Introduction to Descriptive Complexity

George Voutsadakis[1]

[1]Mathematics and Computer Science
Lake Superior State University

LSSU Math 600

1. Second-Order Lower Bounds
   - Second-Order Games
   - SO∃(monadic) Lower Bound on Reachability
   - Lower Bounds Including Ordering

Subsection 1

Second-Order Games

SO∃(monadic) Games

Definition (SO∃(monadic) Games)

Let $\mathcal{A}$, $\mathcal{B}$ be structures of the same vocabulary.
For $c, m \in \mathbb{N}$, define the **second-order** (**monadic**) c-**color**, m-**move game** on $\mathcal{A}$, $\mathcal{B}$ as follows:

- The two players start with the coloring phase in which Spoiler chooses c monadic relations $C_1^{\mathcal{A}}$, $C_2^{\mathcal{A}}$, ..., $C_c^{\mathcal{A}}$ on $|\mathcal{A}|$.
- Duplicator answers with c monadic relations $C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}$ on $|\mathcal{B}|$.

Observe that the coloring phase is not symmetric, in that Spoiler must play on $\mathcal{A}$.
Next, the two players play the m-move Ehrenfeucht-Fraïssé game on the two expanded structures, i.e., they play

$$\mathcal{G}_m((\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}), (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})).$$

Characterization of Duplicator's Win

Theorem

Let $\mathcal{A}$ and $\mathcal{B}$ be two not necessarily finite structures of the same finite, relational vocabulary and let $c, m \in \mathbb{N}$. Then the following two conditions are equivalent:

1. For any formula $\Phi \equiv (\exists C_1^1 \cdots C_c^1)(\varphi)$, with φ first-order of quantifier rank m,
$$\mathcal{A} \vDash \Phi \quad \text{implies} \quad \mathcal{B} \vDash \Phi;$$
2. Duplicator has a winning strategy for the second-order (monadic) c-color, m-move game on $\mathcal{A}, \mathcal{B}$.

- The theorem follows from:
  - The theorem characterizing the Duplicator having a winning strategy in $\mathcal{G}_m^k(\mathcal{A}, \mathcal{B})$;
  - The fact that there are only finitely many inequivalent formulas in $\mathcal{L}_m(\tau \cup \{C_1^1, \ldots, C_c^1\})$.

Characterization of Duplicator's Win (Cont'd)

- Suppose Condition 1 holds.

  Let $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$ be Spoiler's move in the coloring phase.

  Let φ be the conjunction of all quantifier-rank m sentences that are true of $(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}})$.

  By hypothesis, it follows that

  $$\mathcal{B} \models (\exists C_1^1 \cdots C_c^1)\varphi.$$

  Duplicator, thus, can play $C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}$ that are witnesses of φ.

  It follows that

  $$(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}) \equiv_m (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}).$$

  So Duplicator wins the first-order part of the game.

Characterization of Duplicator's Win (Cont'd)

- Conversely, suppose Condition 1 is false.

  So we have $\mathcal{A} \models \Phi$ but $\mathcal{B} \not\models \Phi$.

  Spoiler plays the coloring phase by choosing $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$ witnessing the truth of Φ.

  However Duplicator responds, the two structures

$$(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}),$$
$$(\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})$$

  disagree on the quantifier rank m formula φ.

  So Spoiler wins the first-order part of the game.

Complete Formulas

- We look, next, at a complete methodology, using SO∃(monadic) Ehrenfeucht-Fraïssé games, for determining whether a boolean query is expressible in SO∃(monadic).
- Recall that there are only finitely many inequivalent formulas in a given finite relational language restricted to a given quantifier rank and with a given number of free variables.

Definition

Let $\mathcal{L}$ be a language. We say that φ is a **complete formula of** $\mathcal{L}$ if it is consistent and maximal, in the sense that if $\psi \in \mathcal{L}$ is another consistent formula that implies φ, then φ and ψ are equivalent.

Consider the second-order (monadic) c-color, m-move game on structures of the finite, relational vocabulary τ.

Let $C = C(c, m, \tau)$ be the finite number of such inequivalent, complete formulas in $\mathcal{L}_m(\tau \cup \{C_1^1, \ldots, C_c^1\})$ that have one free variable.

Introducing the Ajtai-Fagin Game

- Suppose we play the second-order (monadic) c-color, m-move game on structures of the finite, relational vocabulary τ.
- The result of Spoiler's coloring a structure with c new monadic relations is that he partitions the universe into a larger, but still finite, number $C = C(c, m, \tau)$ of equivalence classes.
- This equivalence relation can be described as follows.

