On Recognizing Graphs by Numbers of Homomorphisms

Zdeněk Dvořák

INSTITUTE FOR THEORETICAL COMPUTER SCIENCE
CHARLES UNIVERSITY, MALOSTRANSKÉ NÁM. 2/25
118 00 PRAGUE, CZECH REPUBLIC
E-mail: rakdver@kam.mff.cuni.cz

Received September 21, 2006; Revised August 7, 2009

Published online 18 November 2009 in Wiley InterScience (www.interscience.wiley.com).
DOI 10.1002/jgt.20461

Abstract: Let hom (G, H) be the number of homomorphisms from a graph G to a graph H. A well-known result of Lovász states that the function hom $(\cdot, H)$ from all graphs uniquely determines the graph H up to isomorphism. We study this function restricted to smaller classes of graphs. We show that several natural classes (2-degenerate graphs and graphs homomorphic to an arbitrary non-bipartite graph) are sufficient to recognize all graphs, and provide a description of graph properties that are recognizable by graphs with bounded tree-width. J Graph Theory 64: 330–342, 2010

Keywords: *graph homomorphisms; graph reconstruction; number of homomorphisms*

We consider simple undirected graphs without loops and parallel edges, unless otherwise specified. Let $\mathcal{A}$ be the class of all such graphs. Let hom(G,H) be the number of homomorphisms from a graph G to a graph H.

Contract grant sponsor: Czech Ministry of Education; Contract grant number: 1M0545 (to Z. D.).

Lovsz [8] proved that the function $\hom(\cdot,H)$ uniquely determines the graph H up to isomorphism, i.e., that if we know the number of isomorphisms from each graph in $\mathcal{A}$ to H, then we can uniquely reconstruct the graph H. We say that a class of graphs $\mathcal{G}$ *distinguishes* non-isomorphic graphs H_1 and H_2, if there exists a graph $G \in \mathcal{G}$ such that $\hom(G,H_1) \neq \hom(G,H_2)$. If $\mathcal{G}$ distinguishes every pair of non-isomorphic graphs H_1 and H_2, we call $\mathcal{G}$ *distinguishing*. Let us now restate Lovsz's result [8]:

Theorem 1. *The class $\mathcal{A}$ is distinguishing.*

We investigate whether smaller classes of graphs (e.g., graphs with bounded tree-width, chromatic number, etc.) are distinguishing. Fisk [5] studied a related problem—he considered G to distinguish H_1 and H_2 if $\hom(H_1,G) \neq \hom(H_2,G)$. In that setting, $\mathcal{A}$ is still distinguishing; however, the choice of suitable smaller classes is more restricted, since the chromatic number of graphs in such a distinguishing class must be unbounded.

In some cases, we conclude that the chosen class of graphs is not distinguishing (e.g., the class of graphs with tree-width bounded by a fixed constant, see Section 2). In such a case, we still can ask which pairs of graphs can be distinguished. We say that a class of graphs $\mathcal{G}$ *determines* a graph property P, if $\mathcal{G}$ distinguishes all pairs of graphs H_1 and H_2 such that H_1 has the property P and H_2 does not. In other words, the function $\hom(\cdot,H)$ restricted to $\mathcal{G}$ determines whether H has the property P or not.

We use the ideas of Lovsz and Szegedy [9] regarding the algebra of quantum graphs intensively.[1] The set $\{1,2,\ldots,k\}$ is denoted by $[k]$. A *k-labeled graph* G is a graph together with a partial function $\mathrm{lab}_G:[k] \to V(G)$, i.e., labels between 1 and k (but not necessarily all of them) are assigned to some (not necessarily distinct) vertices of G. The *set of labels* of G is the set $L_G = \{i \in [k] : \mathrm{lab}_G(i) \text{is defined}\}$. For a k-labeled graph G, let $\mathrm{base}(G)$ be the same graph without labels. Suppose that G and H are k-labeled graphs such that $L_G \subseteq L_H$. We define $\hom(G,H)$ as the number of homomorphisms $\varphi : V(G) \to V(H)$ that also preserve the labels, i.e., such that $\varphi(\mathrm{lab}_G(i)) = \mathrm{lab}_H(i)$ for each $i \in L_G$.

A (*k-labeled*) *quantum graph* G is a formal finite linear combination with real coefficients of (k-labeled) graphs. If the graphs are labeled, then we require all the graphs in the combination to have the same set of labels S, and define $L_G = S$. If each graph in the linear combination G belongs to a class $\mathcal{G}$ of graphs, we say that G is *$\mathcal{G}$-quantum graph*. We extend the functions $\hom(\cdot,H)$ (where H is a normal, non-quantum graph) linearly to quantum graphs, i.e.,

$$\hom\left(\left(\sum_{i=1}^{t} \alpha_i G_i\right), H\right) = \sum_{i=1}^{t} \alpha_i \hom(G_i,H).$$

[1]Note that we have modified their notation slightly in order to state some of the results in a more elegant way: we allow several labels to appear on the same vertex, we do not require the labels to be consecutive integers, and we define homomorphisms to labeled graphs instead of prescribing the images of labels by a separate function.

