On Recognizing Graphs by Numbers of Homomorphisms

Zdeněk Dvořák

Charles University, Faculty of Mathematics and Physics,
Institute for Theoretical Computer Science (ITI) [1]
Malostranské nám. 2/25, 118 00, Prague, Czech Republic
rakdver@kam.mff.cuni.cz

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 $Hom(., H)$ from all graphs uniquely determines the graph H upto isomorphism. We study this function restricted to smaller classes of graphs. We show that several natural classes (2-degenerated graphs and non-bipartite graphs with bounded chromatic number) are sufficient to recognize all graphs, and provide description of graph properties that are recognizable by other classes (graphs with bounded tree-width and clique-width).

We consider simple undirected graphs without loops and multiple edges, unless specified otherwise. 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. Sometimes we use the empty graph Z (without any vertices and edges). For each H, $Hom(Z, H) = 1$. Let $\mathcal{G}_{\leq H}$ be the class of graphs G that have a homomorphism to H, i.e., such that $Hom(G, H) > 0$.

Lovász [1] has proved that the function $Hom(., H)$ uniquely determines the graph H upto isomorphism, i.e., that if we know the number of isomorphisms from each graph in $\mathcal{A}$ to H, we can uniquely reconstruct the graph H. In fact, the proof of this statement implies that knowledge of $Hom(., H)$ from all graphs on at most $|V(H)|$ vertices is sufficient. We are interested in what graphs and graph properties can be recognized using smaller classes of graphs (independent on H).

[1]Supported as project 1M0545 by the Czech Ministry of Education

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 $Hom(G, H_1) \neq Hom(G, H_2)$. 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 $Hom(., H)$ restricted to $\mathcal{G}$ determines whether H has the property P or not. For example, the Lovász's result [1] shows that the class $\mathcal{A}$ distinguishes all pairs of non-isomorphic graphs. We inverstigate whether smaller classes of graphs (e.g., graphs with bounded tree-width, chromatic number, etc.) are sufficient to distinguish all graphs. We call such classes *distinguishing*. Fisk [6] studied a related problem—he considered G to distinguish between 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.

A simple demonstration of the concepts is the following observation:

Observation 1 *For each H, the class $\mathcal{G}_{\leq H}$ distinguishes all pairs of non-isomorphic graphs in $\mathcal{G}_{\leq H}$.*

Proof. Let $H_1, H_2 \in \mathcal{G}_{\leq H}$ be the graphs we want to distinguish. Let G be an arbitrary graph. If $G \in \mathcal{G}_{\leq H}$, we know both $Hom(G, H_1)$ and $Hom(G, H_2)$. On the other hand, if $G \notin \mathcal{G}_{\leq H}$, then $Hom(G, H_1) = Hom(G, H_2) = 0$. Therefore, we know the functions $Hom(., H_1)$ and $Hom(., H_2)$ completely, and we are able to determine whether H_1 and H_2 are isomorphic by the result of Lovász [1]. $\square$

Note however that $\mathcal{G}_{\leq H}$ does not necessarily determine whether a graph belongs to $\mathcal{G}_{\leq H}$ or not. For example C_6 and $2K_3$ cannot be distinguished using only bipartite graphs, as we will show in Section 5.

We use the ideas of [2] intensively, and we also use a similar notation. The set $\{1, 2, \ldots, k\}$ is denoted by $[k]$. A *k-labeled graph* is a graph G together with a partial function $f : [k] \to V(G)$, i.e., we assign labels between 1 and k (but not necessarily all of them) to some (not necessarily distinct) vertices of G. The *set of labels* of G is the set of i such that $f(i)$ is defined. For a k-labeled graph G, let base(G) be the same graph without labels. Suppose G and H are k-labeled graphs such that the set of labels of H is a superset of the set of labels of G. We define $Hom(G, H)$ as the number of homomorphisms from G to H that also preserve the labels, i.e., the vertex of G with label i is mapped to the vertex of H with the same label i.