  Let $a, a' \in |\mathcal{A}|$, where $\mathcal{A} \in \text{STRUCT}[\tau \cup \{C_1^1, \ldots, C_c^1\}]$.

  Then a, a' are equivalent iff $(\mathcal{A}, a) \sim_m (\mathcal{A}, a')$.
- Since the SO∃(monadic) Ehrenfeucht-Fraïssé game is still difficult for Duplicator to play, Ajtai and Fagin invented an equivalent game.
- The two games are equivalent in that Duplicator has a winning strategy in one iff she has a winning strategy in the other.

Ajtai-Fagin Game

Definition (Ajtai-Fagin Game)

Let $I \subseteq \text{STRUCT}[\tau]$ be a boolean query.
Define the **Ajtai-Fagin game on I** as follows.

1. Spoiler chooses natural numbers c and m.
2. Duplicator chooses a structure $\mathcal{A} \in \text{STRUCT}[\tau]$, such that $\mathcal{A} \in I$.
3. Spoiler chooses c monadic relations $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$ on $|\mathcal{A}|$.
4. Duplicator chooses a structure $\mathcal{B} \in \text{STRUCT}[\tau]$, such that $\mathcal{B} \notin I$.
   Duplicator also chooses c monadic relations $C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}$ on $|\mathcal{B}|$.
5. Finally, the two players play

$$\mathcal{G}_m((\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}), (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})).$$

The Ajtai-Fagin Methodology Theorem

Theorem (Ajtai-Fagin Methodology Theorem)

Let $I \subseteq \text{STRUCT}[\tau]$ be a boolean query. Then the following are equivalent:

1. Duplicator has a winning strategy for the Ajtai-Fagin game on I;
2. $I \notin \text{SO}\exists(\text{monadic})$.

- Suppose $I = \text{MOD}[\Phi]$, where

$$\Phi \equiv (\exists C_1^1 \cdots C_c^1)(\varphi),$$

  φ of quantifier rank m.

  We show that Spoiler has a strategy for winning the Ajtai-Fagin game on I.

The Ajtai-Fagin Methodology Theorem (Converse)

- Spoiler's first move is to choose c, m.

  Let $\mathcal{A} \in I$ be the structure chosen by Duplicator.

  Spoiler chooses colorings $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$ such that

$$(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}) \vDash \varphi.$$

  Duplicator then chooses a structure $\mathcal{B} \notin I$, so $\mathcal{B} \vDash \neg\Phi$.

  Thus, whatever coloring $C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}$ is chosen by Duplicator, we know that

$$(\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}) \vDash \neg\varphi.$$

  So Spoiler wins $G_m((\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}), (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}}))$.

The Ajtai-Fagin Methodology Theorem (Converse Cont'd)

- Conversely, suppose $I \notin \text{SO}\exists(\text{monadic})$.

  We describe a winning strategy for Duplicator.

  Let Spoiler choose the numbers c, m.

  Let S_m be a maximal set of inequivalent sentences of the form

$$\Phi \equiv (\exists C_1^1 \cdots C_c^1)(\varphi),$$

  φ of quantifier rank m, where the first-order part φ is a complete sentence of quantifier rank m.

  S_m is finite, by a previous assertion.

  The sentences that cause us trouble are those that are not satisfied by any structure not in I,

$$T \equiv \{\Phi \in S_m : \forall \mathcal{B} \in (\text{STRUC}[\tau] - I), \mathcal{B} \vDash \neg\Phi\}.$$

  Let Ψ be the disjunction of all sentences in T.

The Ajtai-Fagin Methodology Theorem (Converse Cont'd)

- We let Ψ be the disjunction of all sentences in

$$T \equiv \{\Phi \in S_m : \forall \mathcal{B} \in (\mathrm{STRUC}[\tau] - I), \mathcal{B} \vDash \neg\Phi\}.$$

Note that Ψ is of the form described above, since the disjunction can be pushed through the second-order existential quantifiers.

By assumption, there exists a structure $\mathcal{A} \in I$, such that $\mathcal{A} \vDash \neg\Psi$.

Otherwise, Ψ would express I.

Duplicator should play this $\mathcal{A}$.

Let Spoiler choose colors $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$.

The Ajtai-Fagin Methodology Theorem (Converse Cont'd)

- Let φ_0 be the complete quantifier-rank m sentence satisfied by $(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}})$.

  Let
  $$\Phi_0 = (\exists \overline{C})\varphi_0.$$

  Then $\mathcal{A} \vDash \Phi_0$. So $\Phi_0 \notin T$.

  Therefore, by the definition of T, there exists $\mathcal{B} \in (\mathrm{STRUC}[\tau] - I)$, such that
  $$\mathcal{B} \vDash \Phi_0.$$

  Duplicator plays this $\mathcal{B}$, together with a coloring that witnesses Φ_0.

