The first order definability of finite graphs

Oleg Verbitsky

Humboldt Universität Berlin, Germany and IAPMM Lviv, Ukraine

Bertinoro, October 2009

Based on joint work with
Oleg Pikhurko and Joel Spencer,
with important contributions by
Tom Bohman, Alan Frieze, Martin Grohe, Jeong Han Kim,
Tomasz Łuczak, Clifford Smyth, and Helmut Veith.

Outline

- Basic concepts
  (the logical depth, width, and length of a graph)
- Main tools
  (the Ehrenfeucht game and the Weisfeiler-Lehman algorithm)
- Upper bounds
  - trees and graphs with bounded treewidth
  - planar graphs
  - general case (Ramsey?)
- Random graph
  - applications to the 0-1 law
- Succinctly definable graphs
  - definitions with no quantifier alternation (Ramsey!)

Language of the first order theory of graphs

- variables (x, y, y_1, etc), ranging through the vertex set of a graph;
- the relations $=$ (equality) and $\sim$ (vertex adjacency);
- the quantifiers $\forall$ (universality) and $\exists$ (existence);
- the Boolean connectives $\wedge$ (and), $\vee$ (or), and $\neg$ (negation).

Example. The following first order formula $\Delta_n(x, y)$ says that vertices x any y lie at distance no more than n:

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

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

Basics

A sentence Φ distinguishes a graph G from another graph H if Φ is true on G but false on H.

Example.

1. The sentence $\forall x \forall y \, \Delta_1(x, y)$ distinguishes a complete graph K_n from any other graph H that is not complete.

2. The sentence $\forall x \forall y \, \Delta_{n-1}(x, y)$ distinguishes P_n, a path on n vertices, from any longer path P_m, $m > n$.

Basics

A sentence Φ defines a graph G (up to isomorphism) if
Φ distinguishes G from every non-isomorphic graph H.

Example. P_n is defined by

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

to say that the diameter $= n - 1$

$$\wedge \forall x \neg \exists y_1 \exists y_2 \exists y_3 \left(\bigwedge_{i=1,2,3} x \sim y_i \wedge \bigwedge_{i \neq j} y_i \neq y_j \right)$$

to say that the maximum degree ≤ 2

$$\wedge \exists x \neg \exists y_1 \exists y_2 \left(\bigwedge_{i=1,2} x \sim y_i \wedge y_1 \neq y_2 \right)$$

to say that the minimum degree ≤ 1 (thereby distinguishing from cycles C_{2n-2} and C_{2n-1})

Basics

Succinctness measures of a formula Φ:
the *length* $L(\Phi)$, the quantifier *depth* $D(\Phi)$, and the *width* $W(\Phi)$

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

Example.
$W(\Delta_n) = n + 1$. However, rewriting it as

$\Delta'_n(x, y) \stackrel{\text{def}}{=} \exists z(\Delta_1(x, z) \wedge \Delta'_{n-1}(z, y))$, where

$\Delta'_{n-1}(z, y) \stackrel{\text{def}}{=} \exists x(\Delta_1(z, x) \wedge \Delta'_{n-2}(x, y))$ and so on,

we get $W(\Delta'_n) = 3$.

Basics

Definition.
$D(\Phi)$, the quantifier depth of Φ, is the maximum number of nested quantifiers in Φ.

Example.
$D(\Delta_n) = n - 1$. However, rewriting it as

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

we get $D(\Delta''_n) = \log n + O(1)$.

Main Definition

Definition
(the logical length, depth, and width of a graph).
$L(G)$ (resp. $D(G)$, $W(G)$) is the minimum $L(\Phi)$ (resp. $D(\Phi)$, $W(\Phi)$) over all Φ defining G.

Example.
$W(P_n) \leq 4, \quad D(P_n) \leq \log n + O(1)$.

Basics

Definition.
Let $G \not\cong H$. Then $D(G,H)$ (resp. $W(G,H)$) is the minimum $D(\Phi)$ (resp. $W(\Phi)$) over all Φ distinguishing G from H.

Proposition.