A *(k-labeled) quantum graph* is a formal linear combination with real coefficients of (k-labeled) graphs. We extend the functions $Hom(., H)$ linearly to quantum graphs. Note that if there exists a linear combinations of graphs from class $\mathcal{G}$ that distinguish two graphs H_1 and H_2, then there also exists a graph in $\mathcal{G}$ that distinguishes them. Indeed, at least one of the graphs whose coefficient in the linear combination is nonzero must distinguish H_1 from 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. The sets of labels of G_1 and G_2 do not have to be the same, in particular, if they are disjoint, the product of G_1 and G_2 is just their disjoint union. 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 $\sum_{i,j} \alpha_{1,i}\alpha_{2,j}G_{1,i}G_{2,j}$. In this case, we remove the graphs with loops from this linear combination – this operation preserves the value of $Hom(., H)$ for each loop-less graph 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 each H. We write G^k for product of k copies of a graph G.

1 Complexity remarks

An important open question of the 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 proof that they are non-isomorphic (the graph G such that $Hom(G, H_1) \neq Hom(G, H_2)$). This proof has a polynomial size (it has at most as many vertices as H_1 and H_2 do). 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 proof.

However, for some classes of graphs $\mathcal{G}$ (e.g., graphs with bounded tree-width), it is possible to determine $Hom(G, H)$ in polynomial time for each $G \in \mathcal{G}$. We might thus hope that all graphs can be distinguished by some such $\mathcal{G}$, which (assuming that the graph in $\mathcal{G}$ distinguishing H_1 and H_2 would have polynomial size) would prove that graph non-isomorphism is in NP.

Of course, this turned out not to be the case. The classes of graphs we studied for that the number of homomorphisms can be determined in polynomial time (graphs with bounded tree-width) do not distinguish all non-isomorphic graphs, as we show in the following section. In fact, in the cases we have studied, such a polynomial time algorithm is the base of the proof that the class is not distinguishing.

2 Graphs with Bounded Tree-Width

A graph G has *tree-width* at most k if there exists a tree T whose vertices are subsets of G of size at most $k+1$, satisfying the following conditions:

- each edge e of G is a subset of a vertex of T, and
- vertices of T that contain a vertex v of G induce a connected subgraph in T.

In particular, graphs with tree-width at most 1 are forests. Let us state an equivalent definition that is more suitable for our purposes, and that also extends to labeled graphs. A $(k+1)$-labeled graph G has tree-width at most k if G is an arbitrary graph such that each of the vertices of G has at least one label (thus G has at most $k+1$ vertices), or if G can be obtained by a finite sequence of the following operations:

- If G_1 and G_2 are $(k+1)$-labeled graph with tree-width at most k, then G_1G_2 has tree-width at most k as well.
- If G is a $(k+1)$-labeled graph with tree-width at most k, then a graph obtained by removing the labels from some of the vertices of G has tree-width at most k as well.

It is easy to see that 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 are unions of complete graphs of the same size, their complements, C_5 and the line graph of $K_{3,3}$ (proved by Cameron [4] and independently by Gol'fand [5]). 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.

A *degree refinement* of a graph H is coloring of vertices of H by k distinct vectors $w_i = (n^i_1, n^i_2, \ldots, n^i_k)$ such that for each $1 \leq i, j \leq k$, a 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 permutation of colors. It is often used in algorithms for determining isomorphism of graphs, since there is a polynomial time algorithm that determines the degree refinement in a canonical form that enables to recognize two graphs with different degree refinements. Also, an isomorphism of graphs H_1 and H_2 must preserve colors of the canonical degree refinement.

The concept of the degree refinement can be extended to the classification of k-tuples of vertices, see e.g. [3]. We call this a k-degree refinement. 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. We use the following reformulation that is easier to work with.

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 relations symbols $=$ and E (adjacency), logical connectives $\wedge$, $\vee$ and $\neg$, and quantifiers $\exists$, $\forall$, and $\exists_t$. Note that the variables may be "reused", e.g., formula $(\forall x_1)\,(\exists x_2)\,\big(E(x_1,x_2) \wedge (\exists x_1)\,(x_1 \neq x_2 \wedge \neg E(x_1,x_2))\big)$ 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 in φ. 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 all labels of free variables of φ are present in H. We write $H \models \varphi$ if