  It follows that Duplicator wins the game
  $$G_m((\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}), (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})).$$

Application: Connectivity

Theorem

The connectivity problem for undirected graphs is not expressible in monadic second-order existential logic without numeric relations,

$$\text{CONNECTED} \notin \text{SO}\exists(\text{monadic})(\text{wo}\leq).$$

- We show that Duplicator has a winning strategy for the Ajtai-Fagin game on CONNECTED.

  Suppose that Spoiler chooses constants c, m.

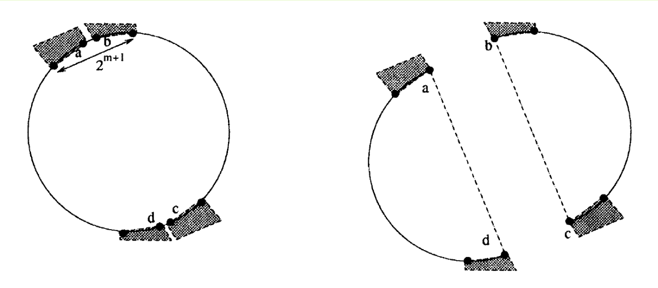
  Duplicator responds with a sufficiently large cycle $\mathcal{A}$.

  Sufficiently large means that

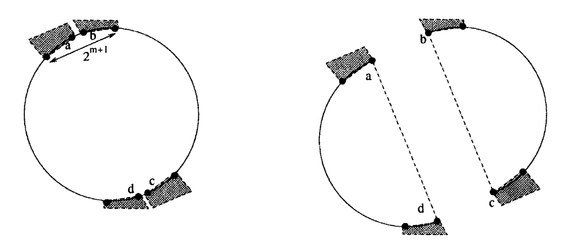
$$\|\mathcal{A}\| \geq h(2^{ch} + 1), \quad h = 2^{m+1} + 1.$$

Application: Connectivity (Cont'd)

- Let $C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}$ be the coloring on $\mathcal{A}$ played by Spoiler.
  The neighborhood $N(a, 2^m)$ of any vertex a contains h vertices.
  The number of possible colorings of such a neighborhood is thus 2^{ch}.
  But $\|\mathcal{A}\|$ is chosen to be at least $h(2^{ch}+1)$.
  So it must contain at least two disjoint neighborhoods $N(a, 2^m)$ and $N(c, 2^m)$ containing identical colorings in clockwise order around the cycle.



Application: Connectivity (Cont'd)



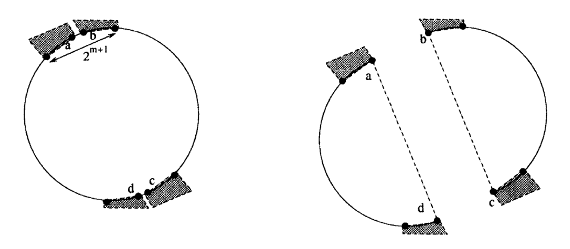
- Let b be the next vertex after a and d the next vertex after c in clockwise order around $\mathcal{A}$.

  To construct $\mathcal{B}$, Duplicator removes edges (a, b) and (c, d) and replaces them by edges (a, d) and (b, c).

  Thus, $\mathcal{B}$ consists of two disjoint cycles.

  Duplicator colors $\mathcal{B}$ exactly as Spoiler has colored $\mathcal{A}$.

Application: Connectivity (Cont'd)



- It follows that the structures

$$(\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}),$$
$$(\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})$$

  have the same number of each 2^m type.
  Thus, by Hanf's Theorem, Duplicator wins the game

$$G_m((\mathcal{A}, C_1^{\mathcal{A}}, C_2^{\mathcal{A}}, \ldots, C_c^{\mathcal{A}}), (\mathcal{B}, C_1^{\mathcal{B}}, C_2^{\mathcal{B}}, \ldots, C_c^{\mathcal{B}})).$$

Negation of Connectivity

- The negation of connectivity is expressible in SO∃(monadic)(wo≤).
- The sentence asserts the existence of a set S, such that:
  - S is not empty;
  - S does not contain all vertices;
  - S contains all the neighbors of its elements.
- So this set S contains a proper connected component of the graph.
- It follows that the graph is not connected.

$$\begin{aligned}\overline{\text{CONNECTED}} \equiv (\exists S^1)[&(\exists xy)(S(x) \wedge \neg S(y)) \wedge \\ &(\forall xy)((S(x) \wedge E(x,y)) \rightarrow S(y))].\end{aligned}$$

Corollary

SO∃(monadic)(wo≤) is not closed under complementation.