1. $D(G) = \max_H D(G,H)$

2. $W(G) = \max_H W(G,H)$

Variations of a logic

Fragments of first order logic

- Bounded number of quantifier alternations (later).
- Bounded number of variables.
  $D^k(G)$ denotes the logical depth of G in the k-variable logic

An extension of first order logic: Counting quantifiers

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

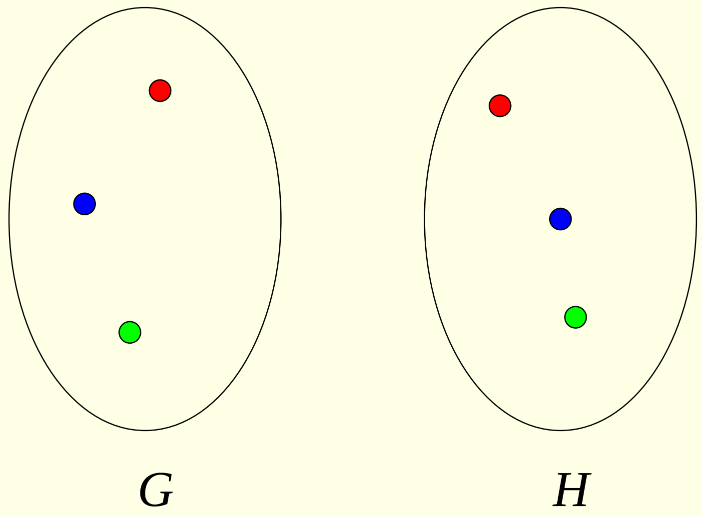
$D_{\#}(G)$ and $W_{\#}(G)$ will denote the logical depth and width of a graph G in the counting logic.

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

Ehrenfeucht's game

Immerman-Poizat: G and H are distinguishable with k variables and quantifier depth r iff Spoiler wins the Ehrenfeucht game with k pebbles in r moves.

Rules of the Game



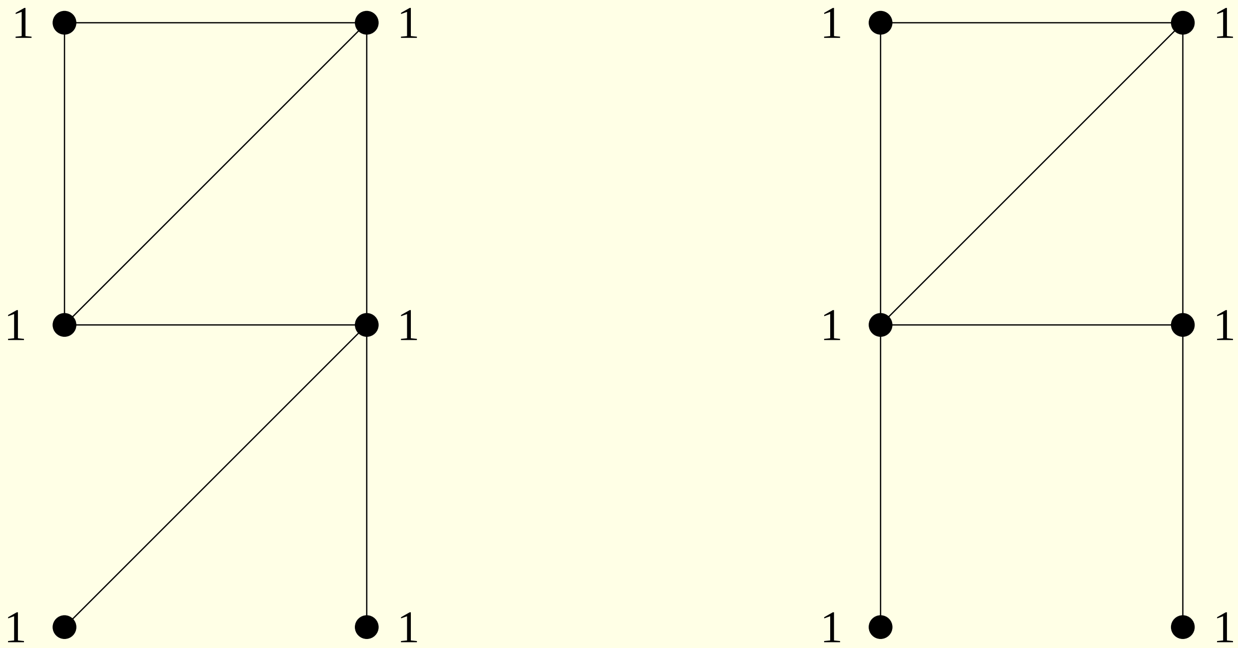
Players: Spoiler and Duplicator
Resources: k pebbles, each in duplicate