Similarly, we extend the function base($\cdot$) to k-labeled quantum graphs. The basic observation is that for the purposes of distinguishing, we can consider $\mathcal{G}$-quantum graphs instead of graphs from the class $\mathcal{G}$:

Lemma 2. *Let $\mathcal{G}$ be a class of graphs and H_1 and H_2 graphs. If there exists a $\mathcal{G}$-quantum graph $G=\sum_{i=1}^{t}\alpha_i G_i$ such that* $\hom(G,H_1)\neq\hom(G,H_2)$, *then $\mathcal{G}$ distinguishes graphs H_1 and H_2.*

Proof. Since $\sum_{i=1}^{t}\alpha_i\hom(G_i,H_1)\neq\sum_{i=1}^{t}\alpha_i\hom(G_i,H_2)$, there exists $i\in\{1,\ldots,t\}$ such that $\hom(G_i,H)\neq\hom(G_i,H_2)$. ■

A *product* G_1G_2 of two k-labeled graphs is a graph constructed by taking a disjoint union of G_1 and G_2, identifying the vertices with the same label, and suppressing the parallel edges that might be created. Note that G_1G_2 may contain loops, e.g., if $\mathrm{lab}_{G_1}(1)=\mathrm{lab}_{G_1}(2)$ and $\mathrm{lab}_{G_2}(1)$ and $\mathrm{lab}_{G_2}(2)$ are adjacent. The sets of labels of G_1 and G_2 do not have to be the same, in particular, if they are disjoint, G_1G_2 is just a disjoint union of G_1 and G_2. For quantum graphs $G_1=\sum_i\alpha_{1,i}G_{1,i}$ and $G_2=\sum_i\alpha_{2,i}G_{2,i}$, we define G_1G_2 as $\sum_{i,j}\beta_{i,j}G_{1,i}G_{2,j}$, where $\beta_{i,j}=\alpha_{1,i}\alpha_{2,j}$ if $G_{1,i}G_{2,j}$ is loop-less and 0 otherwise, i.e., we remove the graphs with loops from the resulting linear combination. Note that if H is loop-less, then ignoring the graphs with loops in the linear combination preserves the value of $\hom(\cdot,H)$. If G_1 and G_2 are two (labeled, quantum) graphs, then $\hom(G_1G_2,H)=\hom(G_1,H)\hom(G_2,H)$ for every loop-less graph H.

For an integer $t\geq 1$, we write G^t for the product of t copies of a (labeled, quantum) graph G. For a set $S\subseteq[k]$, let I_S be the edge-less k-labeled graph with $V(I_S)=S$ and $\mathrm{lab}_{I_S}(i)=i$ for each $i\in S$. We set $G^0=I_{L_G}$. Note that $I_\emptyset$ is the empty graph with no vertices. Let $E_{i,j}$ be the labeled graph consisting of two adjacent vertices v_1 and v_2, such that $L_{E_{i,j}}=\{i,j\}$, $\mathrm{lab}_{E_{i,j}}(i)=v_1$ and $\mathrm{lab}_{E_{i,j}}(j)=v_2$. Let $J_{i,j}$ be the k-labeled graph consisting of a single vertex v such that $L_{J_{i,j}}=\{i,j\}$ and $\mathrm{lab}_{J_{i,j}}(i)=\mathrm{lab}_{J_{i,j}}(j)=v$.

1. COMPLEXITY REMARKS

An important open question of complexity theory is whether the graph non-isomorphism problem is in NP. The fact that $\mathcal{A}$ distinguishes all non-isomorphic graphs "almost" gives answer to this question. If H_1 and H_2 are non-isomorphic, then there exists a witness that they are non-isomorphic (a graph G such that $\hom(G,H_1)\neq\hom(G,H_2)$). This witness may be chosen to have polynomial size (at most $\max(|V(H_1)|,|V(H_2)|)$ vertices). The only problem is that deciding whether $\hom(G,H_1)=\hom(G,H_2)$ is NP-hard (with G, H_1 and H_2 in the input, and even for most fixed pairs of graphs H_1 and H_2), and thus it is not likely that we would be able to find a polynomial-time algorithm to verify this witness.

However, for some classes of graphs $\mathcal{G}$ (by Dalmau and Jonsson [4], exactly the classes of graphs with tree-width bounded by a constant, unless #W[1]=FPT), it is possible to determine $\hom(G,H)$ in polynomial time for every $G\in\mathcal{G}$ and $H\in\mathcal{A}$. We might thus hope that some such class $\mathcal{G}$ is distinguishing, proving (assuming that the graph in $\mathcal{G}$ distinguishing H_1 and H_2 would have polynomial size) that graph non-isomorphism is in NP.

Of course, this turns out not to be the case. The classes of graphs with bounded tree-width are not distinguishing, as we show in the following section. In fact the polynomial-time algorithm for counting the number of homomorphisms from a graph with bounded tree-width is the base of the proof that these classes are not distinguishing.

2. GRAPHS WITH BOUNDED TREE-WIDTH

A graph G has *tree-width* at most k if there exists a tree T satisfying the following conditions:

- the vertices of T are subsets of $V(G)$ of size at most $k+1$,
- each edge of G is a subset of a vertex of T, and
- for each $v \in V(G)$, the vertices of T that contain v induce a connected subgraph of T.

In particular, graphs with tree-width at most one are forests. For algorithmic purposes, the tree T is usually rooted, and its vertices are duplicated in such a way that T is binary. Furthermore, by adding new vertices to T, we may ensure that each vertex of G appears in a leaf of T; see, e.g., Bodlaender [1] for details. Such a rooted tree can be interpreted as a construction of G by a finite sequence of products and label deletions, i.e., a $(k+1)$-labeled graph G has tree-width at most k if

- $V(G)=L_G$ (thus G has at most $k+1$ vertices), or
- $G=G_1G_2$, where G_1 and G_2 are $(k+1)$-labeled graphs of tree-width at most k, or
- G is obtained from a $(k+1)$-labeled graph G' of tree-width at most k by removing some of the labels, i.e., restricting $\text{lab}_{G'}$ to a subset of $L_{G'}$.