- The corollary holds with arbitrary numeric predicates.

ALL-EVEN-DEGREE

- Let ALL-EVEN-DEGREE be true of undirected graphs all of whose vertices have an even number of edges.
- Ajtai showed that ALL-EVEN-DEGREE is not in SO$\exists$, in the presence of arbitrary numeric relations.
- However, in the presence of an ordering relation, ALL-EVEN-DEGREE is expressible in second-order, monadic, universal logic.
- The sentence asserts that, for all two-colorings of the graph, and for all vertices v, if the coloring of the neighbors of v alternates between the two colors, then v's first neighbor has a different color than its last neighbor.
- Note also, that it is an SO$\exists$ property that a particular vertex has even degree.

ALL-EVEN-DEGREE (Cont'd)

Corollary

The language SO∃(monadic) is not closed under either of:

- Complementation;
- First-order quantification.

In particular, ALL-EVEN-DEGREE is expressible in SO∀(monadic) and in the form $(\forall x)$SO∃(monadic), using ordering as the only numeric predicate. However, ALL-EVEN-DEGREE is not expressible in second-order existential, monadic logic in the presence of arbitrary numeric predicates.

Subsection 2

SO∃(monadic) Lower Bound on Reachability

Undirected Graph Reachability in SO∃(monadic)

Proposition

The undirected reachability query is expressible in second order, existential, monadic logic, without numeric relations. In symbols,

$$\text{REACH}_u \in \text{SO}\exists(\text{monadic})(\text{wo}\leq).$$

- We express the existence of an undirected path from s to t.
  The sentence asserts the existence of a set of vertices S, such that:
  1. Vertices s and t are members of S;
  2. Vertices s and t each have unique neighbors in S;
  3. All other members of S have exactly two neighbors in S.

  Clearly, these three conditions are first-order expressible using S.

  Furthermore, any such set S must include a path from s to t.

  Conversely, the vertices along a shortest path from s to t constitute such a set S.

Non-Robustness of Monadic SO Existential Logic

- Connectivity is first-order reducible (in fact quantifier-free reducible) to REACH_u.
- We will see that connectivity is not expressible in monadic, second-order existential logic, even in the presence of ordering.
- Whether this holds in the presence of BIT is an open question.
- The following corollary says that monadic second-order existential logic is not very robust.

Corollary

The language SO∃(monadic)(woBIT) is not closed under quantifier-free reductions.

Reachability and Monadic SO Existential wo $\leq$

Theorem

REACH $\notin$ SO∃(monadic)(wo$\leq$).

- The reader may have noticed that the proof of the preceding proposition does not work for directed graphs.

  The reason is that a graph $G \in$ REACH may have the property that every set of vertices forming a path from s to t admits a "back edge".

  I.e., an edge from a vertex close to t, to a vertex farther away from t.

  We show instead that Duplicator wins the Ajtai-Fagin game on REACH.

Reachability (Cont'd)

- Let Spoiler begin by playing c and m.

  Recall that this indicates Spoiler's intention to:
  - Define c new monadic relations $C_1, \ldots, C_c$;
  - Play the m-move first-order game.

  Let $C = C(c, m, \tau)$ be the number of inequivalent, quantifier-rank m formulas in the language of graphs extended by relations $C_1, \ldots, C_c$.

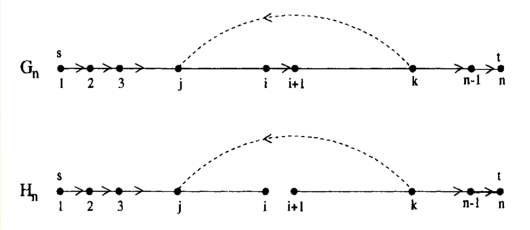
  Consider the set of random graphs G_n, consisting of:
  - A directed path $s = g_1, g_2, \ldots, g_{n-1}, g_n = t$;
  - Some random back edges, (g_i, g_j), for $j < i$.

  These edges will be chosen independently at random with probability $p(n) = n^{\sigma-1}$.

  The small constant σ is chosen by Duplicator, depending on constants m and c chosen by Spoiler.

Reachability (Cont'd)

- The figure shows a drawing of G_n, together with the graph H_n that Duplicator chooses later.



H_n is the same as G_n and colored the same as G_n, except that one forward edge from h_i to h_{i+1} is missing.

Note that $G_n \in$ REACH and $H_n \notin$ REACH.

Reachability (Cont'd)

- Duplicator now plays one of the random graphs G_n meeting the four conditions of the following probabilistic lemma.

  The proof of the lemma is given after the end of this proof.