A round:
Spoiler puts a pebble on a vertex in G or H
Duplicator puts the other copy on a vertex in the other graph

Duplicator's objective: after each round the pebbling should determine a partial isomorphism between G and H

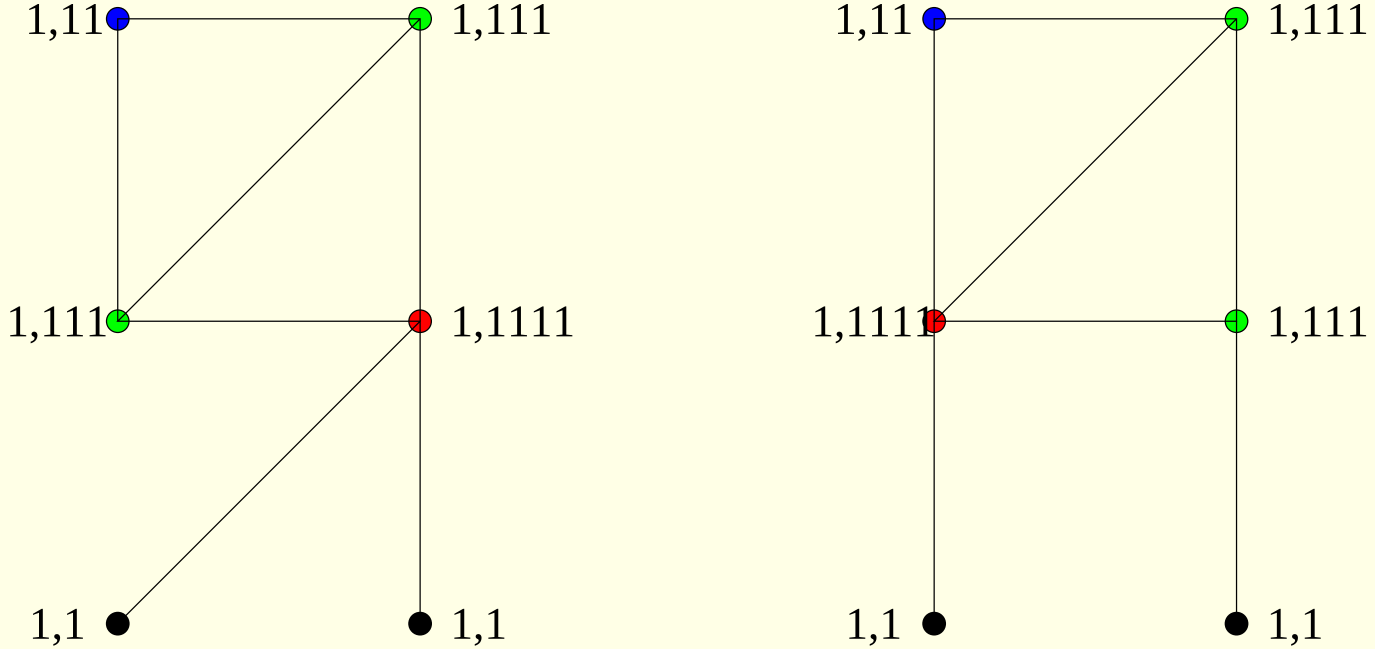
k-dimensional Weisfeiler-Lehman algorithm

