Locality of Queries Definable in Invariant First-Order Logic with Arbitrary Built-In Predicates

Matthew Anderson
University of Wisconsin - Madison

Dieter van Melkebeek
University of Wisconsin - Madison

Nicole Schweikardt
Goethe-Universität Frankfurt am Main

Luc Segoufin
INRIA and ENS-Cachan, LSV

July 7^{th}, 2011

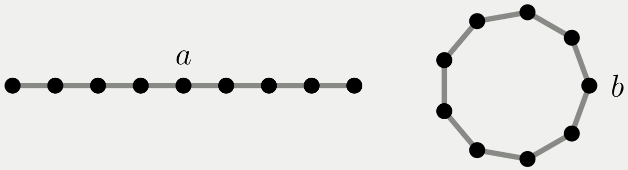
Outline

Motivation

Locality is a powerful tool for proving inexpressibility.

Example

- On-A-Cycle is a property which is not r-local, for $r = o(n)$.



- First-Order logic (FO) is constant-local.
- Therefore, On-A-Cycle is not definable in FO.

What do we know about locality?

- FO is constant-local.
- Order-invariant FO is constant-local [Grohe-Schwentick].
- Arb-invariant FO is polylog(n)-local [**us**].
  Technique: reduction to Boolean circuit lower bounds.

Background – Invariant FO

Example

$$\phi := \exists x \exists y ((x + x = y) \wedge \forall z (z \leq y)).$$

Assume that:

- $\leq$ is interpreted as a linear ordering.

  This induces a bijection $U \rightarrow \{1, 2, \ldots, n\}$.
- $+$ is interpreted in the natural way w.r.t. $\leq$:

$$x, y, z \in U \rightarrow a, b, c \in \{1, 2, \ldots, n\},$$

$$(x, y, z) \in + \text{ iff } a + b = c.$$

Then ϕ expresses whether the universe size is even.

The veracity of ϕ is independent of the actual interpretation of $\leq$.

Background – Invariant FO

Formalizing this intuition:

1. Consider a set of numerical predicates symbols $\mathcal{S}$.
2. For each $n \in \mathbb{N}$ define an interpretation of $\mathcal{S}$ over $\{1, 2, \ldots, n\}$.
3. Selecting an interpretation of $\leq$ as a linear order with respect to a universe U induces an interpretation of $\mathcal{S}$ over U.

Definition

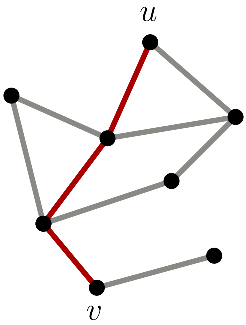
For a fixed interpretation of $\mathcal{S}$ over $\{1, 2, \ldots, n\}$, a formula $\phi(x)$ of FO$(\leq, \mathcal{S})$ is $\mathcal{S}$-invariant if for all graphs G and vertices a, the truth of $\phi(a)$ on G is independent of the particular interpretation of the linear order $\leq$.

We focus on the case where $\mathcal{S}$ is the set of arbitrary numerical predicates (Arb).

Background – Graphs

Graph $G := (V, E)$.

Distance $D(u, v)$ – length of a shortest path between u, v in G.



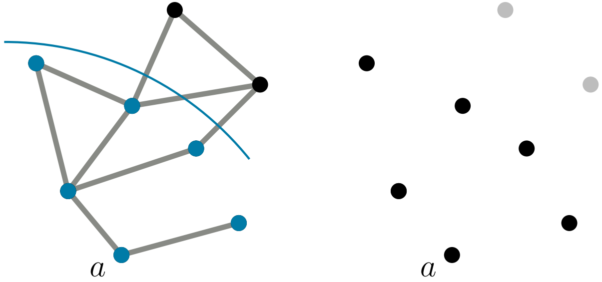
Background – Graphs

Graph $G := (V, E)$.

Distance $D(u, v)$ – length of a shortest path between u, v in G.

Ball $B_r(a)$ of radius r at a in G.

Neighborhood $\mathcal{N}_r(a)$ of radius r at a in G.



