Logical Complexity of Graphs

Oleg Verbitsky

Combinatorics Seminar, University of Warwick
November 23, 2012

Outline

1. Logical depth, width, and length of a graph
2. Relevance to Graph Isomorphism
3. Bounds for particular classes of graphs
4. General bounds
5. Random graphs
6. How succinct are the most succinct definitions?

Outline

Our language

Vocabulary:

$=$ equality of vertices

$\sim$ adjacency of vertices

First-order logic: quantification over vertices;
no quantification over sets.

Example: We can say that vertices x any y lie at distance no more than n:

$$\Delta_1(x, y) \stackrel{\text{def}}{=} x \sim y \vee x = y$$

$$\Delta_n(x, y) \stackrel{\text{def}}{=} \exists z_1 \ldots \exists z_{n-1} \Big(\Delta_1(x, z_1) \wedge \Delta_1(z_1, z_2) \wedge \ldots \wedge \Delta_1(z_{n-2}, z_{n-1}) \wedge \Delta_1(z_{n-1}, y) \Big)$$

Succinctness measures of a formula Φ

Definition

The *width* $W(\Phi)$ is the number of variables used in Φ (different occurrences of the same variable are not counted).

Example: $W(\Delta_n) = n + 1$ but we can economize by recycling just three variables:

$$\Delta'_1(x, y) \stackrel{\text{def}}{=} \Delta_1(x, y)$$

$$\Delta'_n(x, y) \stackrel{\text{def}}{=} \exists z(\Delta'_1(x, z) \wedge \Delta'_{n-1}(z, y)),$$

where $\Delta'_{n-1}(z, y) = \exists x(\ldots)$ getting $W(\Delta'_n) = 3$.

Succinctness measures of a formula Φ

Definition

The *depth* $D(\Phi)$ (or *quantifier rank*) is the maximum number of nested quantifiers in Φ.

Example: $D(\Delta'_n) = n - 1$ but we can economize using the halving strategy:

$$\Delta''_1(x,y) \stackrel{\text{def}}{=} \Delta_1(x,y)$$

$$\Delta''_n(x,y) \stackrel{\text{def}}{=} \exists z \left(\Delta''_{\lfloor n/2 \rfloor}(x,z) \wedge \Delta''_{\lceil n/2 \rceil}(z,y) \right),$$

getting $D(\Delta''_n) = \lceil \log n \rceil$ while keeping $W(\Delta''_n) = 3$.

Succinctness measures of a formula Φ

Definition

The *length* $L(\Phi)$ is the total number of symbols in Φ (each variable symbol contributes 1).

Example: $L(\Delta_n) = O(n)$ and $L(\Delta''_n) = O(n)$ but we can economize

$$\Delta'''_{2n+1}(x,y) \overset{\text{def}}{=} \exists z\,(\Delta_1(x,z) \wedge \Delta_{2n}(z,y))$$

$$\Delta'''_{2n}(x,y) \overset{\text{def}}{=} \exists z \forall u \Big(u = x \vee u = y \to \Delta'''_n(u,z)\Big),$$

getting $L(\Delta'''_n) = O(\log n)$ and still keeping $D(\Delta'''_n) \le 2\log n$ and $W(\Delta'''_n) = 4$.

Definition

A statement *Φ defines a graph G* if Φ is true on G but false on every non-isomorphic graph H.

Example: P_n, the path on n vertices, is defined by

$$\forall x \forall y \Delta_{n-1}(x,y) \wedge \neg \forall x \forall y \Delta_{n-2}(x,y)$$

`% diameter = n-1`

$$\wedge \forall x \forall y_1 \forall y_2 \forall y_3 (x \sim y_1 \wedge x \sim y_2 \wedge x \sim y_3 \rightarrow y_1 = y_2 \vee y_2 = y_3 \vee y_3 = y_1)$$

`% max degree < 3`

$$\wedge \exists x \exists y \forall z \big(x \sim y \wedge (z \sim x \rightarrow z = y) \big)$$

`% min degree = 1`

The logical length, depth, and width of a graph

Definition

$L(G)$ (resp. $D(G)$, $W(G)$) is the minimum $L(\Phi)$ (resp. $D(\Phi)$, $W(\Phi)$) over all Φ defining G.