1-dim WL $=$ color refinement procedure



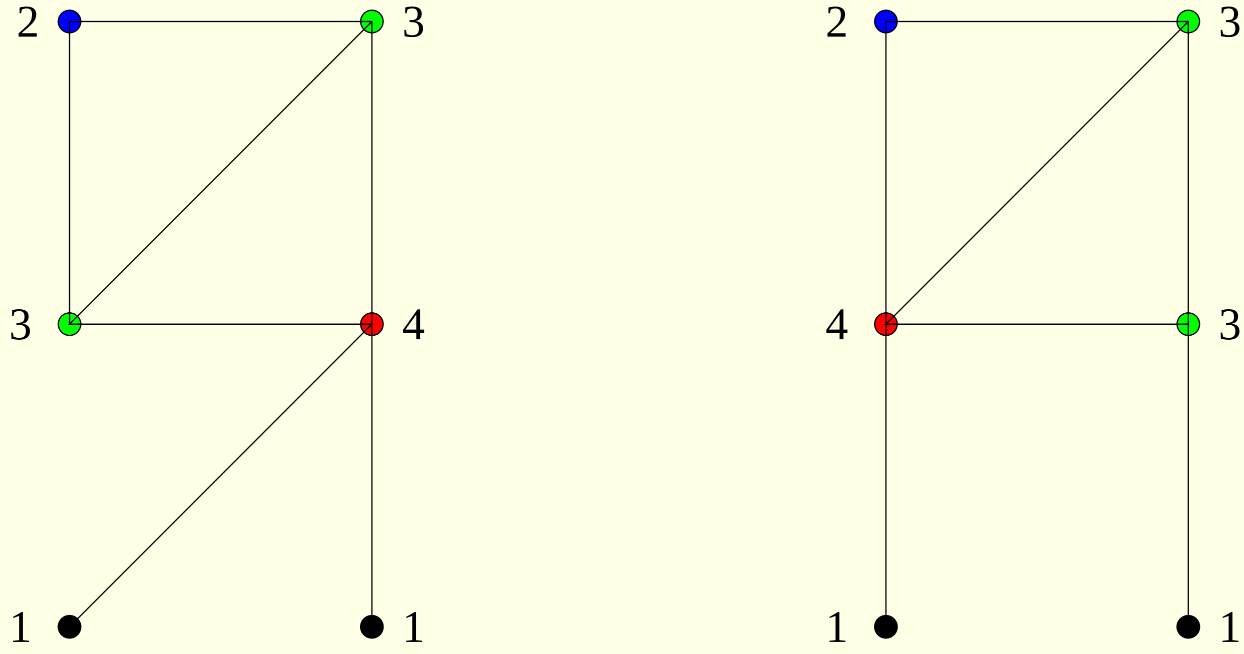
Initial coloring

1-dim WL $=$ color refinement procedure



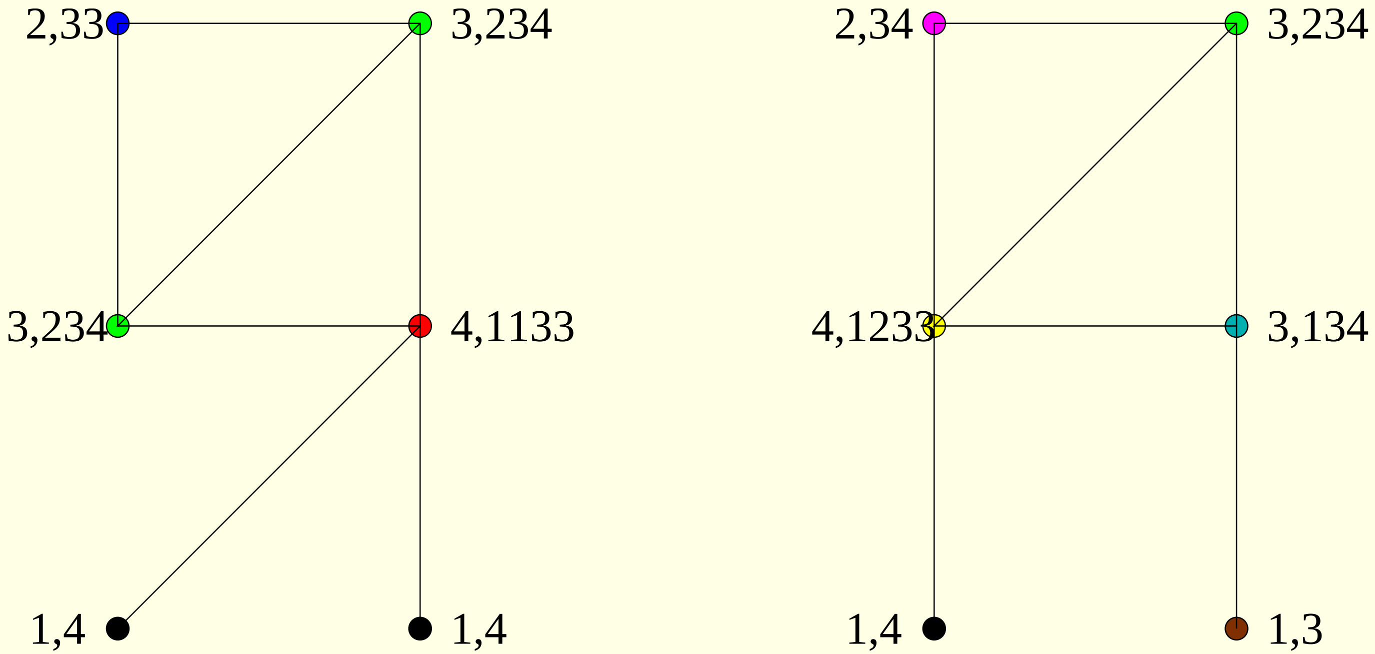
Refine coloring: for each vertex
New Color $=$ Old Color $+$ Old Colors of all neighbors

1-dim WL = color refinement procedure



Simplify color names

1-dim WL = color refinement procedure



Refine coloring again.

The multisets of colors differ,
hence the graphs are non-isomorphic.

k-dim WL

k-dim WL $=$ the same idea, but now we color V^k instead of V. The initial coloring of $(v_1, \ldots, v_k)$ is the isomorphism type of the subgraph induced on $v_1, \ldots, v_k$.

Theorem (Cai, Fürer, and Immerman)
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 for some (G, H) if

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

Theorem. Let $k \geq 2$ be a constant.