  We take for simplicity in the following

  $$n > 100m^2C^2 \quad \text{and} \quad \varepsilon = 0.01.$$

  All discussions of distances, paths and cycles in the following concern undirected paths, i.e., paths in the Gaifman graph.

Reachability (Cont'd)

Lemma

Let C and m be natural numbers, and let $\epsilon > 0$.
Let $\sigma > 0$ be sufficiently small and let n be sufficiently large.
Let G_n be chosen at random, each back edge chosen with probability $n^{\sigma-1}$.
With high probability, the following conditions hold:

1. G_n has fewer than n^ϵ undirected cycles of length at most 2^m.
2. For every vertex $v \in G_n$, the number of vertices of distance at most 2^m from v is less than n^ϵ.
3. No matter how the vertices of G_n are colored, using C colors, a fraction of $(1-\epsilon)$ of the vertices g_i of G_n have at least m back edges to vertices colored the same color as g_{i+1}.
4. No matter how the vertices of G_n are colored, using C colors, a fraction of $(1-\epsilon)$ of the vertices g_{i+1} of G_n have at least m back edges from vertices colored the same color as g_i.

Reachability (Cont'd)

- Spoiler colors G_n with c color relations, forming G_n^c.

  This induces a coloring of each vertex of G_n^c by its quantifier-rank $m2^m$ type.

  Thus, each vertex has one of C colors.

  Call this expansion G_n^C.

  By our choice of G_n and the values of n and ϵ, we know that some vertex v satisfies the following conditions:

  1. The distance of g_i to any cycle of length $\leq 2^m$ is greater than $2^m + 1$.
  2. In G_n^C, the following hold:
     - g_i has $\geq m$ back edges to vertices of the same color as g_{i+1};
     - g_{i+1} has $\geq m$ back edges from vertices of the same color as g_i.

Reachability (Cont'd)

3. There are $2m$ vertices $a_1, \ldots, a_m$, and $b_1, \ldots, b_m$ in G_n^C, such that:
   - Each pair of these vertices are distance greater than $2^m + 1$ from each other and from g_i;
   - The color of the a_i's is the same as g_i;
   - The color of the b_i's is the same as g_{i+1}.

Let H_n^c be G_n^c with the edge from g_i to g_{i+1} removed.

Duplicator plays H_n^c.

Reachability (Cont'd)

Claim: $G_n^c \sim_m H_n^c$.

Each vertex v in G_n^C has been colored by the m-color of v.

That is, v is colored by a complete quantifier-rank m description of v.

We now show that the coloring of each vertex v of H_n^C is still the unchanged quantifier-rank m description of v.

This follows from the above three conditions.

Especially important is that g_i is not within distance $2^m + 1$ of a cycle of length at most 2^m.

It follows that for the m-move game, the set of vertices near g_i, g_{i+1} and h_i, h_{i+1} form a forest.

Furthermore, the m-colors of g_i and g_{i+1} guarantee this property of not being near a small cycle.

Thus, the a_i's and b_i's from Condition 3 above also have this property.

That is, their 2^m neighborhoods are disjoint trees.

Reachability (Cont'd)

- Duplicator's winning strategy for the m-move game on G_n^c and H_n^c is now clear.

  At the first move, she answers any play v by Spoiler with a vertex v' of the same color.

  From then on, Duplicator answers almost any move by Spoiler according to her winning strategy in the game on (G_n, v), (G_n, v').

  The exception is if this strategy would call for a move near g_i or g_{i+1} when a vertex near the other of these points has already been played.

  If the move is near g_i, Duplicator should answer by substituting for g_i a "new" vertex w, where:

  - w is one of the a_i's, if Duplicator is answering a vertex not near g_{i+1};
  - w is one of the vertices of the same color as g_i having a back edge to g_{i+1}, otherwise.

  By assumption, m such vertices are available.

Proof of the Lemma

Fact (The Weak Law of Large Numbers)

Suppose there are n independent trials, and each has probability p of success. Let S_n be the number of successful trials. Let $M = pn$ be the expected number of successful trials. Then, for any $\rho > 0$,

$$\lim_{n\to\infty} \text{Prob}[|S_n - M| \geq \rho M] = 0.$$

- Conditions (1) and (2) of the lemma are easily met by letting $\sigma \leq \frac{\epsilon}{4^m}$, and then letting n grow.

  In particular, the expected number of cycles of length at most 2^m is less than $n^{2^m}(\frac{n^\sigma+2}{n})^{2^m}$. And this is less than $n^{\epsilon/2}$, for large n.

  Similarly, the expected number of neighbors of distance at most 2^m of any vertex v is less than $(n^\sigma + 2)^{2^m} < n^{\epsilon/2}$.