Remark

$W(G) \le D(G) < L(G)$

Theorem (Pikhurko, Spencer, V. 06)

$L(G) < Tower(D(G) + \log^* D(G) + 2)$. *This bound is tight in the sence that* $L(G) \ge Tower(D(G) - 7)$ *for infinitely many* G.

Example (a path)

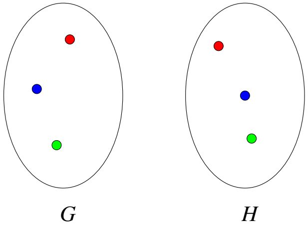
- $W(P_n) \leq 4$ (in fact, $W(P_n) = 3$ if $n \geq 2$)
- $D(P_n) < \log n + 3$ (and $D(P_n) \geq \log n - 2$).

How to determine $W(G)$ or $D(G)$?

1. $D(G) = \max_{H \not\cong G} D(G, H)$, where $D(G, H)$ is the minimum quantifier depth needed to distinguish between G and H. Similarly for $W(G)$.
2. $D(G, H)$ and $W(G, H)$ are characterized in terms of a combinatorial game.

The Ehrenfeucht game

Barwise; Immerman 82; Poizat 82: *G* and *H* are distinguishable with *k* variables and quantifier depth *r* iff Spoiler wins the *k*-pebble Ehrenfeucht game in *r* rounds.



Rules of the game

Players: Spoiler and Duplicator
Resources: *k* pebbles,
each in duplicate

A round:

Spoiler puts a pebble on a vertex in *G* or *H*.
Duplicator puts the other copy in the other graph.
Duplicator's objective: after each round the pebbling should determine a partial isomorphism between *G* and *H*.

Example (a path)

$L(P_n) = O(\log n)$

Remark: This is tight up to a multiplicative constant because $L(P_n) > D(P_n) \geq \log n - 2$.

Variations of logic: bounded number of variables

$D^k(G)$ denotes the logical depth of G in the k-variable logic (assuming $W(G) \le k$).

Example (a path)

- $D^3(P_n) \le \log n + 3$
- $L^4(P_n) = O(\log n)$

Theorem (Grohe, Schweikardt 05)

$L^3(P_n) > \sqrt{n}$

Variations of logic: counting quantifiers

$\exists^m x\Psi(x)$ means that there are at least m vertices x having property Ψ.

The counting quantifier $\exists^m$ contributes 1 in the quantifier depth whatever m.

$D_{\#}(G)$ and $W_{\#}(G)$ denote the logical depth and width of a graph G in the counting logic.

$D^k_{\#}(G)$ denotes the variant of $D^k(G)$ for the k-variable counting logic.

Counting move in the Ehrenfeucht game

- Spoiler exhibits a set $A \subset V(G)$ of “good” vertices.
- Duplicator responds with $B \subset V(H)$ such that $|B| = |A|$.
- Spoiler selects $b \in B$ and puts a pebble on it.
- Duplicator selects $a \in A$ and puts the other pebble on it.

Power of counting

Example

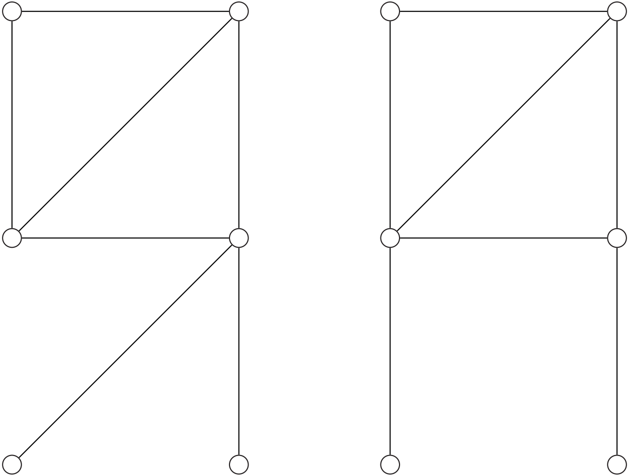
$W_{\#}(K_n) = 2$ while $W(K_n) = n + 1$.