1. Let C be a class of graphs G with $D_{\#}^{k}(G) = O(\log n)$. Then Graph Isomorphism for C is solvable in TC^1.

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

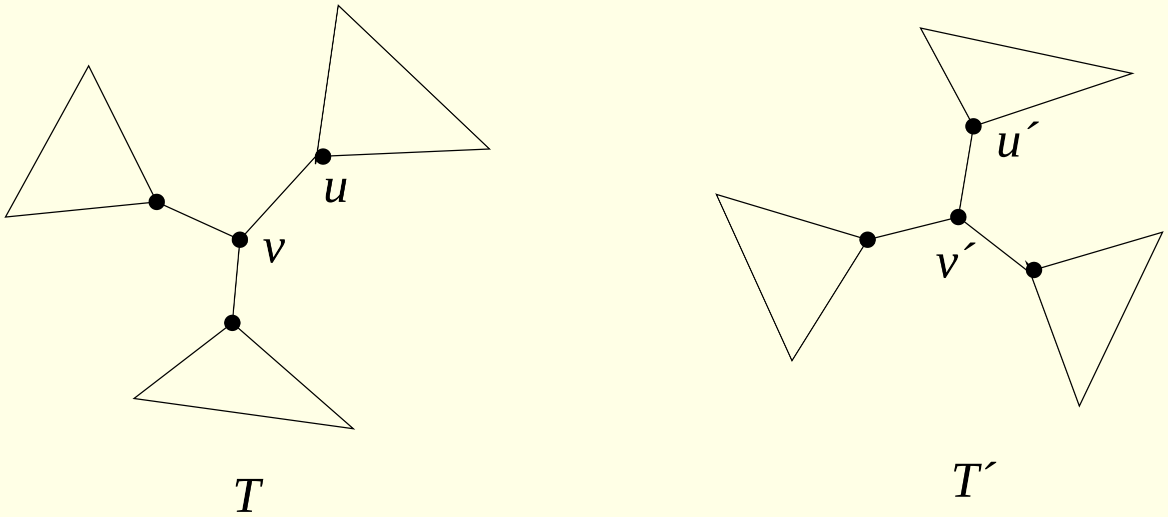
Example: Trees

Theorem. $D^3_{\#}(T) \le 3\log n + 2$ for every tree T on n vertices.

Proof (a separator strategy): Let $T' \not\cong T$ (and assume T' is a tree too). 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).

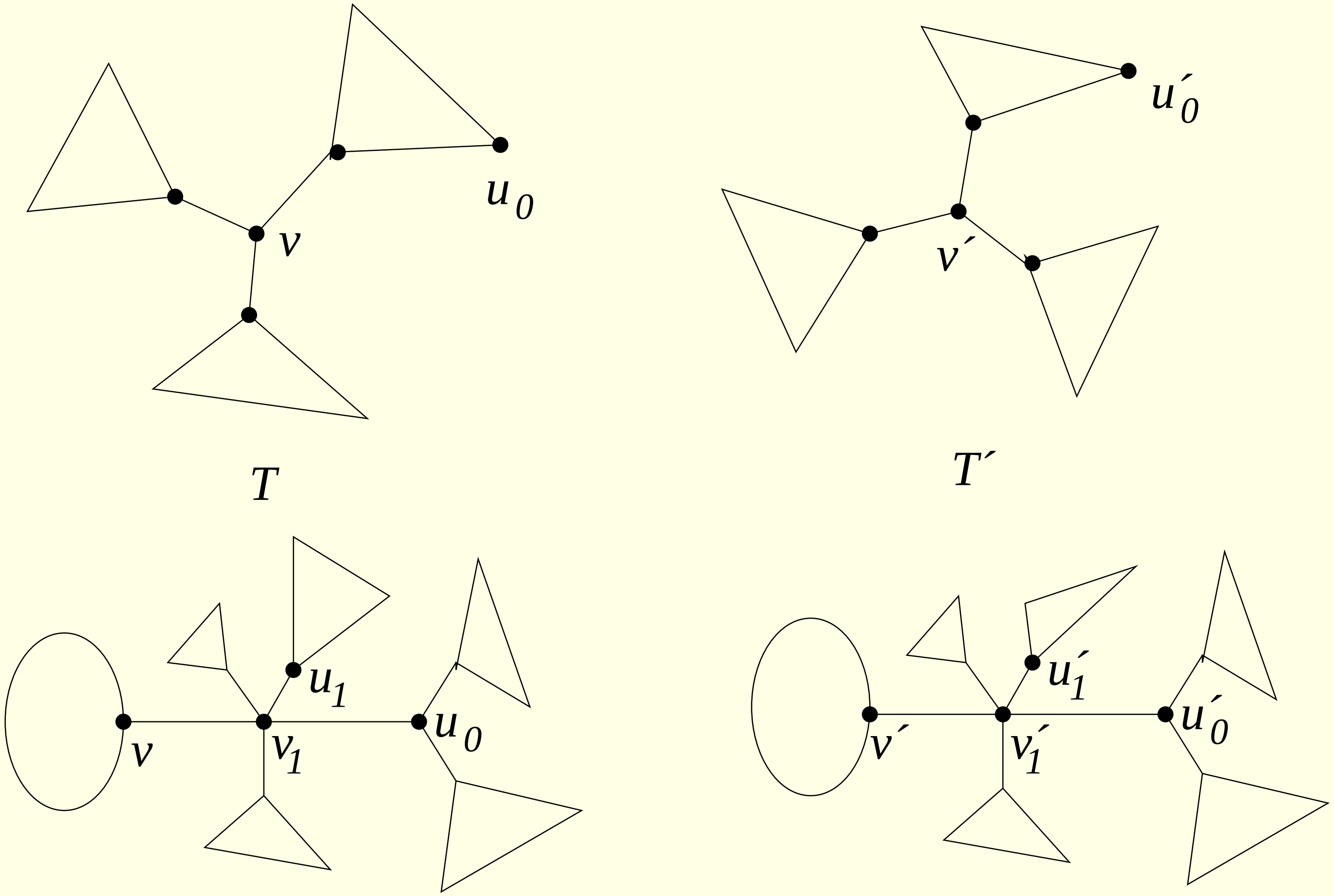
Step 2. Spoiler ensures pebbling $u \in \Gamma(v)$ and $u' \in \Gamma(v')$ so that the corresponding components are non-isomorphic rooted trees.



