{array}{rcl}\gamma_k(\tau) & \equiv & (\forall x_1 \ldots x_{k-1}.\mathrm{distinct}(x_1, \ldots, x_{k-1})) \\ & & \bigwedge_{S \subseteq A}((\exists x_k.\mathrm{distinct}(x_1, \ldots, x_k))(\bigwedge_{\alpha \in S} \alpha \wedge \bigwedge_{\alpha \in A-S} \neg\alpha)).\end{array}$$

- The preceding lemmas go through for any such $\gamma_k(\tau)$.

Zero-One Law

- We see that any property expressible by a set of sentences from $\mathcal{L}^k(\tau)$ is true in almost all structures, or false in almost all structures.
- This is sometimes known as the **zero-one law for** $\mathcal{L}^{\omega}_{\infty\omega}$.

Theorem (Zero-One Law)

Let $S \subseteq \mathcal{L}^k$ be any set of k variable sentences over a finite vocabulary τ, with no constant or function symbols. Then the limit

$$\lim_{n\to\infty} \mu_n(S)$$

exists and is equal to zero or one.

Zero-One Law (Cont'd)

- By the preceding lemma, for every sentence $\varphi \in S$,

$$\gamma_k \vdash \varphi \quad \text{or} \quad \gamma_k \vdash \neg\varphi.$$

Thus, by a previous lemma,

$$\lim_{n \to \infty} \mu_n(S)$$

exists and:

(a) If γ_k implies every sentence in S,

$$\lim_{n \to \infty} \mu_n(S) = 1;$$

(b) If γ_k implies the negation of some sentence in S,

$$\lim_{n \to \infty} \mu_n(S) = 0.$$

Another Zero-One Law

Corollary

Assume that no constant symbols occur.
Then a zero-one law holds for the language FO(wo≤).
Furthermore, a zero-one law holds for the languages

$$\text{FO(wo}\leq\text{)(TC)}, \quad \text{FO(wo}\leq\text{)(LFP)}, \quad \text{FO(wo}\leq\text{)(PFP)},$$

where the operators TC and PFP are formally defined in later chapters.

- The key step involves showing that any sentence in one of these languages is equivalent to an infinite disjunction of sentences from $\mathcal{L}^k(\tau)$, for some k and τ.

  Now γ_k determines the truth of any sentence in $\mathcal{L}^k(\tau)$.

  So it also determines the truth of any infinite disjunction of such sentences.

Bounded Expressive Power on Average

- Suppose first-order logic has a zero-one law for the class $\mathcal{C}$ of structures.
- It can be shown that this implies that, for each k, $\mathcal{L}^k$ has **bounded expressive power on average**.
- This means that, there exists a fixed bound b, such that almost all elements of $\mathcal{C}$ fall in one of b $\mathcal{L}^k$-equivalence classes.
- That is, when talking about typical structures, $\mathcal{L}^k$ can express only a bounded number of facts.

Constants

- The zero-one laws do not hold for ordered structures or for structures with constants.

  Example: Consider the language of graphs with a constant 0.

  Then we have
  $$\mu_n(E(0,0)) = \frac{1}{2}.$$

  Example: Consider ordered structures $\mathcal{A}$ and $\mathcal{B}$.

  If $\mathcal{A} \equiv^2 \mathcal{B}$, then $\|\mathcal{A}\| = \|\mathcal{B}\|$.

  Thus, for $k \geq 2$, $\mathcal{L}^k$ is not bounded.

Subsection 6

Ehrenfeucht-Fraïssé Games with Ordering

Upper Bounding the Complexity Lower Bounds

Proposition

Let G and H be ordered graphs and let $n = \max(\|G\|, \|H\|)$. Then

$$G \sim^3_{\lceil \log(n-1) \rceil + 1} H \quad \text{implies} \quad G = H.$$

- Assume, to the contrary, that $G \sim^3_{\lceil \log(n-1) \rceil + 1} H$ but $G \neq H$.

  Let $n = \|G\|$ and $m = \|H\|$ and suppose it was the case that $n < m$.

  Let $\text{PATH}_{<_d}(x, y)$ mean that there is a path of length at most d from x to y, where each step is given by the less than relation.

  Thus,

  $$G \models \text{PATH}_{<_{n-1}}(0, \max) \quad \text{but} \quad H \models \neg \text{PATH}_{<_{n-1}}(0, \max).$$

  By a previous proposition, $\text{PATH}_{<_{n-1}} \in \mathcal{L}^3_{\lceil \log(n-1) \rceil}$.

  We conclude that $n = \|G\| = \|H\|$.

Upper Bounding the Complexity Lower Bounds (Cont'd)

- By our assumption, $G \neq H$.

  So there must be a pair of vertices i, j, such that there is an edge from vertex i to vertex j in one of the graphs but not the other.

  Consider the game $\mathcal{G}^3_{\lceil \log(n-1) \rceil + 1}(G, H)$.

  Spoiler plays the vertices i and j in G in his first two moves.
  - Suppose Duplicator answers with vertices i and j from H.
    In this case, she loses immediately.
  - Suppose Duplicator does not answer with these vertices.
    Then (G, α_2) and (H, β_2) disagree on a formula of the form

    $$\mathrm{PATH}_{<d}(x_k, c),$$

    for $k \in \{1, 2\}$, $c \in \{\max, 0\}$, and $d \leq \frac{n-1}{2}$.

  So Spoiler wins the remaining $\lceil \log(n-1) \rceil - 1$ move game.

EVEN With Ordering Without BIT

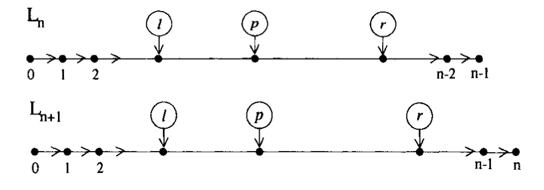
- We saw that, without ordering, we needed $n+1$ variables to say that a structure has size exactly n or that it has even cardinality.
- Thus, without ordering, even languages as strong as FO(LFP) cannot express very simple queries.
- We now show that, with ordering but still without BIT, quantifier rank $\log n$ is necessary to count even mod 2.

EVEN With Ordering Without BIT (Cont'd)

Proposition

The sentence EVEN, meaning that the cardinality of the universe is even, is not expressible in quantifier rank $\lceil \log(n-1) \rceil - 1$ with ordering, but without BIT.

- Consider the graphs L_n and L_{n+1} shown in the figure.



L_n and L_{n+1} are $(\lceil \log(n-1) \rceil - 1)$-equivalent.

However, they disagree on property EVEN.

REACH With Ordering Without BIT

Corollary

Boolean query REACH is not expressible in quantifier rank $\lceil \log(n-1) \rceil - 1$ with ordering, but without BIT.

- Define G_n and G_{n+1} to be graphs that have the same universe and ordering relation as L_n and L_{n+1}, respectively.

  Let $s = 0$ and $t = \max$.

  Replace the edge predicate by the following relation, meaning that the points are two steps apart in the ordering,

  $$E(x,y) \equiv (\exists z)(\mathrm{SUC}(x,z) \wedge \mathrm{SUC}(z,y)).$$

  Thus, REACH holds for one of G_n, G_{n+1} and not the other.

  However, G_n and G_{n+1} are still $\lceil \log(n-1) \rceil - 2$ equivalent.

  To see this, note that any win by Spoiler in $\mathcal{G}_r(G_n, G_{n+1})$ can be converted in one more move to a win by Spoiler in $G_{r+1}(L_n, L_{n+1})$.