Proof of the Lemma (Condition 3)

- To prove Condition (3), let

$$\alpha = \frac{\epsilon}{2C}.$$

Let A be a set of αn vertices from G_n.

Let v be a vertex to the right of all the vertices in A.

The expected number of back edges from v to A is αn^σ.

Let S_v be the number of such back edges in a randomly chosen G_n.

Let

$$\delta = \text{Prob}\left[S_v < \frac{\alpha}{2} n^\sigma\right].$$

By the Law of Large Numbers, we can choose n so large that δ is as small as we like.

We choose n so that

$$\delta^\alpha < \frac{1}{8}.$$

Proof of the Lemma (Condition 3)

- Let B be a set of αn vertices all to the right of all the vertices in A. Let $E(A,B)$ be the event that

$$S_v < \frac{\alpha}{2} n^{\sigma}, \quad \text{for all } v \text{ in } B.$$

Thus,

$$\text{Prob}[E(A,B)] \leq \delta^{\alpha n}.$$

Clearly, there are fewer than 2^{2n} choices of such sets A and B. Thus, the probability that any of them satisfy event $E(A,B)$ is

$$\leq 2^{2n}\delta^{\alpha n} \leq 2^{2n}\left(\frac{1}{8}\right)^n = \frac{1}{2^n}.$$

Thus, with high probability, there are no such sets A and B.

Proof of the Lemma (Condition 3 Cont'd)

- Let G_n be a graph with random back edges having no sets A and B satisfying $E(A, B)$.

  Suppose that each vertex in G_n is colored one of C colors.

  Let S be a set of vertices g_i of G_n such that:

  - g_{i+i} has color C_ℓ;
  - g_i has fewer than m back edges to vertices of color C_ℓ.

  Suppose, for the sake of contradiction, that

$$|S| > \frac{n\epsilon}{C}.$$

  Let B be the rightmost half of the vertices of S.

Proof of the Lemma (Condition 3 Cont'd)

- Let A consist of all vertices of color C_ℓ lying to the left of all the vertices of B.

  Thus,
  $$|A|, |B| \geq \frac{n\epsilon}{2C} = \alpha n.$$

  It would now follow that $E(A, B)$ holds.

  This is impossible since G_n was chosen with no such sets A and B.

  It follows that
  $$|S| \leq \frac{n\epsilon}{C}.$$

  Thus, the number of vertices g_i that do not have at least m back edges to vertices of the same color as g_{i+1} is at most $n\epsilon$.

  This proves Condition (3).

  Condition (4) can be proved similarly.

Subsection 3

Lower Bounds Including Ordering

Introduction

- We strengthen the Connectivity Theorem.
- We show that, even in the presence of ordering, graph connectivity is not expressible in monadic, second-order existential logic.
- The argument is subtle in that every two vertices appear together in a tuple in the ordering relation.
- Thus, every Gaifman graph has diameter one.
- It follows that a proof using Hanf's Theorem is not possible.
- The main interest in this result is that it introduces a new way for Duplicator to win a game in the presence of ordering.
- Tight complexity lower bounds on nondeterministic time can be proved in the presence of ordering and addition.

The Graph P_s^n

- Consider the set of all permutations on n objects

$$S_n = \{\sigma_1, \sigma_2, \ldots, \sigma_{n!}\}.$$

- Let $s = \pi_1, \ldots, \pi_r$ be a sequence of elements of S_n.
- Define the ordered graph $P_s^n = (V_s^n, E_s^n)$ by setting:

$$\begin{aligned} V_s^n &= \{1, 2, \ldots, r+1\} \times \{1, 2, \ldots, n\}; \\ E_s^n &= \{(\langle i, j\rangle, \langle i+1, \pi_i(j)\rangle) : 1 \leq i \leq r, 1 \leq j \leq n\}. \end{aligned}$$

- Thus, P_s^n consists of n disjoint paths of length r.
- The ordering on P_s^n is the lexicographic ordering.

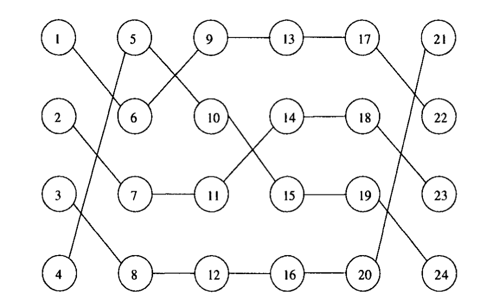
Example: The Graph P_s^4

- Consider S_4 and let

$$s = (1234), (12), (23), e, (1234),$$

where e is the identity permutation.

The figure shows a drawing of P_s^4.