Trees (or forests) are not sufficient to distinguish all graphs—any two d-regular graphs on the same number of vertices have the matching numbers of homomorphisms from all trees. Similarly, two strongly regular graphs with the same parameters cannot be distinguished using graphs with tree-width at most two. It might seem that we could proceed in a similar manner with the other classes of graphs with bounded tree-width by simply strengthening the constraints on the regularity of the graphs that cannot be recognized; however, for graphs of tree-width at least 5, the only sufficiently regular graphs would be unions of complete graphs of the same size, their complements, C_5 and the line graph of $K_{3,3}$ (proved by Cameron [3] and independently by Gol'fand [7]). Therefore, we need a more precise characterization of the graphs that cannot be distinguished by graphs with small tree-width. Let us start with a few definitions.

The *degree refinement* of a graph H is coloring of vertices of H by m distinct vectors $w_i=(n^i_1,n^i_2,\ldots,n^i_m)$ such that for each $1\leq i,j\leq m$, each vertex with color w_i has exactly n^i_j neighbors with color w_j, and k is the smallest possible. The degree refinement of a graph is unique up to a permutation of colors, and there exists a polynomial time algorithm that determines the degree refinement in a canonical form.

This makes the concept useful for isomorphism testing, as any isomorphism of graphs H_1 and H_2 preserves colors of their canonical degree refinements; see, e.g., Cai et al. [2] for more details.

The concept of the degree refinement can be extended to the classification of k-tuples of vertices (appearing in Weisfeiler-Lehman method for isomorphism testing [11, 12]). Consider a coloring γ of k-tuples of vertices of a graph H. Given a k-tuple $U=(u_1,u_2,\ldots,u_k)$ and a vertex v, let $e(\gamma,U,v)$ be the k-tuple of colors $(\gamma(v,u_2,\ldots,u_k),\gamma(u_1,v,\ldots,u_k),\ldots,\gamma(u_1,u_2,\ldots,v))$. For $k>1$, a *k-degree refinement* of H is a coloring γ of the k-tuples of the vertices of H together with a mapping ψ from the colors to graphs with at most k vertices, such that γ uses the smallest possible number of colors, and

- for each k-tuple U of vertices of H, the subgraph of H induced by U is equal to $\psi(\gamma(U))$, and
- for any two k-tuples U_1 and U_2 of vertices of H such that $\gamma(U_1)=\gamma(U_2)$ and for any k-tuple C of colors, $|\{v \in V(H) : e(\gamma,U_1,v) = C\}| = |\{v \in V(H) : e(\gamma,U_2,v)=C\}|$.

Let us define the 1-*degree refinement* to be the degree refinement (note that the definition of k-degree refinement in the previous paragraph would give a different result if used with $k=1$, since for 1-degree refinement, we need to add the information about edges explicitly). The main result of this section states that two graphs are distinguished by graphs with tree-width at most k if and only if their k-degree refinements are different.

A *k-variable first-order formula with counting* is a formula φ built in the usual way from variables $x_1,\ldots,x_k$ (that stand for vertices), the relation symbols $=$ and E (adjacency), true and false, logical connectives $\wedge$, $\vee$, and $\neg$, and quantifiers $\exists$, $\forall$, and $\exists_t$, where t may be any non-negative integer. Note that the variables may be "reused," e.g., formula $(\forall x_1)(\exists x_2)(E(x_1,x_2)\wedge(\exists x_1)(x_1\neq x_2\wedge\neg E(x_1,x_2)))$ says that each vertex has a neighbor that is not universal. The set of all such formulas is denoted by $\mathcal{C}_k$. A variable is *free* in φ if it has a non-quantified occurrence. A formula is called *closed* if it has no free variables. The semantics is defined as follows. Let H be a k-labeled graph such that L_H contains indices of all free variables of φ. Given a label i and a vertex $v\in V(H)$, let $H(i\to v)$ be the graph H' obtained from H by setting $\mathrm{lab}_{H'}(i):=v$. We write $H\vDash\varphi$ if

- $\varphi=$true; or
- φ is $x_i=x_j$ and $\mathrm{lab}_H(i)=\mathrm{lab}_H(j)$; or
- φ is $E(x_i,x_j)$ and $\mathrm{lab}_H(i)$ and $\mathrm{lab}_H(j)$ are adjacent in H; or
- φ is $\varphi_1\wedge\varphi_2$, and $H\vDash\varphi_1$ and $H\vDash\varphi_2$; or
- φ is $\varphi_1\vee\varphi_2$, and $H\vDash\varphi_1$ or $H\vDash\varphi_2$; or
- φ is $\neg\varphi_1$, and not $H\vDash\varphi_1$; or
- φ is $(\exists x_i)\varphi_1$, and there exists a vertex $v\in V(H)$ such that $H(i\to v)\vDash\varphi_1$; or
- φ is $(\forall x_i)\varphi_1$, and all vertices $v\in V(H)$ satisfy $H(i\to v)\vDash\varphi_1$; or
- φ is $(\exists_t x_i)\varphi_1$, and there exist at least t vertices $v_1,v_2,\ldots,v_t\in V(H)$ such $H(i\to v_j)\vDash\varphi_1$ for every $1\le j\le t$.