- φ is $x_i = x_j$ and the labels i and j are on the same vertex in H.
- φ is $E(x_i, x_j)$ and the labels i and j are on adjacent vertices in H.
- φ is $\varphi_1 \wedge \varphi_2$, and $H \models \varphi_1$ and $H \models \varphi_2$.
- φ is $\varphi_1 \vee \varphi_2$, and $H \models \varphi_1$ or $H \models \varphi_2$.
- φ is $\neg\varphi_1$, and $H \not\models \varphi_1$.
- φ is $(\exists x_i)\,\varphi_1$, and there exists a vertex v of H such that if H_1 is obtained from H by moving label i to v, then $H_1 \models \varphi_1$.
- φ is $(\forall x_i)\,\varphi_1$, and for each vertex v of H, if H_1 is obtained from H by moving label i to v, then $H_1 \models \varphi_1$.
- φ is $(\exists_t) x_i\,\varphi_1$, and there exists at least t vertices v of H such that if H_1 is obtained from H by moving label i to v, then $H_1 \models \varphi_1$.

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 only by working with k-tuples of vertices of H. The following theorem was proved in [3]:

Theorem 2 *Two graphs H_1 and H_2 have the same k-degree refinement if and only if for each closed formula $\varphi \in \mathcal{C}_{k+1}$, $H_1 \models \varphi$ iff $H_2 \models \varphi$.*

Let us start with the following lemma that basically states that number of homomorphisms from a graph with bounded tree-width can be described by a formula with a few variables.

Lemma 3 *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 whose set of labels is superset of the set of labels of G, $H \models \varphi$ 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 such that each vertex of G has a label. The set of labels of H must be a superset of the set of labels of G, so that $Hom(G, H)$ is defined. If $m > 1$, then φ is $false$. If $m = 1$, then the formula φ is just a conjunction of terms $E(x_i, x_j)$ for each two labels i and j of G such that the corresponding vertices are adjacent in G and $x_i = x_j$ for each two labels i and j that appear on the same vertex in G. If $m = 0$, then φ is negation of this conjunction.

Suppose now that G is obtained from G' by removing a label i. Let φ_n be a formula such that $H' \models \varphi_n$ iff $Hom(G', H') = n$ for each H' that satisfies assumptions of this lemma. If $m > 0$, then let $m = \sum_i c_i m_i$ be a decomposition of m such that the numbers c_i are positive integers and the numbers m_i are distinct positive integers, and set $c = \sum_i c_i$. For this decomposition we construct a formula $(\exists_{!c} x_i)\, \neg\varphi_0 \wedge \bigwedge_i (\exists_{!c_i} x_i)\, \varphi_{m_i}$. The formula φ is a disjunction of such sub-formulas for all possible decompositions of m. In case $m = 0$, we let $\varphi = (\forall x_i)\, \varphi_0$.

Finally, suppose that $G = G_1 G_2$, hence

$$Hom(G, H) = Hom(G_1, H) Hom(G_2, H).$$

Let φ_n^i be the formula such that $H \models \varphi_n^i$ iff $Hom(G_i, H) = n$. If $m \neq 0$, then the formula φ is a 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 $m = m_1 m_2$. If $m = 0$, then $\varphi = \varphi_0^1 \vee \varphi_0^2$. □

We now want to prove that we can simulate formulas with few variables by number of homomorphisms from a graph with bounded tree-width. We first need the following observation (made for series-parallel graphs in [2], Claim 4.1).

Observation 4 *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 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 S be the set of labels of G. Let $p(x) = \sum_{i=0}^{k} a_i x^i$ be a polynomial such that $p(x) = 0$ for $x \in X_0$ and $p(x) = 1$ for $x \in X_1$. We set $G[X_0, X_1] = \sum_{i=0}^{k} a_i G^i$, where G^0 is the edge-less graph on $|S|$ vertices, with vertices labeled with elements of S. □

We say that a (labeled) quantum graph G *models* φ *for graphs of size* n if the labels of G correspond to the free variables of φ, and for each graph H on n vertices, if $H \models \varphi$ then $Hom(G, H) = 1$ and if $H \not\models \varphi$ then $Hom(G, H) = 0$.

Lemma 5 *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 $\varphi = (x_i = x_j)$, then we let G be a graph with a single vertex with labels i and j. If $\varphi = E(x_i, x_j)$ and $i \neq j$, then we set $G = K_2$ with one vertex with label i and the other one j. If $i = j$, we let $G = 0$, since H is loop-less and this predicate can never be satisfied.