Definitions

- For any permutation $\sigma_i \in S_n$, let Q_i be the sequence consisting of 2^m copies of the identity, σ_i and another 2^m copies of the identity,

$$Q_i = \underbrace{e, \ldots, e}_{2^m}, \sigma_i, \underbrace{e, \ldots, e}_{2^m}.$$

- Let σ_{i_0} be the inverse of the product of all $n!$ permutations in S_n,

$$\sigma_{i_0} = \left(\prod_{i=1}^{n!} \sigma_i \right)^{-1}.$$

- Define the following sequence of permutations,

$$T = Q_{i_0}, Q_1, Q_2, \ldots, Q_{n!}.$$

- The graph P_T^n consists of n disjoint paths of length

$$\ell = (2^{m+1} + 1)(n! + 1).$$

Definitions (Cont'd)

- The product of sequence T is the identity permutation.
- So, taking into account the lexicographic ordering of the vertices, these paths connect vertices i and $n\ell + i$, for $i = 1, 2, \ldots, n$.
- Let sequence Z consist of N copies of T followed by a single copy of Q_{i_1}, where σ_{i_1} is the n-cycle $(12\cdots n)$,

$$Z = \underbrace{T, \ldots, T}_{N}, Q_{i_1}.$$

- The length of this sequence is

$$L = N\ell + (2^{m+1} + 1).$$

Definitions (Cont'd)

- The product of sequence Z is just the n-cycle $(12\cdots n)$.
- Thus, P_Z^n consists of n paths of length L, connecting vertex i on the left to vertex $i+1$ mod n on the right, for $i = 1, 2, \ldots, n$.
- Let $\mathcal{A}_n$ be the graph P_Z^n together with the n edges

$$\{(i, nL + i) : 1 \leq i \leq n\},$$

  connecting the first and last vertices in each row.
- $\mathcal{A}_n$ consists of a single long cycle and is thus connected.

The Ajtai-Fagin Game on Boolean Query CONNECTED

- We play the Ajtai-Fagin game on boolean query CONNECTED.
- At the first move, Spoiler chooses constants c and m.
- Duplicator plays the graph $\mathcal{A}_n$.
- The numbers n and N will be specified later to be sufficiently large.
- Let Spoiler now choose a coloring of $\mathcal{A}_n$ using c new color relations, $C_1, \ldots, C_c$.
- Let $\mathcal{A}_n^c$ be the structure $\mathcal{A}_n$ together with these new color relations.
- Each vertex $v \in |\mathcal{A}_n^c|$ has one of at most $C = C(c, m, \langle E^2, \leq^2 \rangle)$ complete descriptions in the language $\mathcal{L}_m(C_1, \ldots, C_c)$.
- Consider $\mathcal{A}_n^C$ as colored with these complete descriptions.
- Thus, each vertex has one of C possible colorings.

The Ajtai-Fagin Game (Cont'd)

- The number of possible colorings of a copy of P^n_T in $\mathcal{A}^C_n$ is $C^{n\ell}$.
- Suppose we choose $N > n!C^{n\ell}$.
- Then, there are at least $n!$ identically colored copies of P^n_T in $\mathcal{A}^C_n$.
- Let T^c be such a copy of P^n_T in $\mathcal{A}^c_n$.
- For each part P_{Q_i} of T^c, there are $C^{n(2^{m+1}+1)}$ possible colorings of all the vertices in P_{Q_i}.
- Note that $C^{n(2^{m+1}+1)} \leq B^n$ for some constant B.
- For sufficiently large n, $n!$ is much greater than B^n.
- So there exists some set of permutations $A \subseteq S_n$ of size at least $\frac{n!}{B^n}$, such that, for all $\sigma_i, \sigma_j \in A$, the colorings of P_{Q_i} and P_{Q_j} in T^c are identical.

The Interchange Lemma

- The following lemma asserts that Duplicator can interchange any such P_{Q_i}'s and P_{Q_j}'s without detection.

Lemma

Let $\mathcal{B}_n^c$ result from $\mathcal{A}_n^c$ by replacing any number of parts P_{Q_i} in a copy of T^c by the part P_{Q_j}, for pairs $\sigma_i, \sigma_j \in A$. Then

$$\mathcal{A}_n^c \sim_m \mathcal{B}_n^c.$$

- We show that Duplicator wins the m-move game on $\mathcal{A}_n^c$ and $\mathcal{B}_n^c$.

  The only difference between the two structures is in the middle two columns of any P_{Q_i} that has been changed to P_{Q_j}.

  With r moves to go, we say that a newly pebbled vertex is "near" another chosen vertex if the distance between their respective columns is at most 2^r.