We also use $\exists_{!t}$ to mean that there are exactly t vertices with the property, i.e., $(\exists_{!t}x_i)\,\varphi$ is a shorthand for $(\exists_t x_i)\varphi\wedge\neg(\exists_{t+1}x_i)\varphi$. Informally, a formula in $\mathcal{C}_k$ describes a property that can be determined by working exclusively with k-tuples of vertices of H. The following theorem was proved by Cai et al. [2]:

Theorem 3. *For any graphs H_1 and H_2, the following two statements are equivalent*:

- *H_1 and H_2 have the same k-degree refinement (up to a permutation of colors).*
- *For each closed formula $\varphi \in \mathcal{C}_{k+1}$, $H_1 \vDash \varphi$ if and only if $H_2 \vDash \varphi$.*

Let us start with the following lemma, showing that graphs having a fixed number of homomorphisms from a graph with tree-width at most k can be described by a formula with $k+1$ variables.

Lemma 4. *If G is a $(k+1)$-labeled graph of tree-width at most k, and m is a non-negative integer, then there exists a formula $\varphi \in \mathcal{C}_{k+1}$ such that for each $(k+1)$-labeled graph H with $L_G \subseteq L_H$, $H \vDash \varphi$ if and only if* $\hom(G,H)=m$.

Proof. We proceed inductively by the recursive construction of G. The basic case is that G is a graph with $V(G)=L_G$. As $L_G \subseteq L_H$, there exists at most one homomorphism from G to H. If $m>1$, then we set $\varphi=$false. If $m=1$, then we let φ be the conjunction of terms $E(x_i,x_j)$ for each two labels $i,j \in L_G$ such that $\mathrm{lab}_G(i)$ is adjacent to $\mathrm{lab}_G(j)$ in G, and of terms $x_i = x_j$ for each $i,j \in L_G$ such that $\mathrm{lab}_G(i)=\mathrm{lab}_G(j)$. If $m=0$, then φ is the negation of this conjunction.

Suppose now that G is obtained from a graph G' of tree-width at most k by removing a label l. By induction hypothesis, for each integer n there exists a formula φ_n such that for each H with $L_{G'} \subseteq L_H$, $H \vDash \varphi_n$ if and only if $\hom(G',H)=n$. Suppose first that $m>0$. Consider a decomposition $m=\sum_{i=1}^{t} c_i m_i$ such that c_i is a positive integer for $1 \le i \le t$, and the numbers $m_1, m_2, \ldots, m_t$ are mutually distinct positive integers. Let $c=\sum_{i=1}^{t} c_i$ and $\varphi_{(c_1,m_1,\ldots,c_t,m_t)} = (\exists!_c x_l)\neg\varphi_0 \wedge \bigwedge_{i=1}^{t} (\exists!_{c_i} x_l)\varphi_{m_i}$. Let φ be the disjunction of formulas $\varphi_{(c_1,m_1,\ldots,c_t,m_t)}$ taken over all such decompositions $\sum_{i=1}^{t} c_i m_i$ of m. Observe that $\hom(G,H)=m$ if and only if there exists a decomposition $m=\sum_{i=1}^{t} c_i m_i$ such that $\hom(G',H(l \to v)) \neq 0$ for exactly $\sum_{i=1}^{t} c_i$ vertices $v \in V(H)$, and $\hom(G',H(l \to v))=m_i$ for exactly c_i vertices $v \in V(H)$ for each $i \in \{1,\ldots,t\}$. We conclude that $\hom(G,H)=m$ if and only if $H \vDash \varphi$. Similarly, if $m=0$, then let $\varphi=(\forall x_l)\varphi_0$.

Finally, suppose that $G=G_1G_2$, where G_1 and G_2 have tree-width at most k. By induction hypothesis, there exist formulas φ_n^i for $i \in \{1,2\}$ and any integer n, such that $H \vDash \varphi_n^i$ if and only if $\hom(G_i,H)=n$ for any graph H with $L_{G_i} \subseteq L_H$. If $m \neq 0$, then we let φ be the disjunction of terms $\varphi_{m_1}^1 \wedge \varphi_{m_2}^2$ for each pair of positive integers m_1 and m_2 such that $m=m_1m_2$. If $m=0$, then we let $\varphi=\varphi_0^1 \vee \varphi_0^2$. Analogically to the previous paragraph, we conclude that $\hom(G,H)=m$ if and only if $H \vDash \varphi$. ∎

Now, we prove that the properties expressed by formulas with at most $k+1$ variables are determined by the class of graphs with tree-width at most k. We first need the following observation (made for series-parallel graphs by Lovsz and Szegedy [9], Claim 4.1).

Lemma 5. *Let G be a $(k+1)$-labeled quantum graph of tree-width at most k. If X_0 and X_1 are disjoint finite sets of real numbers, then there exists a $(k+1)$-labeled quantum graph $G[X_0,X_1]$ of tree-width at most k such that for each graph H,*

- *if* $\hom(G,H)\in X_0$ *then* $\hom(G[X_0,X_1],H)=0$, *and*
- *if* $\hom(G,H)\in X_1$ *then* $\hom(G[X_0,X_1],H)=1$.

Proof. Let $p(x)=\sum_{i=0}^{t} a_i x^i$ be a polynomial such that $p(x)=0$ for every $x\in X_0$ and $p(x)=1$ for every $x\in X_1$. We set $G[X_0,X_1]=\sum_{i=0}^{t} a_i G^i$. The claim of the lemma follows from the fact that $\hom(G^i,H)=(\hom(G,H))^i$ for each integer $i\geq 0$. ∎

For a formula φ, a (labeled) quantum graph G *models* φ *for graphs of size* n if L_G consists of the indices of the free variables of φ, and for each graph H on n vertices with $L_G\subseteq L_H$,

- if $H\vDash\varphi$ then $\hom(G,H)=1$, and
- if $H\nvDash\varphi$ then $\hom(G,H)=0$.