If $\varphi = \varphi_1 \wedge \varphi_2$, G_1 models φ_1 and G_2 models φ_2, then $G_1 G_2$ models φ. Similarly, $(G_1 + G_2)[\{0\}, \{1, 2\}]$ models $\varphi_1 \vee \varphi_2$, and $G_1[\{1\}, \{0\}]$ models $\neg\varphi_1$.

If $\varphi = (\exists x_i)\, \varphi_1$, and G_1 models φ_1, then let G_1' be the graph G_1 without the label i. The graph $G_1'[\{0\}, \{1, 2, \ldots, n\}]$ models φ. 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 6 *The following two conditions are equivalent:*

1. *There exists a closed formula $\varphi \in \mathcal{C}_{k+1}$ such that $H_1 \models \varphi$ and $H_2 \not\models \varphi$.*
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) $\implies$ (2). If $|V(H_1)| \neq |V(H_2)|$, then $G = K_1$. Otherwise, if $n = |V(H_1)| = |V(H_2)|$, then let G' be a quantum graph that models φ for graphs of size n. That means that $Hom(G', H_1) = 1$ and $Hom(G', H_2) = 0$. Therefore, there exists a graph G in the formal linear combination that defines G' such that $Hom(G, H_1) \neq Hom(G, H_2)$.

Now let us prove that (2) $\implies$ (1). By Lemma 3, there exists a formula $\varphi \in \mathcal{C}_{k+1}$ such that $H \models \varphi$ if and only if $Hom(G, H) = Hom(G, H_1)$. Thus, $H_1 \models \varphi$ and $H_2 \not\models \varphi$. □

Cai et al.[3] have proved that for each k, there exists non-isomorphic graphs H_1 and H_2 such that for each $\varphi \in \mathcal{C}_{k+1}$, $H_1 \models \varphi$ if and only if $H_2 \models \varphi$. These graphs thus cannot be distinguished using only graphs of tree-width at most k.

3 Graphs with Bounded Clique-Width

Clique-width is another width parameter of graphs, in some sense stronger than tree-width (a graph with bounded tree-width has also a bounded clique-width). We show that concerning recognition of graphs, graphs with bounded clique-width are not significantly more powerful than graphs with bounded tree-width.

The clique-width is defined via a construction that manipulates labels of vertices. In order to avoid confusion with the labels that we defined in the previous section, we call the labels used during construction of a graph with bounded clique-width *marks*. Unlike labels, each vertex has precisely one mark, and several vertices may have the same mark.

Definition 1 *A graph G in that each vertex has a mark in $[k]$ has clique-width at most k if G is a single vertex, or if G can be obtained by a finite sequence of the following operations:*

- *If marked graphs G_1 and G_2 have clique-width at most k, then their disjoint union also has clique-width at most k.*
- *If a marked graph G' has clique-width at most k, and $i, j \in [k]$, then the graph obtained from G' by changing all marks i to j has clique-width at most k.*
- *If a marked graph G' has clique-width at most k, and $i \neq j \in [k]$, and G'' is the graph obtained from G' by adding all edges $\{u, v\}$ (that are not already present in G') such that mark of u is i and mark of v is j, then G'' has clique-width at most k.*

For example, graphs of clique-width 1 are exactly the disjoint unions of cliques, and complete bipartite graphs have clique-width 2. Similarly as for the tree-width, there are many problems that are NP-complete in general, but become polynomial when restricted to graphs with bounded clique-width. In particular, determining the chromatic number for graphs with bounded clique-width is polynomial. But it is hard to determine $Hom(G, H)$ even if G is a graph with clique-width one (since determining the size of maximal clique in a graph H is NP-complete). If we restrict H to be triangle-free, this argument fails. Still, determining the size of the maximal bipartite clique in a bipartite graph H is NP-complete, which shows that determining $Hom(G, H)$ remains hard even if H is bipartite and the clique-width of G is at most two. However, if we restrict our attention to the case when H has girth at least five, the number can be determined in polynomial time. To simplify the argumentation a bit, we state the following equivalent definition of a graph with bounded clique-width.