The Interchange Lemma (Cont'd)

- Let Spoiler put the first pebble on any vertex in either structure.

  Duplicator should answer with the vertex of the same number in the other structure.

  Let the chosen vertices be $a = \alpha_1(x_1)$ and $b = \beta_1(x_1)$.

  Suppose that a and b are inside parts P_{Q_i} and P_{Q_j}, for $i \neq j$.

  Since $\sigma_i, \sigma_j \in A$, a and b have the same complete description in $\mathcal{L}_{m-1}$.

  Thus, Duplicator has a winning strategy in

  $$\mathcal{G}_{m-1}(\mathcal{A}_n^c, a, \mathcal{A}_n^c, b'),$$

  where b' is the piece corresponding to b in part P_{Q_j} of $\mathcal{A}_n^c$.

The Interchange Lemma (Cont'd)

- From now on, for moves near already chosen points in this part, Duplicator should answer according to her winning strategy in

  $$\mathcal{G}_{m-1}(\mathcal{A}_n^c, a, \mathcal{A}_n^c, b').$$

  Suppose, on the other hand, that the chosen vertices are inside unchanged, and therefore identical, parts of $\mathcal{A}_n^c$ and $\mathcal{B}_n^c$.

  In this case, from now on, for moves near this part, Duplicator will keep moving according to the isomorphism between these two parts.

The Interchange Lemma (Cont'd)

- In successive moves, if the newly pebbled point is near an already chosen point, then Duplicator should answer according to her winning strategy in the subgame of the already chosen point.

  If the newly pebbled point is not near any such subgame, then Duplicator answers with the vertex of the same number in the other structure.

  This pair establishes a new subgame.

  Duplicator, thus, wins all the subgames.

  Furthermore, if $\alpha_m(x_u)$ and $\alpha_m(x_v)$ belong to different subgames, then these subgames were each started with points of the same number in the two structures. Thus,

$$\alpha_m(x_u) < \alpha_m(x_v) \Leftrightarrow \beta_m(x_u) < \beta_m(x_v).$$

  It follows that Duplicator wins the whole game.

The Ajtai-Fagin Game (Cont'd)

- The lemma tells us that, when we transplant a part of the structure P_{Q_j} in place of the different, but identically colored, P_{Q_i}, then the colors, i.e., the complete descriptions in $\mathcal{L}_{m-1}$ of the vertices, remain the same!
- This transplanting changes the product of the corresponding permutations, but it is not detectable in language $\mathcal{L}_m$.
- The reason we defined the sequences Q_i to have a length 2^m buffer on each side was so that Duplicator's winning strategy for the game $\mathcal{G}_{m-1}(\mathcal{A}_n^c, a, \mathcal{A}_n^c, b')$ can be used for the subgames.
- If we can change the product enough so that it is no longer an n-cycle, then $\mathcal{B}_n^c$ will not be connected and the theorem would follow.

The Ajtai-Fagin Game (Cont'd)

Fact

For sufficiently large n, suppose that H is a subgroup of S_n, such that for all $h \in H$, the product $(12\cdots n)h$ is an n-cycle. Then

$$|H| \le n!\left(\frac{6}{\log n}\right)^n.$$

- Fix $\sigma_i \in A$. For each $\sigma_j \in A$, let

  $$T_j = Q_{i_0}, Q_1, Q_2, \ldots, Q_{i-1}, Q_j, Q_{i+1}, \ldots, Q_{n!}$$

  be the sequence T with Q_i replaced by Q_j.
  Let ρ_j be the product of the sequence T_j.
  Define H to be the subgroup of S_n generated by all the ρ_j's.
  Obviously H is at least as big as A and, thus, of size at least $\frac{n!}{B^n}$.
  By Fact, there exists $h \in H$, such that $(12\cdots n)h$ is not a cycle.

The Ajtai-Fagin Game (Cont'd)

- There exists $h \in H$, such that $(12\cdots n)h$ is not a cycle.
  By the definition of H, we can write h as the product
  $$h = \rho_{j_1} \cdot \rho_{j_2} \cdots \rho_{j_t}.$$
  We know that $t \leq n!$.
  Define $\mathcal{B}_n^c$ as the stucture resulting by replacing P_{Q_i} by $P_{Q_{j_k}}$ in t successive copies of T^c in $\mathcal{A}_n^c$.
  It follows from the lemma that
  $$\mathcal{A}_n^c \sim_m \mathcal{B}_n^c.$$
  However, $\mathcal{A}_n$ is connected and $\mathcal{B}_n$ is not.

Theorem

Connectivity is not expressible in monadic, second-order existential logic with ordering as the only numeric predicate.