Lemma 6. *For each formula $\varphi\in\mathcal{C}_{k+1}$ and for each positive integer n, there exists a quantum graph G of tree-width at most k such that G models φ for graphs of size n.*

Proof. Let us proceed inductively by the structure of φ. If $\varphi=$true, then let $G=I_\emptyset$. If $\varphi=$false, then let $G=0$ (i.e., the linear combination in that no coefficient is nonzero). If $\varphi=(x_i=x_j)$, then let $G=J_{i,j}$. If $\varphi=E(x_i,x_j)$ and $i\neq j$, then let $G=E_{i,j}$. If $i=j$, then let $G=0$, since H is loop-less and thus the predicate can never be satisfied.

If $\varphi=\varphi_1\wedge\varphi_2$, then by induction there exist quantum graphs G_1 and G_2 of tree-width at most k such that G_i models φ_i for graphs of size n for $i\in\{1,2\}$. Observe that G_1G_2 models φ for graphs of size n. Similarly, $(G_1+G_2)[\{0\},\{1,2\}]$ models $\varphi_1\vee\varphi_2$, and $G_1[\{1\},\{0\}]$ models $\neg\varphi_1$.

If $\varphi=(\exists x_i)\varphi_1$, then by induction there exists a quantum graph G_1 that models φ_1 for graphs of size n. Let G_1' be the graph G_1 without the label i. The graph $G_1'[\{0\},\{1,2,\ldots,n\}]$ models φ for graphs of size n. Similarly, $G_1'[\{0,1,\ldots,t-1\},\{t,t+1,\ldots,n\}]$ models $(\exists_t x_i)\varphi_1$ and $G_1'[\{0,1,\ldots,n-1\},\{n\}]$ models $(\forall x_i)\varphi_1$. ∎

Now we can state the main result of this section:

Theorem 7. *For any two graphs H_1 and H_2, the following conditions are equivalent:*

1. *There exists a closed formula $\varphi\in\mathcal{C}_{k+1}$ such that $H_1\vDash\varphi$ and $H_2\nvDash\varphi$.*
2. *There exists a graph G of tree-width at most k such that* $\hom(G,H_1)\neq\hom(G,H_2)$.

Proof. Let us first prove (1)$\Longrightarrow$(2). If $|V(H_1)|\neq|V(H_2)|$, then we let $G=K_1$. Therefore, assume that $n=|V(H_1)|=|V(H_2)|$. By Lemma 6, let G' be a quantum graph of tree-width at most k that models φ for graphs of size n. It follows that $\hom(G',H_1)=1$ and $\hom(G',H_2)=0$. By Lemma 2, there exists a graph G of tree-width at most k such that $\hom(G,H_1)\neq\hom(G,H_2)$.

Now, let us prove that (2)$\Longrightarrow$(1). Let $m=\hom(G,H_1)$. By Lemma 4, there exists a formula $\varphi\in\mathcal{C}_{k+1}$ such that for any graph H, $H\vDash\varphi$ if and only if $\hom(G,H)=m$. Therefore, $H_1\vDash\varphi$ and $H_2\nvDash\varphi$. ∎

Cai et al. [2] proved that for each k, there exist non-isomorphic graphs H_1 and H_2 such that for each $\varphi \in \mathcal{C}_{k+1}$, $H_1 \vDash \varphi$ if and only if $H_2 \vDash \varphi$. These graphs thus cannot be distinguished using only graphs of tree-width at most k.

Another well-known width parameter is clique-width. Any graph with bounded tree-width has also a bounded clique-width, but not vice versa, thus we might hope a class of graphs with clique-width bounded by a constant is distinguishing. However, Gurski and Wanke [6] showed the following statement:

Theorem 8. *Let $t > 1$ be an arbitrary integer. If a graph G does not contain $K_{t,t}$ as a subgraph, then* $\mathrm{tw}(G) \leq 3(t-1)\mathrm{cw}(G) - 1$.

In particular, a graph of clique-width at most k and girth at least 5 has tree-width at most $3k-1$. Let us now consider the non-isomorphic graphs H_1 and H_2 that cannot be distinguished by formulas in $\mathcal{C}_{6k}$ and let H_i' be the graph obtained from H_i by subdividing each edge once, for $i \in \{1,2\}$. Observe that H_1' and H_2' are not isomorphic, and that it is not possible to distinguish them by formulas in $\mathcal{C}_{3k}$. It follows that $\hom(G, H_1') = \hom(G, H_2')$ for any graph G of clique-width at most k and girth at least 5. Furthermore, as the girth of H_i' is at least 6, $\hom(G, H_1') = \hom(G, H_2') = 0$ for any graph G of girth at most 5. We conclude that H_1' and H_2' cannot be distinguished using graphs with clique-width at most k. Therefore, graphs with bounded clique-width are not significantly more powerful than graphs with bounded tree-width with regards to distinguishing graphs.

3. GRAPHS WITH BOUNDED DEGENERACY

A graph G is *k-degenerate* if each subgraph of G contains a vertex of degree at most k. Every graph with tree-width k is k-degenerate and 1-degenerate graphs are exactly forests (graphs of tree-width 1), but there are 2-degenerate graphs with arbitrarily large tree-width. It turns out that 2-degenerate graphs are sufficient to distinguish all graphs.