Definition 2 *A graph G in that each vertex has a mark (not necessarily in $[k]$) has has clique-width at most k if G is a single vertex, or if G can be obtained by a finite sequence of the following operations:*

- *If marked graphs G_1 and G_2 have clique-width at most k and their sets of marks are disjoint, then their disjoint union also has clique-width at most k.*
- *If a marked graph G' has clique-width at most k, and i and j are marks, then the graph obtained from G' by changing all marks i to j has clique-width at most k.*
- *If a marked graph G' has clique-width at most k, all its marks belong to $[k]$, $i \neq j \in [k]$, and G'' is the graph obtained from G' by adding all edges $\{u, v\}$ such that mark of u is i and mark of v is j, then G'' has clique-width at most k.*

Given a construction of a graph according to Definition 1, we obtain a construction according to Definition 2 by changing marks of G_2 from i to $k + i$ before each disjoint union operation and by changing them back after it. Thus we need only marks in $[2k]$, and the new construction will be longer by at most $2k|V(G)|$ operations.

Theorem 7 *If H is a graph with girth at least 5 and G is a graph of clique-width at most k, then $Hom(G, H)$ can be determined in time $O(|H|^{2k}|G|k^2)$, assuming that the construction of G that witnesses its clique-width is given.*

Proof. If we add edges between vertices with marks i and j during the construction of G, we force the subgraph induced by them to map to a complete bipartite subgraph of H. Since the girth of H is at least five, this subgraph must be a star, and thus vertices with marks i or j must map to a single vertex x, while the ones with the other mark map to neighbors of x. Given a graph G' of clique-width at most k and a homomorphism f from G' to H, we let the *signature* s of this homomorphism be a function that assigns the following values to the marks:

- If there is no vertex with mark i in G, then $s(i) = \emptyset$.
- If f maps all vertices with mark i to a single vertex x, then $s(i) = (=, x)$.
- If f does not map vertices with mark i to a single vertex, but maps them to to neighbors of a single vertex x, then $s(i) = (E, x)$. Note that x is determined uniquely, because of the girth of H.
- Otherwise, $s(i) = \star$.

Throughout the construction of G (according to Definition 2, using marks in $[2k]$), we count the number of homomorphisms to H for each possible signature. The total number of homomorphisms is then determined by summing the numbers over all signatures. Determining these numbers for graphs with a single vertex is trivial. Let us consider the operations by that G is constructed:

- If G is a disjoint union of G_1 and G_2, then we go through all signatures s for G. Let s_1 be the signature for G_1 such that $s_1(i) = s(i)$ for marks used in G_1 and $s_1(i) = \emptyset$ otherwise, and let s_2 be the signature for G_2 obtained in the same way. Suppose that there are n_1 homomorphisms from G_1 to H with signature s_1, and n_2 homomorphisms from G_2 with signature s_2. The number of homomorphism from $G_1 \cup G_2$ to H with signature s is $n_1 n_2$.
- If G is obtained from G' by changing marks i to j, then the set of homomorphisms does not change. If a homomorphism from G' to H used to have signature s', it has the signature s obtained in the following way after the change: $s(t) = s'(t)$ for $t \neq i, j$, $s(i) = \emptyset$, and

  $s(j) = s'(j)$ if $s'(i) = \emptyset$, or $s'(i) = s'(j)$, or $s'(j) = (E, x)$ and $s'(i) = (=, y)$ and x and y are adjacent,

  $s(j) = s'(i)$ if $s'(j) = \emptyset$, or $s'(j) = (=, x)$ and $s'(i) = (E, y)$ and x and y are adjacent,

$s(j) = (E, x)$ if $s'(i) = (=, y)$ and $s'(j) = (=, z)$ and y and z ($y \neq z$) are neighbors of x,

$s(j) = \star$ otherwise.

Thus it suffices to go over the signatures s' and add the numbers of homomorphisms to the signatures s obtained in this way.

- Suppose that G is obtained from G' by adding edges between vertices with marks i and j. If there are no vertices with marks i or j, this does not change the graph and does not affect numbers of homomorphisms. Otherwise, we need to set to zero the numbers of homomorphisms for all signatures except for those with $s(i) = (=, x)$ and $s(j) = (E, x)$ for some x, or $s(i) = (E, x)$ and $s(j) = (=, x)$ for some x, or $s(i) = (=, x)$ and $s(j) = (=, y)$ for some adjacent vertices x and y.

