Game-based Notions of Locality

Marcelo Arenas Pablo Barceló Leonid Libkin
University of Toronto
{marenas, pablo, libkin}@cs.toronto.edu

Abstract

Locality is a very useful tool in finite model theory, that allows one to prove inexpressibility results by avoiding complicated combinatorial reasoning typically associated with proofs involving Ehrenfeucht-Fraïssé games. It is applicable to first-order logic (FO) and some of its rather powerful counting extensions. In this short report on an on-going research project, we present a more general – and perhaps a more natural – notion of locality, and give a few examples of its applicability.

1 Introduction and motivation

The idea of locality in finite model theory is as follows. Suppose we have a formula $\phi(\vec{x})$ in some logic, a structure $\mathfrak{A}$, and two tuples $\vec{a}$ and $\vec{b}$. We want to show that ϕ cannot distinguish $\vec{a}$ from $\vec{b}$. For that, it suffices to establish that for some $r \geq 0$, the r-neighborhoods around $\vec{a}$ and $\vec{b}$ are isomorphic.

A typical example of using locality (assuming we proved such a property for a logic) is inexpressibility of the transitive closure. Assume, to the contrary, that the transitive closure is expressible by a formula $\phi(x, y)$. Pick r as above, and consider a "long" chain (successor relation), with two points a and b at distance at least $2r + 1$ from each other and the endpoints. Then, the r-neighborhoods of (a, b) and (b, a) are isomorphic, and hence ϕ cannot express the transitive closure, which clearly sees the difference between (a, b) and (b, a).

To formulate this precisely, we need the notion of the *Gaifman graph.* Given a structure $\mathfrak{A}$ of a relational vocabulary with universe A, its Gaifman graph is a graph whose vertices are the elements of A and where two nodes a and b are adjacent if there is a tuple in some relation of $\mathfrak{A}$ containing both a and b. The distance between points a and b, denoted by $d(a, b)$, is the distance in the Gaifman graph (if $a = b$ we assume that $d(a, b) = 0$). We define $d((a_1, \ldots, a_n), b) = \min_i d(a_i, b)$, and let $B_r(\vec{a}) = \{b \mid d(\vec{a}, b) \leq r\}$. The *$r$-neighborhood of $\vec{a}$*, $N_r(\vec{a})$ is the substructure of $\mathfrak{A}$ whose universe is $B_r(\vec{a})$, extended with constants interpreted as $\vec{a}$. That is, for $N_r(\vec{a})$ and $N_r(\vec{b})$ to be isomorphic (written $N_r(\vec{a}) \cong N_r(\vec{b})$), we must have an isomorphism h such that $h(\vec{a}) = \vec{b}$.

Suppose we have an *m-ary query*, that is, a mapping Q that associates with each structure $\mathfrak{A}$ a subset of A^m. We say that Q is *Gaifman-local* [4] if there exists a number r such that for any structure $\mathfrak{A}$ and any $\vec{a}, \vec{b} \in A^m$,

$$N_r(\vec{a}) \cong N_r(\vec{b}) \implies \vec{a} \in Q(\mathfrak{A}) \text{ iff } \vec{b} \in Q(\mathfrak{A}).$$

Intuitively, if Q is a local query and two tuples look alike in some neighborhood, then they are indistinguishable by Q.

Gaifman's locality theorem [2] implies that every FO-definable query is Gaifman-local. More recently [4, 5] it was shown that Gaifman-locality extends to logics with very powerful counting mechanisms. While Gaifman-locality is very useful for proving expressivity bounds, its limitations are well understood [4]; in particular, it does not apply to the extension of FO with counting quantifiers *and* a linear order.

Instead of giving up hope of proving new bounds by locality beyond those currently known, we suggest a re-examination of the notion. The intuition behind locality is as follows: "if the neighborhoods of $\vec{a}$ and $\vec{b}$ look the same, then a formula ϕ cannot distinguish $(\mathfrak{A}, \vec{a})$ and $(\mathfrak{A}, \vec{b})$". However, our notion of "look the same" is extremely strong: it says that the neighborhoods must be isomorphic. We use this strong notion to derive that the difference between $\vec{a}$ and $\vec{b}$ cannot be observed in some logic. It appears more natural to use a symmetric definition of the form: "if the difference of neighborhoods of $\vec{a}$ and $\vec{b}$ cannot be observed in a logic, then a formula ϕ cannot distinguish $(\mathfrak{A}, \vec{a})$ and $(\mathfrak{A}, \vec{b})$". This is the relaxation of the notion of locality we propose. We define it in the next section, and prove a few initial results about it.

2 Locality under games

For many logics, elementary equivalence can be characterized by Ehrenfeucht-Fraïssé games: if the duplicator (player II) has a winning strategy in the k-round game on $\mathfrak{A}$ and $\mathfrak{B}$, then $\mathfrak{A}$ and $\mathfrak{B}$ agree on all formulae of quantifier rank up to k. The idea of the new notion of locality is to change isomorphism by the existence of a winning strategy for the duplicator. We present the definition and the initial results for the case of FO and the usual Ehrenfeucht-Fraïssé game, and for the case of an infinitary counting logic characterized by *bijective* games [3].