Let us call a 2-labeled quantum graph that is a linear combination of paths of length at least two joining the vertex with label 1 with the vertex with label 2 a *quantum path*. A 2-labeled quantum graph C is a *connector* for a graph H if

$$\hom(C, H') = \hom(E_{1,2}, H') \tag{1}$$

for every 2-labeled graph H' with $\mathrm{base}(H') = H$. Trivially, $E_{1,2}$ is a connector. More interestingly, Lovsz and Szegedy [9, Theorem 1.4] proved the following:

Theorem 9. *For each H, there exists a quantum path C such that C is a connector for H.*

If C is a connector for H, then for each 2-labeled graph G,

$$\begin{aligned}\hom(\mathrm{base}(GC), H) &= \sum_{u,v \in V(H)} \hom(GC, H(1 \to u, 2 \to v)) \\ &= \sum_{u,v \in V(H)} \hom(G, H(1 \to u, 2 \to v)) \hom(C, H(1 \to u, 2 \to v))\end{aligned}$$

$$= \sum_{u,v\in V(H)} \hom(G, H(1\to u, 2\to v))\hom(E_{1,2}, H(1\to u, 2\to v))$$

$$= \sum_{u,v\in V(H)} \hom(GE_{1,2}, H(1\to u, 2\to v))$$

$$= \hom(\mathrm{base}(GE_{1,2}), H)$$

That is, the connector can be used to replace an edge of a graph without affecting the number of homomorphisms to H. Lovsz and Szegedy [9] use this property as the definition of the connector, and claim (without a proof) that it is equivalent to Equation (1). However, it is not immediately obvious that the equation hom(base (GC),H)=hom(base($GE_{1,2}$),H) implies Equation (1), thus we chose to use the latter as a definition instead (let us point out that the proof of Theorem 9 finds a quantum path that satisfies this definition). Let us now prove a slight variation of Theorem 9:

Lemma 10. *If H_1 and H_2 are unlabeled graphs, then there exists a quantum path C such that C is a connector for both H_1 and H_2.*

Proof. Let C be a quantum path connector for the disjoint union $H = H_1 H_2$ that exists by Theorem 9. Let u and v be two vertices of H_1, and let $H_1' = H_1(1\to u, 2\to v)$. Then, $\hom(C, H_1') = \hom(C, H_1' H_2)$, since C is connected, and $\hom(C, H_1' H_2) = \hom(E_{1,2}, H_1' H_2) = \hom(E_{1,2}, H_1')$. Therefore, C is a connector for H_1, and by symmetry also for H_2. ■

We use the existence of a common path connector to show that we may subdivide edges in each graph that distinguishes H_1 and H_2.

Lemma 11. *Let G, H_1 and H_2 be graphs such that* $\hom(G, H_1) \neq \hom(G, H_2)$ *and e an edge of G. There exists a graph G' obtained from G by replacing e by a path of length at least two, such that* $\hom(G', H_1) \neq \hom(G', H_2)$.

Proof. Let C be the quantum path connector for H_1 and H_2 obtained using Lemma 10. Let G''' be the graph $G - e$ with the vertices of e labeled with 1 and 2, and let $G'' = \mathrm{base}(G'''C)$ be the quantum graph obtained from G by replacing e with C. By the definition of the connector, $\hom(G'', H_1) = \hom(G, H_1) \neq \hom(G, H_2) = \hom(G'', H_2)$. By Lemma 2, we have $\hom(G', H_1) \neq \hom(G', H_2)$ for at least one graph G' in the linear combination G'', and G' satisfies the claim of the lemma. ■

The main result of this section is the following:

Theorem 12. *If H_1 and H_2 are not isomorphic, then there exists a 2-degenerate graph G such that* $\hom(G, H_1) \neq \hom(G, H_2)$.

Proof. By Theorem 1, there exists a graph G' that distinguishes H_1 and H_2. Using Lemma 11, we construct the graph G by replacing edges of G' by paths of length at least two, while maintaining $\hom(G, H_1) \neq \hom(G, H_2)$. The graph G is 2-degenerate, as each subgraph of G that contains at least one edge also contains at least one of the vertices of $V(G) \setminus V(G')$ of degree at most two. ■

We may continue subdividing the edges of G in Theorem 12, making the resulting graph arbitrarily sparse. Using the formalism introduced recently by Neetil and Ossona de Mendez [10], this shows that there exists a distinguishing class of graphs with bounded expansion.

4. GRAPHS HOMOMORPHIC TO A FIXED GRAPH

Let $\mathcal{A}_{\leq M}$ be the class of graphs G that are homomorphic to M, i.e., such that $\hom(G,M)>0$. The results of the previous section imply that if M is not bipartite, then $\mathcal{A}_{\leq M}$ is distinguishing (since a sufficiently fine subdivision of every graph is M-colorable, for any fixed non-bipartite graph M). On the other hand, it is easy to find graphs that are not distinguished by the class $\mathcal{A}_{\leq K_2}$ of bipartite graphs. However, it is interesting to derive these results in a more systematic way.

We let $H_1 \times H_2$ denote the categorical product of two graphs (the graph with $V(H_1 \times H_2) = V(H_1) \times V(H_2)$ and $\{\langle u_1,u_2\rangle,\langle v_1,v_2\rangle\} \in E(H_1 \times H_2)$ if and only if $\{u_1,v_1\} \in E(H_1)$ and $\{u_2,v_2\} \in E(H_1)$), and let $\pi_1^{H_1 \times H_2}: V(H_1 \times H_2) \to V(H_1)$ and $\pi_2^{H_1 \times H_2}: V(H_1 \times H_2) \to V(H_2)$ be the associated projections.