Question

Is it true that $W(G) = O(W_{\#}(G) \log n)$ *if* G *is asymmetric, i.e., has no nontrivial automorphism?*

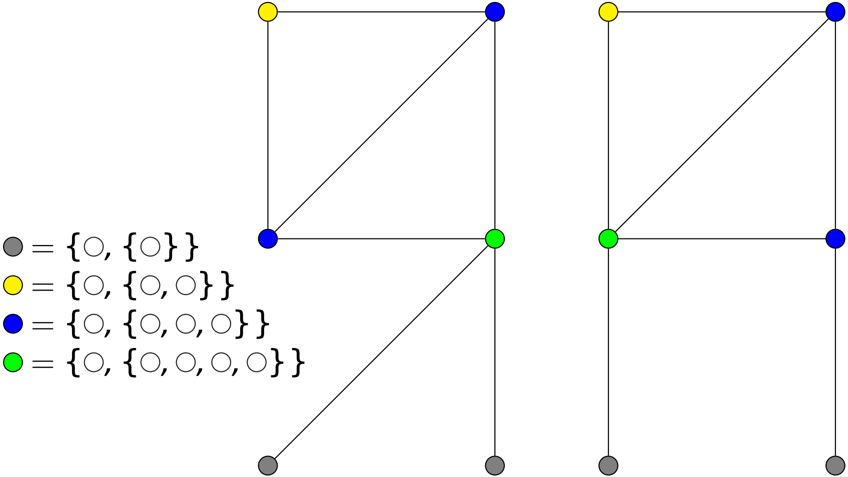
Outline

Color refinement algorithm



Initial coloring is monochromatic.

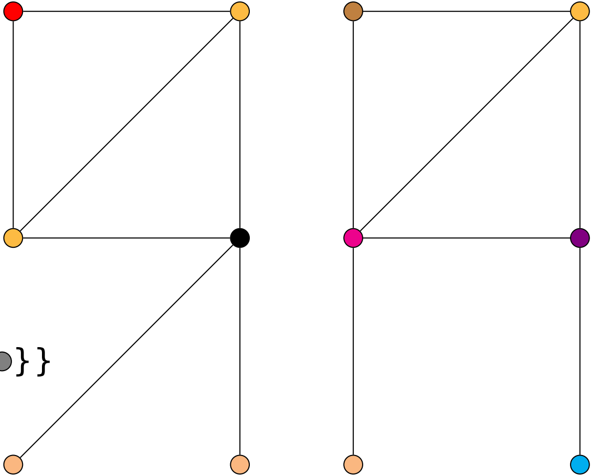
Color refinement



New color of a vertex =

old color + old colors of all neighbours.

Next refinement



the graphs are non-isomorphic

k-dimensional Weisfeiler-Lehman algorithm

- 1-dim WL $=$ the color refinement algorithm
- k-dim WL colors $V(G)^k$
- Initial coloring: $C^1(\bar{u})$ $=$ the equality type of $\bar{u} \in V(G)^k$ and the isomorphism type of the spanned subgraph
- Color refinement: $C^i(\bar{u}) = \{C^{i-1}(\bar{u}), \{(C^{i-1}(\bar{u}^{1,x}), \ldots, C^{i-1}(\bar{u}^{k,x}))\}_{x \in V}\}$, where $(u_1, \ldots, u_i, \ldots, u_k)^{i,x} = (u_1, \ldots, x, \ldots, u_k)$

The Weisfeiler-Lehman algorithm

- purports to decide if input graphs G and H are isomorphic,
  - If $G \cong H$, the output is correct.
  - If $G \ncong H$, the output can be wrong.
- has two parameters: *dimension* and *number of rounds*.

Theorem (Cai, Fürer, Immerman 92)

The r-round k-dim WL works correctly on any pair (G, H) if

$$k = W_{\#}(G) - 1 \text{ and } r = D_{\#}^{k+1}(G) - 1.$$

On the other hand, it is wrong on (G, H) for some H if

$$k < W_{\#}(G) - 1, \text{ whatever } r.$$

The Weisfeiler-Lehman algorithm

Theorem (Cai, Fürer, Immerman 92)

Let C be a class of graphs G with $W_{\#}(G) \leq k$ for a constant k. Then Graph Isomorphism for C is solvable in P.

Theorem (Grohe, V. 06)

1. *Let C be a class of graphs G with*
$$D^k_{\#}(G) = O(\log n).$$
Then Graph Isomorphism for C is solvable in $\mathrm{TC}^1 \subseteq \mathrm{NC}^2 \subseteq \mathrm{AC}^2$.
2. *Let C be a class of graphs G with*
$$D^k(G) = O(\log n).$$
Then Graph Isomorphism for C is solvable in $\mathrm{AC}^1 \subseteq \mathrm{TC}^1$.