Let $\mathfrak{A} \equiv_k \mathfrak{B}$ mean that the duplicator has a winning strategy in the k-round Ehrenfeucht-Fraïssé game on $\mathfrak{A}$ and $\mathfrak{B}$. We say that an m-ary query Q is *Gaifman-local under games* if there exist numbers $r, k \geq 0$ such that for any structure $\mathfrak{A}$ and any $\vec{a}, \vec{b} \in A^m$,

$$N_r(\vec{a}) \equiv_k N_r(\vec{b}) \implies \vec{a} \in Q(\mathfrak{A}) \text{ iff } \vec{b} \in Q(\mathfrak{A}).$$

That every FO-definable query is Gaifman-local under games is a corollary of Gaifman's theorem [2]. However, the standard Gaifman-locality for FO can be shown by a simple inductive argument, which extends to other logics [4]. We thus looked for a similar direct proof of this new notion of locality for FO, and we found it. We established the following:

Proposition 1 *If Q is definable by an* FO *formula ϕ, then Q is Gaifman-local under games, and r is $O(4^q)$, where q is the quantifier rank of ϕ.*

This proposition is proved by an inductive argument, using Ehrenfeucht-Fraïssé games.

Example. Let $\mathfrak{A}$ contain an equivalence relation $\sim$, together with a graph R on the equivalence classes. Suppose all $\sim$-classes have different cardinalities. How can we prove that the transitive closure of R is not expressible? The usual locality does not help, since for a, b with isomorphic r-neighborhoods, $r > 0$, we must have $a \sim b$. The new notion, however, is easily applicable, and implies the result.

We now move to a different logic, $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$. It adds, for every k and m, a quantifier $\exists^k \vec{x} \phi(\vec{x}, \cdot)$, stating that there are at least k m-tuples $\vec{x}$ such that $\phi(\vec{x}, \cdot)$ is true. Furthermore, it adds infinitary connectives $\bigvee$ and $\bigwedge$, but only allows formulae of *finite* quantifier rank.

This logic is known to be Gaifman-local [5], and it is captured by bijective Ehrenfeucht-Fraïssé games [3]: in such a game, before each round, the duplicator selects a bijection $f : \mathfrak{A} \to \mathfrak{B}$, and if the spoiler plays $a \in \mathfrak{A}$, the duplicator responds by $f(a) \in \mathfrak{B}$. It is known that $\mathfrak{A}$ and $\mathfrak{B}$ agree on all $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$ sentences of quantifier rank up to k iff the duplicator wins the k-round bijective game, which we denote by $\mathfrak{A} \equiv^{bij}_k \mathfrak{B}$.

We now say that an m-ary query Q is *Gaifman-local under bijective games* if there exist $r, k \geq 0$ such that for any structure $\mathfrak{A}$ and any $\vec{a}, \vec{b} \in A^m$,

$$N_r(\vec{a}) \equiv^{bij}_k N_r(\vec{b}) \implies \vec{a} \in Q(\mathfrak{A}) \text{ iff } \vec{b} \in Q(\mathfrak{A}).$$

Proposition 2 *Every $\mathcal{L}^*_{\infty\omega}(\mathbf{C})$-definable query is Gaifman-local under bijective games, and r is $O(2^q)$, where q is the quantifier rank of ϕ.*

3 Ongoing Work

- Propositions 1 and 2 suggest that there might be a general result saying that if $N_r(\vec{a})$ and $N_r(\vec{b})$ are indistinguishable for k rounds of a game for a local logic $\mathcal{L}$, then $(\mathfrak{A}, \vec{a})$ and $(\mathfrak{B}, \vec{b})$ are indistinguishable in l rounds of the game. The two proofs we have, however, are quite different, and we do not yet know how far we can push the results.
- For the FO case, our bound on r is better than Gaifman's $O(7^q)$ and matches the bound of Lifsches and Shelah [6]. For the standard notion of Gaifman-locality, an $O(2^q)$ bound is known [5] (and it matches the lower bound). We would like to bridge the gap between 2^q and 4^q.

4 Remarks

Other notions of locality known in the literature are Hanf-locality and its variation, threshold equivalence [1]. The latter is only valid for structures of small degrees, where the win for the duplicator with sufficiently many moves implies isomorphism of neighborhoods. For Hanf-locality, we can show that there are FO-definable queries that violate the natural extension of the notion that uses games instead of isomorphism.

References

[1] R. Fagin, L. Stockmeyer, and M. Vardi. On monadic NP vs. monadic co-NP. *Information and Computation*, 120(1):78–92, 1995.

[2] H. Gaifman. On local and non-local properties. In *Proceedings of the Herbrand Symposium, Logic Colloquium*, pages 105–135, 1981.

[3] L. Hella. Logical hierarchies in PTIME. *Information and Computation*, 129 (1996), 1–19.

[4] L. Hella, L. Libkin, and J. Nurmonen. Notions of locality and their logical characterizations over finite models. *Journal of Symbolic Logic*, 64:1751–1773, 1999.

[5] L. Libkin. Logics with counting and local properties. *ACM TOCL*, 1(1):33–59, 2000.

[6] S. Lifsches and S. Shelah. Distorted sums of models. Unpublished manuscript, 1996.