Theorem 13. *Let H_1, H_2, and M be graphs. Then,* $\hom(G,H_1)=\hom(G,H_2)$ *holds for every $G \in \mathcal{A}_{\leq M}$ if and only if there exists an isomorphism $f: H_1 \times M \to H_2 \times M$ such that $\pi_2^{H_1 \times M} = \pi_2^{H_2 \times M} f$ (i.e., for each $v \in V(H_1)$ and $m \in V(M)$, $f(\langle v,m\rangle) = \langle w,m\rangle$ for some $w \in V(H_2)$).*

Proof. The theorem is obvious if M has no edges. Let us assume in the rest of the proof that $K_2 \in \mathcal{A}_{\leq M}$. Suppose first that there exists the isomorphism f with the required properties, and let G be a graph from $\mathcal{A}_{\leq M}$. Let c be a fixed homomorphism from G to M. We construct a bijection between the homomorphisms from G to H_1 and the homomorphisms from G to H_2, thus showing that $\hom(G,H_1)=\hom(G,H_2)$. Given a homomorphism $g_1: V(G) \to V(H_1)$, we define the function $g_2: V(G) \to V(H_2)$ by $g_2(v) = \pi_1^{H_2 \times M}(f(\langle g_1(v), c(v)\rangle))$. The function g_2 is a homomorphism: if $\{u,v\}$ is an edge of G, then $\{g_1(u),g_1(v)\} \in E(H_1)$ and $\{c(u),c(v)\} \in E(M)$, thus $\{\langle g_1(u),c(u)\rangle,\langle g_1(v),c(v)\rangle\} \in E(H_1 \times M)$, and $\{g_2(u),g_2(v)\}$ is an edge of H_2. The mapping from g_1 to g_2 is a bijection, since $g_1(v) = \pi_1^{H_1 \times M}(f^{-1}(\langle g_2(v),c(v)\rangle))$.

Suppose now that $\hom(G,H_1)=\hom(G,H_2)$ for every $G \in \mathcal{A}_{\leq M}$. We use the idea of Lovsz [8]. For a graph H, let $I(H,H_2)$ be the number of homomorphisms $g: V(H \times M) \to V(H_2)$ such that for each two vertices $x \neq y$ of H and each $m \in V(M)$, $g(\langle x,m\rangle) \neq g(\langle y,m\rangle)$. Let $A_{x,y,m}$ be the set of homomorphisms g from $H \times M$ to H_2 such that $g(\langle x,m\rangle) = g(\langle y,m\rangle)$. By the principle of inclusion and exclusion,

$$I(H,H_2) = \left| \bigcap_{i \in V(H)^2 \times V(M)} \overline{A_i} \right| = \hom(H \times M, H_2) - \sum_{\emptyset \neq I \subseteq V(H)^2 \times V(M)} (-1)^{|I|+1} \left| \bigcap_{i \in I} A_i \right|.$$

However,

$$\left|\bigcap_{i\in I} A_i\right| = \hom(H_I, H_2),$$

where H_I is the graph obtained from $H\times M$ by identifying all pairs of vertices $\langle x,m\rangle$ and $\langle y,m\rangle$ such that $\langle x,y,m\rangle\in I$. Since we only identify the vertices with the same projection to M, H_I is homomorphic to M. Thus, $\hom(H_I,H_2)=\hom(H_I,H_1)$, and $I(H,H_1)=I(H,H_2)$ for any graph H. In particular, this means that $I(H_1,H_1)=I(H_1,H_2)$. Since the projection $\pi_1^{H_1\times M}$ is one of the homomorphisms counted by $I(H_1,H_1)$, it follows that $I(H_1,H_2)=I(H_1,H_1)>0$, hence there exists a homomorphism g from $H_1\times M$ to H_2 such that $g(\langle x,m\rangle)\neq g(\langle y,m\rangle)$ for each $x\neq y$ and each m. We define the function f by $f(\langle v,m\rangle)=\langle g(\langle v,m\rangle),m\rangle$.

Let us show that f is a homomorphism. Suppose that $\{\langle u,m_1\rangle,\langle v,m_2\rangle\}$ is an edge of $H_1\times M$. Then $\{m_1,m_2\}\in E(M)$, and since g is a homomorphism, $\{g(\langle u,m_1\rangle),g(\langle v,m_2\rangle)\}\in E(H_2)$. Therefore, $\{f(\langle u,m_1\rangle),f(\langle v,m_2\rangle)\}$ is an edge of $H_2\times M$.

The function f is obviously injective. Since $\hom(K_1,H_1)=\hom(K_1,H_2)$, the graphs H_1 and H_2 have the same number of vertices, hence f is surjective. Similarly, since $\hom(K_2,H_1)=\hom(K_2,H_2)$, the graphs H_1 and H_2 have the same number of edges. It follows that f is an isomorphism that satisfies the requirements of the theorem. ■

For any graph G, consider the graphs $H_1=2G$ (i.e., the disjoint union of two copies of G) and $H_2=G\times K_2$. If G is not bipartite, then these graphs are non-isomorphic, as H_2 is bipartite but H_1 is not. Let us find the isomorphism of $H_1\times K_2$ and $H_2\times K_2$: For a vertex v of G, let v^0 and v^1 be the corresponding vertices in $2G$. Then the function f is defined by $f(\langle v^i,j\rangle)=\langle\langle v,(i+j) \bmod 2\rangle,j\rangle$, and f satisfies the assumptions of Theorem 13. Therefore, H_1 and H_2 cannot be distinguished using just bipartite graphs. Let us now consider non-bipartite graphs.