Outline

Trees

Theorem (Immerman, Lander 90)

$W_{\#}(T) \le 2$ for every tree T.

Remark: $D^2_{\#}(P_n) = \frac{n}{2} - O(1)$

Speed-up: an extra variable $\mapsto$ logarithmic depth

Theorem

If T is a tree on n vertices, then $D^3_{\#}(T) \le 3\log n + 2$.

Proof-sketch

We can easily distinguish between T and $T' \not\cong T$ if T'

- is disconnected;
- has different number of vertices;
- has the same number of vertices, is connected but has a cycle;
- has larger maximum degree.

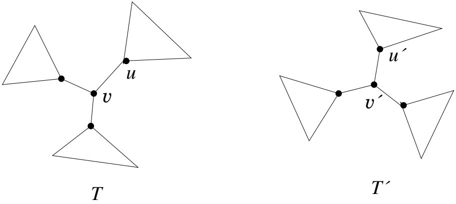
It remains the case that T' is a tree with the same maximum degree. For simplicity, assume that the maximum degree is 3 (then no counting quantifiers are needed).

Proof cont'd (a separator strategy)

We need to show that Spoiler wins the 3-pebble game on T and T' in $3 \log n + 2$ moves.

Step 1. Spoiler pebbles a separator v in T (every component of $T - v$ has $\le n/2$ vertices).

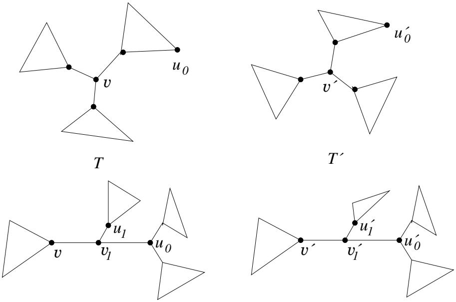
Step 2. Spoiler ensures pebbling $u \in N(v)$ and $u' \in N(v')$ so that the corresponding components are non-isomorphic rooted trees.



Spoiler forces further play on these components and applies the same strategy again.

Proof cont'd

A complication: the strategy is now applied to a graph with one vertex pebbled and we may need more than 3 pebbles. Assume that u_0 and u'_0 were pebbled earlier and $T - v$ and $T' - v'$ differ only by the components containing u_0 and u'_0. Suppose that $d(v, u_0) = d(v', u'_0)$.



Step 3. Spoiler pebbles v_1 in the v-u_0-path such that $T - v_1$ and $T' - v'_1$ differ by components with no pebble (assuming that $d(v, v_1) = d(v', v'_1)$).

Isomorphism of trees (history revision)

Theorem

If T is a tree on n vertices, then $D^3_\#(T) \le 3\log n + 2$.

Testing isomorphism of trees is

- in Log-Space *Lindell 92*
- in AC^1 *Miller-Reif 91*
- in AC^1 if $\Delta = O(\log n)$ *Ruzzo 81*
- in Lin-Time by 1-WL ($W_\#(T) = 2$) *Edmonds 65*

Miller and Reif [SIAM J. Comput. 91]: *"No polylogarithmic parallel algorithm was previously known for isomorphism of unbounded-degree trees."*

However, the $3\log n$-round 2-WL solves it in TC^1 and is known since 68 !