If S is the number of possible signatures, then each of the operations can be performed in time $O(kS)$. Since the number of signatures S is at most $O(|H|^{2k})$, the time complexity of determining the number of homomorphisms is at most $O(k|H|^{2k}t)$, where t is number of the operations during the construction of G. We can assume that $t \leq k|G|$, thus the time complexity is $O(k^2|H|^{2k}|G|)$. □

This algorithm is also a basis for the following result. A *partial signature* p is a function that assign to each mark (from $[2k]$) one of the symbols $\emptyset$, $=$, E or $\star$. Given a $2k$-labeled graph H, a *completion* of the partial signature p is the signature $\overline{p}$ such that if $p(i) = \emptyset$ or $p(i) = \star$, then $\overline{p}(i) = p(i)$, otherwise $\overline{p}(i) = (p(i), x)$, where x is the vertex of H with label x.

Lemma 8 *If G is a graph of clique-width at most k, p a partial signature and n is a positive integer, then there exists a formula $\varphi \in \mathcal{C}_{2k+2}$ such that for each $2k$-labeled graph H of girth at least 5, $H \models \varphi$ iff the number of homomorphisms from G to H with signature $\overline{p}$ is exactly n.*

Proof. The operations described in the proof of Theorem 7 only need the knowledge of neighborhoods of the vertices selected by each signature, and thus they can be emulated using the formulas in that x_i stands for the vertex selected by the signature for vertices with mark i. The additional variables x_{2k+1} and x_{2k+2} are needed to express facts of type "x_i has two distinct neighbors x_{2k+1} and x_{2k+2} such that the number of homomorphisms from G_1 with signatures s_1 such that $s_1(i) = (=, x_{2k+1})$ is n_1 and the number of homomorphisms from G_2 with signatures s_2 such that $s_2(i) = (=, x_{2k+2})$ is n_2." that are needed to emulate changing marks. □

Theorem 9 *Suppose that H_1 and H_2 are graphs of girth at least 5, such that for each closed formula $\varphi \in \mathcal{C}_{2k+2}$, $H_1 \models \varphi$ if and only if $H_2 \models \varphi$. The graphs H_1 and H_2 cannot be distinguished using graphs of clique-width at most k.*

Proof. Suppose that G is a graph of clique-width at most k such that $Hom(G, H_1) \neq Hom(G, H_2)$. We may assume that all vertices of G have mark 1. G must contain at least one edge, thus it can map neither to a single vertex nor to a neighborhood of a single vertex in H_1 or H_2. Thus $Hom(G, H_i)$ (for $i = 1, 2$) is equal to the number of homomorphisms from G to H_i with (partial) signature $s(1) = \star$, $s(t) = \emptyset$ for $t > 1$. This contradicts Lemma 8. □

Let us take two arbitrary non-isomorphic graphs H_1' and H_2' that cannot be distinguished using $\mathcal{C}_{4k+4}$, and construct graphs H_1 and H_2 by subdividing each edge of H_1' and H_2' with a vertex. The graphs H_1 and H_2 cannot be distinguished using $\mathcal{C}_{2k+2}$, since given a formula $\varphi \in \mathcal{C}_{2k+2}$ that speaks about H_1 and H_2, we can construct an equivalent formula $\varphi' \in \mathcal{C}_{4k+4}$ that speaks about H_1' and H_2'. We do that by introducing a pair of variables x_i' and x_i'' for each variable x_i of φ. If x_i is an original vertex of H_1' or H_2', then $x_i' = x_i'' = x_i$, if x_i is a vertex that splits an edge $\{u, v\}$ of H_1' or H_2', then $x_i' = u$ and $x_i'' = v$. The graphs H_1 and H_2 are non-isomorphic and they have girth at least six, thus they cannot be distinguished using only graphs with clique-width at most k.