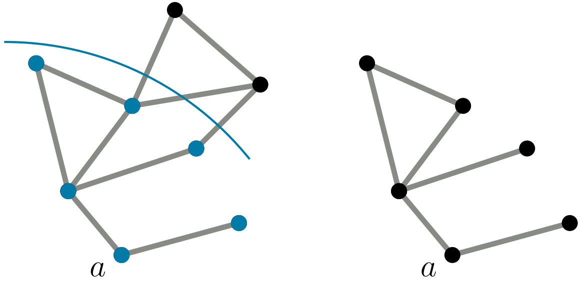
Background – Graphs

Graph $G := (V, E)$.

Distance $D(u, v)$ – length of a shortest path between u, v in G.

Ball $B_r(a)$ of radius r at a in G.

Neighborhood $\mathcal{N}_r(a)$ of radius r at a in G.

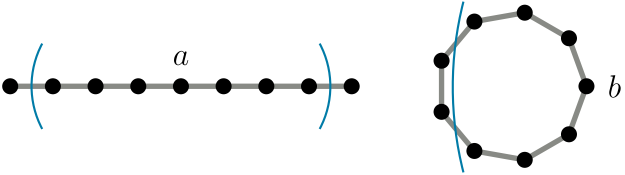


Background – Gaifman Locality

We say that two neighborhoods are isomorphic

$$\mathcal{N}_r(a) \cong \mathcal{N}_r(b),$$

if there is an isomorphism π between the two that maps a to b.

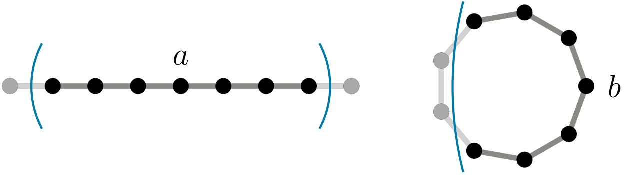


Background – Gaifman Locality

We say that two neighborhoods are isomorphic

$$\mathcal{N}_r(a) \cong \mathcal{N}_r(b),$$

if there is an isomorphism π between the two that maps a to b.

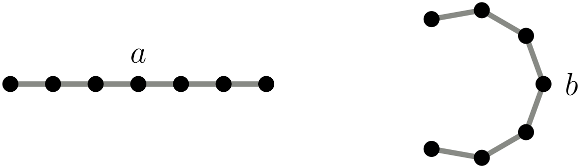


Background – Gaifman Locality

We say that two neighborhoods are isomorphic

$$\mathcal{N}_r(a) \cong \mathcal{N}_r(b),$$

if there is an isomorphism π between the two that maps a to b.



Definition

Let $f : \mathbb{N} \to \mathbb{R}_{\geq 0}$. A formula $\phi(x)$ is Gaifman f-local if for any large enough graph G with n vertices, and any two vertices a and b:

$$\mathcal{N}_{f(n)}(a) \cong \mathcal{N}_{f(n)}(b) \implies G \models \phi(a) \text{ iff } G \models \phi(b).$$

Main Result

Theorem (Main)

1. *For each Arb-invariant* FO *formula* $\phi(\boldsymbol{x})$ *there is a* $c \in \mathbb{N}$ *such that the formula is Gaifman* $(\log n)^c$*-local.*
2. *For each* $c \in \mathbb{N}$ *there is an Arb-invariant* FO *formula* $\phi(\boldsymbol{x})$ *that is not Gaifman* $(\log n)^c$*-local.*

Proof Outline

Proof Sketch of Part 1.

1. Suppose otherwise, then there is a formula $\phi(x)$

   that is not f-local for G, a and b.

2. Derive a formula $\phi'(y)$

   that is not $\Omega(f)$-local for G', a', and b' with $D(a', b') = \Omega(f)$.