Graphs of bounded tree-width

Theorem

For a graph G of tree-width k on n vertices

$W_{\#}(G) \leq k + 2$ *[Grohe, Mariño 99];*

$D_{\#}^{4k+4}(G) < 2(k+1)\log n + 8k + 9$ *[Grohe, V. 06].*

Planar graphs

Theorem

For a planar graph G on n vertices

$W_{\#}(G) = O(1)$ *[Grohe 98].*

If G is, moreover, 3-connected, then

$D^{15}(G) < 11 \log n + 45$ *[V. 07].*

Interval graphs

Theorem

For an interval graph G on n vertices

$W_{\#}(G) \le 4$ *[Evdokimov et al. 00, Laubner 10];*

$D^{15}_{\#}(G) < 9 \log n + 8$ *[Köbler, Kuhnert, Laubner, V. 11].*

Our approach to interval graphs

- The *clique hypergraph* $\mathcal{C}(G)$ of a graph G has vertices as in G and the maxcliques in G as hyperedges.
- $G =$ the Gaifman graph of $\mathcal{C}(G)$.
- $G \cong$ the intersection graph of the dual $\mathcal{C}(G)^*$.
- Laubner 10: If G is interval, $\mathcal{C}(G)^*$ is constructible (definable) from G because any maxclique is then the common neighborhood of some two vertices.
- If G is interval, any minimal interval model of G is isomorphic to $\mathcal{C}(G)^*$; hence, $\mathcal{C}(G)^*$ is an interval hypergraph.
- Then $\mathcal{C}(G)^*$ is decomposable into a tree, known in algorithmics as *PQ-tree*.

Circular-arc graphs

Question

Is the bound $W_{\#}(G) = O(1)$ true for circular-arc graphs?

The approach used for interval graphs fails because circular-arc graphs can have exponentially many maxcliques.

In fact, the status of the isomorphism problem for circular-arc graphs is open.
Curtis et al. [arXiv, March 12] found a bug in the only known Hsu's algorithm.

Graphs with an excluded minor

Theorem (Grohe 11)

For each H, if G excludes H as a minor, then

$$W_{\#}(G) = O(1).$$

Question

Is it then true that $D^k_{\#}(G) = O(\log n)$ for some constant k?

Theorem (a version of Dawar, Lindell, Weinstein 95)

If $W(G) \leq k$*, then* $D^k(G) < n^{k-1} + k$*.*

Question

How tight is this bound?

We have $D^k(G) = O(\log n)$ *or* $D^k_{\#}(G) = O(\log n)$ *for some classes of graphs. Can one formulate some general conditions under which this is true?*

Outline

Remark

Every finite graph G is definable

by the following generic formula:

$$\exists x_1 \dots \exists x_n \Big(\mathrm{Distinct}(x_1, \dots, x_n) \\ \wedge \mathrm{Adj}(x_1, \dots, x_n) \Big) \\ \wedge \ \forall x_1 \dots \forall x_{n+1} \neg \mathrm{Distinct}(x_1, \dots, x_{n+1})$$

Thus, for any G on n vertices

$$W(G) \leq D(G) \leq n+1, \qquad L(G) = O(n^2)$$

Bad news: $W(K_n) = n+1$

Very bad news:

Theorem (Cai, Fürer, Immerman 92)

There are graphs on n vertices, even of maximum degree 3, such that

$$W_{\#}(G) > 0.004\,n.$$

Any good news? Well,...

Exercise: $D(G) \leq n$ for all G on n vertices except K_n and $\overline{K_n}$.

Exercise: $D(G) \leq n-1$ for all G on n vertices except $K_n, \overline{K_n}, K_{1,n-1}, \overline{K_{1,n-1}}, \ldots$, altogether 10 exceptional graphs (each having at least $n-2$ *twins*).

Definition

Two vertices are *twins* if they are both adjacent or both non-adjacent to any third vertex.

Theorem (Pikhurko, Veith, V. 06)

For a graph G on n vertices, it is easy to recognize whether or not

$$D(G) > n - t,$$

as long as $t \leq \frac{n-5}{2}$.

Theorem (Pikhurko, Veith, V. 06)

If G is a twin-free graph on n vertices, then

$$D(G) \le \frac{n+5}{2}.$$

Definition. Let $X \subset V(G)$ and $y \notin X$. The set X *sifts out* y if $N(y) \cap X \neq N(z) \cap X$ for any other $z \notin X$.
$S(X)$ consists of X and all y sifted out by X.
X is a *sieve* if $S(X) = V(G)$.
X is a *weak sieve* if $S(S(X)) = V(G)$.

Exercise 1. Let $G \not\cong H$. If X is a sieve in G, then Spoiler wins the Ehrenfeucht game on G and H in $|X| + 2$ moves.

Exercise 2. If X is a weak sieve in G, then Spoiler wins the Ehrenfeucht game on G and H in $|X| + 3$ moves.

Exercise 3. Any twin-free graph G on n vertices has a weak sieve X with $|X| \le (n-1)/2$.

By a similar argument:

Theorem (Pikhurko, Veith, V. 06)

Any two non-isomorphic G and H on n vertices can be distinguished by a statement of quantifier depth at most $\frac{n+3}{2}$.