4 Graphs with Bounded Degeneracy

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

Let K be the 2-labeled graph on two vertices connected by an edge, with one vertex with label 1 and the other one with label 2. A 2-labeled quantum graph C is a *connector* for unlabeled graph H, if for each 2-labeled graph G, $Hom(\text{base}(GC), H) = Hom(\text{base}(GK), H)$. Equivalently, for each choice of labels 1 and 2 in H, $Hom(C, H) = Hom(K, H)$. Trivially, K is a connector. More interestingly, Lovász and Szegedy [2] (Theorem 1.4) have proved that for each H, there exists a connector that is a linear combination of paths with at least three vertices (with the end vertices labeled with 1 and 2).

Lemma 10 *If H_1 and H_2 are arbitrary unlabeled graphs, then there exists a quantum graph C that is a linear combination of paths with at least three vertices, such that C is connector for both H_1 and H_2.*

Proof. Let C be a connector for the disjoint union of H_1 and H_2, such that C is a linear combination of paths, that exists by [2]. The graph C is also a connector for H_1 and H_2—this follows from the definition of the connector and the fact that C is connected. □

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

Lemma 11 *Let G, H_1 and H_2 be graphs such that $Hom(G,H_1) \neq Hom(G,H_2)$. If e is an edge of G, then there exists a graph G' obtained from G by replacing e by a path with at least one inner vertex, such that $Hom(G', H_1) \neq Hom(G', H_2)$.*

Proof. Let C be the linear combination of paths that is a common connector for H_1 and H_2. Let G'' be the quantum graph obtained from G by replacing e with C. By the definition of the connector, $Hom(G'', H_1) = Hom(G, H_1) \neq Hom(G, H_2) = Hom(G'', H_2)$. Therefore, for at least one graph G' with nonzero coefficient in the linear combination G'', $Hom(G',H_1) \neq Hom(G',H_2)$, and G' satisfies the requirements of this 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-degenerated graph G such that $Hom(G, H_1) \neq Hom(G, H_2)$.*

Proof. There exists a graph G' that distinguishes H_1 and H_2. Using Lemma 11, we construct the graph G by replacing each edge of G' by paths with at least one vertex, while preserving that $Hom(G, H_1) \neq Hom(G, H_2)$. The graph G is 2-degenerated, as each subgraph of G that contains at least one edge also contains at least one of the vertices of $V(G) \setminus V(G')$ that have degree at most two. □

Another corollary of Lemma 11 is that there exists a class of graphs with bounded expansion that distinguishes all graphs. This graph parameter was recently introduced and studied by Nešetřil and Mendez [7]. A graph H is *rank r contraction* of a graph G if there exists a set S of vertex disjoint subgraphs of G such that each member of S has diameter at most r, and H is a simple graph obtained from G by contracting all edges of the subgraphs that belong to S (the arising parallel edges are suppressed). For example, the only rank 0 contraction of G is G itself, and a rank 1 contraction is obtained from G by contracting edges of a partial matching. *Maximum average degree* of graph G is the maximum of average degrees over all subgraphs of G. *Greatest reduced average density of rank* r of G (denoted by $\nabla_r(G)$) is the maximum of maximum average degrees over all rank r contractions of G. A class G has *bounded expansion* if for each $G \in \mathcal{G}$, $\nabla_r(G)$ is bounded by a constant c_r for each r. In particular, nontrivial minor closed classes, as well as graphs with bounded maximum degree, have bounded expansion.

Let $\mathrm{sd}(G)$ be the set of graphs that can be obtained from G by subdividing each edge at least $|V(G)|$ times. Consider a class $\mathcal{A}' = \bigcup_{G \in \mathcal{A}} \mathrm{sd}(G)$. This class has bounded expansion, because for each r, there is only a finite number of graphs $G \in \mathcal{A}'$ such that a rank r contraction of G is not 2-degenerated. For any G that distinguishes graphs H_1 and H_2, we can use Lemma 11 to repeatedly subdivide edges of G to obtain a graph $G' \in \mathcal{A}$ that distinguishes H_1 and H_2 as well. Therefore, $\mathcal{A}'$ distinguishes all pairs of graphs.