3. Using $\phi'(y)$, construct a small constant-depth Boolean circuit computing parity on $\Omega(f)$ bits.

4. For some c depending on ϕ, allowing $f = \Omega((\log n)^c)$ contradicts known lower bounds. ∎

Background – Circuit Complexity

Fact

Let $\phi(x)$ be an Arb-invariant FO *formula. There exists $d \in \mathbb{N}$ and a Boolean circuit family $(C_n)_{n\in\mathbb{N}}$ with depth d and size poly(n) such that for each graph G of size n, and vertex a,*

$$C_n(G, a) = 1 \textit{ iff } G \models \phi(a).$$

Lemma (Håstad)

For each $d \in \mathbb{N}$ and large enough m there is no Boolean circuit with depth d and size $2^{km^{1/(d-1)}}$ computing parity on m bits.

Disjoint Case

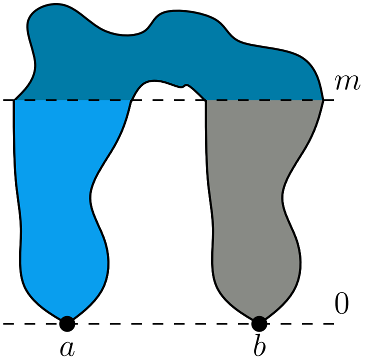
Lemma

Let $\phi(x)$ be an Arb-invariant FO formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:



Disjoint Case

Lemma

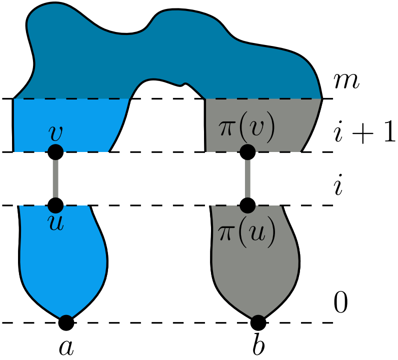
Let $\phi(x)$ be an Arb-invariant FO *formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.*

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.



Disjoint Case

Lemma

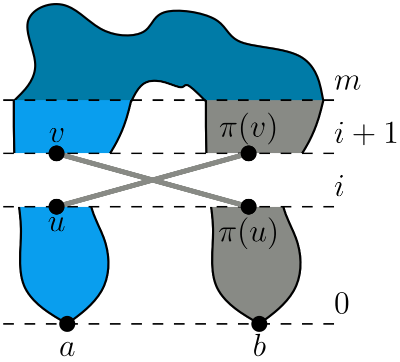
Let $\phi(x)$ be an Arb-invariant FO formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.



Disjoint Case

Lemma

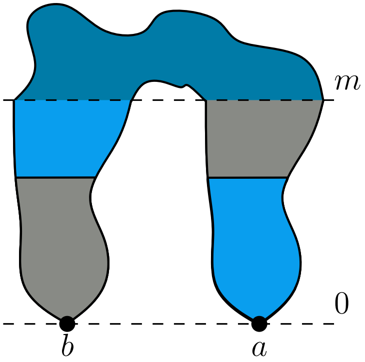
Let $\phi(x)$ be an Arb-invariant FO *formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.*

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.



Disjoint Case

Lemma

Let $\phi(x)$ be an Arb-invariant FO formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.

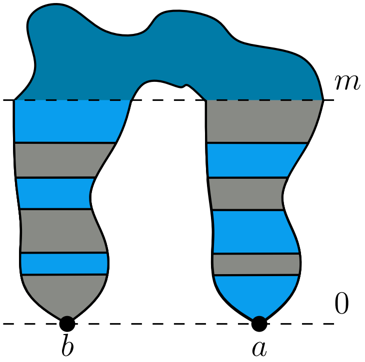
Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.

The resulting graph $G_w \cong G$.



Disjoint Case

Lemma

Let $\phi(x)$ be an Arb-invariant FO *formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.*

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

> *Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.*

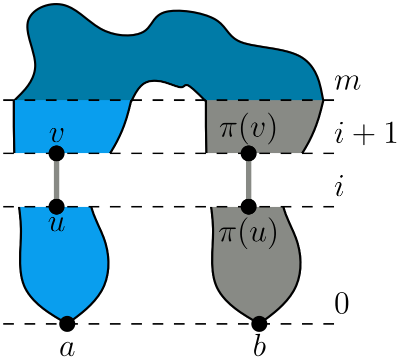
The resulting graph $G_w \cong G$.

$$(G_w, a) \cong \begin{cases} (G, a), & \bigoplus w = 0 \\ (G, b), & \bigoplus w = 1 \end{cases}$$

$\phi(x)$ distinguishes these cases.