Corollary

$D_{\#}(G) \leq \frac{1}{2}n + 3$ *for any G on n vertices.*

Question

$W_{\#}(G) \leq (\frac{1}{2} - \epsilon)n$ *for any G on n vertices?*

Question

$W_{\#}(G) = o(n)$ *for any asymmetric G on n vertices??*

Outline

Random graphs (counting logic)

Theorem (Babai, Erdős, Selkow 80)

With probability more than $1 - 1/\sqrt[7]{n}$, the 1-dim 3-round WL works correctly on a random graph $G_{n,1/2}$ and all H. Therefore,

$$D^2_{\#}(G_{n,1/2}) \le 4$$

with this probability.

Theorem

With high probability,

$$D^2_{\#}(G_{n,1/2}) = 4 \ \text{ and } \ 3 \le D_{\#}(G_{n,1/2}) \le 4$$

Question

What is the typical value of $D_{\#}(G_{n,1/2})$?

Random graphs (no counting)

Theorem (Kim, Pikhurko, Spencer, V. 05)

With high probability

$$\log n - 2\log\log n + 1 < W(G_{n,1/2}) \le D(G_{n,1/2}) \le \log n - \log\log n + \omega,$$

for each (arbitrarily slowly) increasing function $\omega = \omega(n)$.

Theorem (Kim, Pikhurko, Spencer, V. 05)

For infinitely many n

$$D(G_{n,1/2}) \le \log n - 2\log\log n + 5 + \log\log \mathrm{e} + o(1)$$

with high probability.

An application to the 0-1-law

Let $p_n(\Phi) = \mathbb{P}\left[G_{n,1/2} \models \Phi\right]$.

Theorem (Glebskii et al. 69, Fagin 76)

$p_n(\Phi) \to p(\Phi)$ as $n \to \infty$, where $p(\Phi) \in \{0, 1\}$.

Define the *convergence rate function* by

$$R(k, n) = \max_{\Phi} \left\{ |p_n(\Phi) - p(\Phi)| : D(\Phi) \le k \right\}.$$

Thus, $R(k, n) \to 0$ as $n \to \infty$ for any fixed k.

Theorem

Let $k(n) = \log n - 2 \log\log n + c$.

1. *Set $c = 1$. Then $R(k(n), n) \to 0$ as $n \to \infty$.*
2. *The claim does not hold true for $c = 6$.*

Remark

With high probability,

$$\Omega\left(\frac{n^2}{\log n}\right) \le L(G_{n,1/2}) \le O(n^2).$$

Question

Where is $L(G_{n,1/2})$ concentrated?

Outline

The most succinct definitions

Definition (succinctness function)

$s(n) = \min \{ D(G) : G \text{ has } n \text{ vertices} \}$

$s(n) \to \infty$ as $n \to \infty$ but its values can be inconceivably small if compared to n.

Theorem (Pikhurko, Spencer, V. 06)

There is no total recursive function f such that

$$f(s(n)) \geq n \quad \text{for all} \quad n.$$

Nevertheless ...

Definition (smoothed succinctness function)

$s^*(n) = \max_{m \le n} s(m)$, the least monotone nondecreasing function bounding $s(n)$ from above.

Theorem (Pikhurko, Spencer, V. 06)

$\log^* n - \log^* \log^* n - 2 \le s^*(n) \le \log^* n + 4$

Succinctness function over trees

Let $t(n) = \min \{ D(T) : T \text{ is a tree on } n \text{ vertices} \}$.

Theorem (Pikhurko, Spencer, V. 06)

$\log^* n - \log^* \log^* n - 4 \le t(n) \le \log^* n + 4$

Theorem (Dawar, Grohe, Kreutzer, Schweikardt 07)

For infinitely many n, there is a tree T on n vertices with $L(T) = O((\log^* n)^4)$.

Conjecture. The first-order theory of a class of graphs $\mathcal{C}$ is decidable iff the succinctness function over $\mathcal{C}$ admits a total recursive lower bound.

A more detailed exposition can be found in:

O. Pikhurko and O. Verbitsky.
Logical complexity of graphs: a survey.
In: *Model Theoretic Methods in Finite Combinatorics*, J. Makowsky and M. Grohe Eds. Contemporary Mathematics, vol. 558, Amer. Math. Soc., Providence, RI, pp. 129–179 (2011).