Theorem 14. *Let M be a non-bipartite graph, G a graph and $\pi_2: V(G)\to M$ a homomorphism from G to M. Then there exists at most one graph H (up to isomorphism) with the following property: there exists an isomorphism $f: V(G)\to V(H\times M)$ such that $\pi_2=\pi_2^{H\times M} f$.*

Proof. If no such graph H exists, then the claim is true, hence assume that such a graph H and an isomorphism f exist. Let $m_1m_2\dots m_k$ be an odd cycle in M, and let $V_i\subseteq V(G)$ consist of the vertices v such that $\pi_2(v)=m_i$, for $1\le i\le k$. Note that $|V_i|=|V(H)|$ for each i. Let $N_i(x)=\{y\in V_i|\{x,y\}\in E(G)\}$, for each $x\in V(G)$ and $1\le i\le k$. For two vertices $x_1,x_2\in V_i$, let us write $x_1\equiv x_2$ if $N_{i+1}(x_1)=N_{i+1}(x_2)$ (where $N_{k+1}(x)=N_1(x)$). Observe that $x_1\equiv x_2$ if and only if $\pi_1^{H\times M}(f(x_1))$ and $\pi_1^{H\times M}(f(x_2))$ are twins in H. Let U_i be the partition of V_i to the classes of equivalence of $\equiv$. Let P_i be the perfect matching between U_i and U_{i+2} (where $U_{k+1}=U_1$ and $U_{k+2}=U_2$) such that if $\{X,Y\}\in P_i$, then $N_{i+1}(x)=N_{i+1}(y)$ for each $x\in X$ and $y\in Y$. The graph $P=\bigcup_{i=1}^k P_i$ is a union of vertex-disjoint k-cycles. Let Q be a perfect matching between V_1 and V_2 such that if $xy\in Q$, then the $\equiv$-class of x and the $\equiv$-class of y belong to the same cycle of P. Let H' be the subgraph of G induced by $V_1\cup V_2$. Observe that the graph obtained from H' by identifying the pairs of vertices matched by Q is isomorphic to H. We conclude that H is determined uniquely up to isomorphism. ■

Theorems 13 and 14 together imply that any class $\mathcal{A}_{\leq M}$ for non-bipartite M is distinguishing.

5. CONCLUSIONS

There are many other classes of graphs that might be interesting to study. One natural example is graphs with bounded maximum degree. Lovsz and Szegedy [9] found series-parallel *connectors* (2-labeled quantum graphs with labels on distinct vertices, equivalent to $J_{1,2}$) for number of homomorphisms into any graph. If contractors with bounded maximum degree exist, then the graphs with bounded maximum degree distinguish all graphs.

Other possibility is to consider directed graphs. In particular, Theorem 13 is true also for directed graphs (and more generally, for arbitrary finite relational structures), but the characterization similar to Theorem 14 seems harder to obtain.

Finally, one might consider determining some other properties of graphs using numbers of homomorphisms. For example, since $2K_3$ and C_6 cannot be distinguished using bipartite graphs, it is not possible to determine whether a graph is connected or not, or whether a graph is bipartite or not, using bipartite graphs alone. Somewhat curiously, given a connected graph, it is possible to determine whether it is bipartite using only paths and even cycles (in limit, connected bipartite graphs have twice as large probability that a walk of even length starts and ends in the same vertex). Of course, it is also possible to determine whether a graph is bipartite using just odd cycles. Is it possible to determine whether a graph is connected using only paths and cycles?

ACKNOWLEDGMENTS

I would like to thank Jaroslav Neetil for pointing my attention to the problem and for useful discussions about it. I am also indebted to the referees for valuable suggestions regarding the presentation of the results.

REFERENCES

[1] H. L. Bodlaender, A tourist guide through treewidth, Acta Cybernet 11 (1993), 1–21.

[2] J. Cai, M. Frer, and N. Immerman, An optimal lower bound on the number of variables for graph identification, Combinatorica 12(4) (1992), 389–410.

[3] P. J. Cameron, 6-transitive graphs, J Combin Theory B 28 (1980), 168–179.

[4] V. Dalmau and P. Jonsson, The complexity of counting homomorphisms seen from the other side, Theoret Comput Sci 329 (2004), 315–323.
[5] S. Fisk, Distinguishing graphs by the number of homomorphisms, Discuss Math Graph Theory 15(1) (1995), 73–75.
[6] F. Gurski and E. Wanke, The tree-width of clique-width bounded graphs without $K_{n,n}$, Graph-theoretic Concepts in Computer Science, Lecture Notes in Computer Science 1928, Konstanz, Springer, Berlin, 2000, pp. 196–205.
[7] J. J. Gol'fand and M. H. Klin, On k-homogeneous graphs (in Russian), Algorithmic Investigations in Combinatorics, Nauka, Moscow, 1978, pp. 76–85.
[8] L. Lovsz, Operations with structures, Acta Math Hungar 18 (1967), 321–328.
[9] L. Lovsz and B. Szegedy, Contractors and connectors in graph algebras, J Combin Theory B, to appear.
[10] J. Neetil and P. Ossona de Mendez, Linear time low tree-width partitions and algorithmic consequences, STOC'06, Proceedings of the 38th Annual ACM Symposium on Theory of Computing, ACM Press, New York, 2006, pp. 391–400.
[11] B. Weisfeiler, On Construction and Identification of Graphs, Lecture Notes in Mathematics 558, Springer, New York, 1976.
[12] B. Weisfeiler and A. A. Lehman, A reduction of a Graph to a Canonical Form and an Algebra Arising during this Reduction (in Russian), Nauchno—Technic Inf, Ser 2 9 (1968), 12–16.