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

Proof - continuation

A complication: the strategy is now applied to a graph with one vertex pebbled and we may need more than 3 pebbles.



Step 3. If $T - v$ and $T' - v'$ differ only by the components with pebbled vertices u_0 and u'_0, then Spoiler pebbles a v_1 in the v-u_0-path such that $T - v_1$ and $T' - v'_1$ differ by components with no pebble. (The case that $d(v, u_0) \neq d(v', u'_0)$ or $d(v, v_1) \neq d(v', v'_1)$ is even easier for Spoiler.)

Isomorphism of trees (a revision of the history)

GI for trees is solvable in

- 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 ($W_\#(T) = 2$) *Aho-Hopcroft-Ullman 74*

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 TREE ISO in $\mathrm{TC}^1 \subseteq \mathrm{NC}^2$ and is known since 68 !

Estimates for particular classes

Theorem. If a graph G on n vertices has treewidth k, then

$$D_{\#}^{4k+4}(G) < 2(k+1)\log n + 8k + 9.$$

Consequently, isomorphism of graphs whose treewidth does not exceed k is recognizable by the $(4k+3)$-dim WL in $\mathrm{TC}^1 \subseteq \mathrm{NC}^2$.

Theorem. For a 3-connected planar graph G on n vertices we have

$$D^{15}(G) < 11\,\log_2 n + 45.$$

Consequently, the isomorphism problem for 3-connected planar graphs is solvable by the 14-dim WL in AC^1.

General bounds

Consider $G \ncong H$, both with n vertices.
It is easy to find example where

$$W(G,H) \geq \frac{n+1}{2}.$$

Moreover:

Theorem (Cai, Fürer, and Immerman) There are pairs of graphs such that

$$W_{\#}(G,H) = \Omega(n).$$

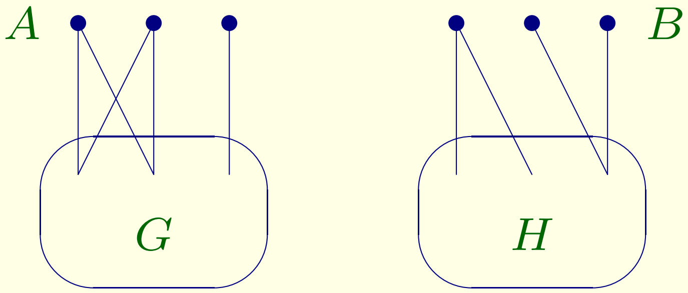
On the other hand:

How to show that $D(G,H) < n$ for all such pairs?

An initial approach with the Ramsey theorem

$$D(G,H) < n - \frac{1}{4}\log_2 n.$$

Indeed, as $D(G,H) = D(\bar{G},\bar{H})$ we can assume that G contains an independent set A with $|A| > \frac{1}{2}\log_2 n$.



- Spoiler selects all $V(G) \setminus A$.
- Duplicator selects $V(H) \setminus B$ for some B, $|B| = |A|$.

Identify $V(G) \setminus A$ and $V(H) \setminus B$ according to the partial isomorphism established. Suppose B is independent too for else Spoiler wins in 2 moves.

Given $X \subseteq V(G) \setminus A$, let $m_G(X) = |\{v \in A : \Gamma(v) = X\}|$
and $m_H(X) = |\{v \in B : \Gamma(v) = X\}|$.