A small circuit computes $\phi(x)$.

A small circuit computes $\bigoplus w$. ∎



Disjoint Case

Lemma

Let $\phi(x)$ be an Arb-invariant FO *formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.*

Proof Sketch.

Consider $w \in \{0,1\}^m$.

For $i \in \{0, 1, \ldots, m-1\}$ with $w_i = 1$:

Swap the endpoints of the edges leaving $B_i(a)$ with the corresponding endpoints of the edges leaving $B_i(b)$.

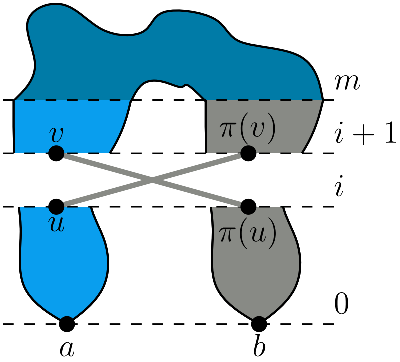
The resulting graph $G_w \cong G$.

$$(G_w, a) \cong \begin{cases} (G, a), & \bigoplus w = 0 \\ (G, b), & \bigoplus w = 1 \end{cases}$$

$\phi(x)$ distinguishes these cases.

A small circuit computes $\phi(x)$.

A small circuit computes $\bigoplus w$. ∎



Disjoint Case

Lemma

Let $\phi(x)$ be an Arb-invariant FO *formula where $G \models \phi(a) \wedge \neg\phi(b)$, $\mathcal{N}_m(a) \cong \mathcal{N}_m(b)$, and $D(a,b) > 2m$. For $d \in \mathbb{N}$ depending on ϕ there is a Boolean circuit of depth d and size poly(n) computing parity on m bits.*

Lemma (Håstad)

For each $d \in \mathbb{N}$ and large enough m there is no Boolean circuit with depth d and size $2^{km^{1/(d-1)}}$ computing parity on m bits.

Selecting $m(n) = f = \Omega((\log n)^c)$ for $c > (d-1)$ induces a contradiction.

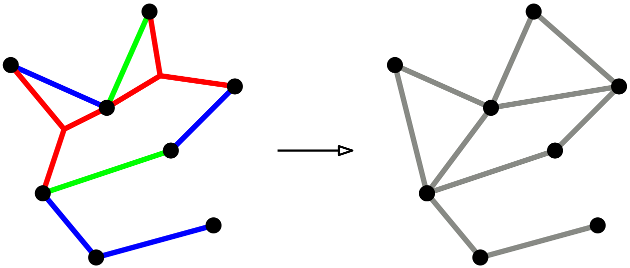
Extensions

1. unary $\longrightarrow$ k-ary:

Lemma (informal)

Let $\phi(\boldsymbol{x})$ be a k-ary Arb-invariant FO *formula that is not Gaifman f-local. For some $k' < k$, there is a k'-ary Arb-invariant* FO *formula $\phi'(\boldsymbol{y})$, that is not Gaifman $\Omega(f)$-local.*

2. Graphs $\longrightarrow$ Structures:



Measure distance on the Gaifman graph of the structure.

Summary

Theorem (Gaifman Locality)

1. *For each Arb-invariant* FO *formula* $\phi(\boldsymbol{x})$ *there is a* $c \in \mathbb{N}$ *such that the formula is Gaifman* $(\log n)^c$*-local.*
2. *For each* $c \in \mathbb{N}$ *there is an Arb-invariant* FO *formula* $\phi(\boldsymbol{x})$ *that is not Gaifman* $(\log n)^c$*-local.*

Theorem (Hanf Locality)

1. *For each Arb-invariant* FO *formula over strings there is a* $c \in \mathbb{N}$ *such that the formula is Hanf* $(\log n)^c$*-local.*
2. *For each* $c \in \mathbb{N}$ *there is an Arb-invariant* FO *formula over strings that is not Hanf* $(\log n)^c$*-local.*