Since $G \not\cong H$, there is X with $m_G(X) \neq m_H(X)$. Spoiler can now win in $\min\{m_G(X), m_H(X)\}+1$ moves. Since $\sum_X m_G(X) = \sum_X m_H(X)$, there are at least two such sets X_1 and X_2 and for one of them $\min\{m_G(X_i), m_H(X_i)\} < |A|/2$.

The final bound

Theorem. $D(G,H) \le \frac{n+3}{2}$ for all non-isomorphic G and H on n vertices.

Average case bounds

Theorem (Babai, Erdős, and Selkow). 2 color refinements split a random graph $G_{n,1/2}$ into color classes which are singletons with probability more than $1 - 1/\sqrt[7]{n}$, for all large enough n. Consequently, $D^2_{\#}(G_{n,1/2}) \le 4$ with this probability.

Theorem. With high probability we have

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

where $\omega = \omega(n)$ is an arbitrarily slowly increasing function.

Theorem. For infinitely many n we have

$$D_2(G_{n,1/2}) \le \log n - 2\log\log n + 5 + \log\log \mathrm{e} + o(1)$$

with high probability.

An application: The convergency rate in the 0-1 law

Let $p_n(\Phi) = \mathbb{P}\left[G_{n,1/2} \models \Phi\right]$.
0-1-law (Glebskii-et-al.–Fagin):
$p_n(\Phi)$ approaches 0 or 1 as $n \to \infty$.
Denote the limit by $p(\Phi)$.

Define the convergency rate function for the 0-1-law by

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

The standard version of the 0-1-law gives only the following: $R(k,n) \to 0$ 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 $c = 6$.

Succinctly definable graphs

How small can the logical depth be?

Definition (the succinctness functions).

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

$s(n) \to \infty$ as $n \to \infty$ but admits no recursive lower bound.

Theorem.
There is no general recursive function f such that

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

Definitions with no quantifier alternation

Consider the first order logic with no quantifier alternation (purely existential and universal formulas and their monotone Boolean combinations). This fragment of first order logic is provably weak.

FINITE SATISFIABILITY problem: Given a f.o. sentence about graphs, decide whether it is true on at least one finite graph.

Lavrov (63): FINITE SATISFIABILITY is unsolvable even for sentences without equality.

A sentence with no quantifier alternation is equivalent to a sentence in the Bernays-Schönfinkel class, i.e., to some sentence

$$\Phi = \exists x_1 \ldots \exists x_k \forall y_1 \ldots \forall y_l \Psi(\bar{x}, \bar{y}) \qquad (*)$$

FINITE SATISFIABILITY for the Bernays-Schönfinkel class is solvable. Moreover, let $\mathrm{Spectrum}(\Phi)$ consists of all those n such that there is a graph on n vertices satisfying Φ.

The logical Ramsey theorem (the case of graphs). Let Φ have form $(*)$. If $\mathrm{Spectrum}(\Phi)$ contains some $n \geq 2^k 4^l$, then $\mathrm{Spectrum}(\Phi)$ contains all $n \geq k + l$.

Definitions with no quantifier alternation

Nevertheless,

- $D_0(G_{n,1/2}) \le (2 + o(1)) \log_2 n$ with high probability;
- $D_0(G, H) \le \frac{n+5}{2}$ for all non-isomorphic graphs G and H of the same order n,

where $D_0(G)$ and $D_0(G, H)$ are analogs of $D(G)$ and $D(G, H)$ in the logic with no quantifier alternation.

Definitions with no quantifier alternation

Define

$$s_0(n) = \min \{ D_0(G) : G \text{ has } n \text{ vertices} \}.$$

Theorem. For all n we have

$$\log^* n - \log^* \log^* n - 2 \le s_0(n) \le \log^* n + 22.$$

Corollary. A gap in relation $D(G) \le D_0(G)$ can be super-recursive: There is no general recursive function f such that $D_0(G) \le f(D(G))$.

Thank you!