5 Graphs Homomorphic to a Fixed Graph

In this section, we consider classes $\mathcal{G}_{\leq M}$ of M-colorable graphs, with M fixed. This class is very natural in this context. As we mentioned in the introduction, bipartite graphs (case $M = K_2$) do not distinguish all graphs. The results of the previous section on the other hand show that if M is not bipartite, then $\mathcal{G}_{\leq M}$ distinguishes all graphs (since a sufficiently fine subdivision of every graph is M-colorable, for any fixed non-bipartite graph M). However, it is interesting to derive these results in a more systematic way.

We let $H_1 \times H_2$ denote a categorical product of two graphs, and let $\pi_1^{H_1 \times H_2}$ and $\pi_2^{H_1 \times H_2}$ be the associated projections.

Theorem 13 *If H_1, H_2 and M are graphs, then $Hom(G, H_1) = Hom(G, H_2)$ holds for each $G \in \mathcal{G}_{\leq M}$ if and only if there exists an isomorphism f of $H_1 \times M$ and $H_2 \times M$ such that $\pi_2^{H_1 \times M} = \pi_2^{H_2 \times M} f$.*

Proof. Suppose first that there exists the isomorphism f with the required properties. We construct a bijection between the homomorphisms from G to H_1 and the homomorphisms from G to H_2, thus showing that their numbers are the same. Let c be a fixed homomorphism from G to M. Given a homomorphism g_1 from G to H_1, we define the function g_2 from G to 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, since if uv is an edge of G, then $g_1(u)g_1(v)$ and $c(u)c(v)$ are edges, thus

$\langle g_1(u), c(u)\rangle\langle g_1(v), c(v)\rangle$ is an edge, 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 each $G \in \mathcal{G}_{\leq M}$. We use the idea of [1]. Let $I(H, H_2)$ be the number of homomorphisms g from $H \times M$ to 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,

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

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 m, H_I is homomorphic to M. Thus, $Hom(H_I, H_2) = Hom(H_I, H_1)$, and thus $I(H, H_1) = I(H, H_2)$ 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 $I(H_1, H_1)$, these numbers are both nonzero, 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 $f(\langle v, m\rangle) = \langle g(\langle v, m\rangle), m\rangle$. The function f is obviously injective. It is surjective, since $Hom(K_1, H_1) = Hom(K_1, H_2)$ and thus the graphs H_1 and H_2 have the same number of vertices. 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, thus if f is a homomorphism, it is an isomorphism as well. Suppose that $\langle u, m_1\rangle\langle v, m_2\rangle$ is an edge of $H_1 \times M$. Then m_1m_2 is an edge of M, and since g is a homomorphism, $g(\langle u, m_1\rangle)g(\langle v, m_2\rangle)$ is an edge of H_2. Therefore also $f(\langle u, m_1\rangle)f(\langle v, m_2\rangle)$ is an edge, and f is the isomorphism we look for. □

For example, 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. For a vertex v of G, let v^0 and v^1 be the corresponding vertices in $2G$. Then the function f defined by $f(\langle v^i, j\rangle) = \langle\langle v, (i + j) \bmod 2\rangle, j\rangle$ is an isomorphism of $H_1 \times K_2$ and $H_2 \times K_2$ that satisfies conditions of Theorem 13. Therefore, H_1 and H_2 cannot be distinguished using just bipartite graphs. It remains to show that non-bipartite graphs are sufficient to distinguish all graphs. To do this, we need to prove that if we are given a graph $H \times M$ and the projection $\pi_2^{H \times M}$, then the graph H is uniquely determined.

Theorem 14 *Suppose that M is a non-bipartite graph. If we know $H \times M$ and the projection $\pi_2^{H \times M}$, then the graph H is uniquely determined up to isomorphism.*

Proof. Let $m_1 m_2 \ldots m_k$ be an odd cycle in M. Let x be an arbitrary vertex in $H \times M$ such that $\pi_2(x) = m_1$. A vertex $\langle \pi_1(x), m_3 \rangle$ must have the same set of neighbors y such that $\pi_2(y) = m_2$ as the vertex x, thus we can determine this vertex (up to permutation of twin vertices of H). Similarly, using the neighbors with $\pi_2(y) = m_4$, we can find the vertex $\langle \pi_1(x), m_5 \rangle$, etc. Since k is odd, finally we find the vertex $\langle \pi_1(x), m_2 \rangle$. The graph H is isomorphic to the graph obtained by taking the subgraph of $H \times M$ induced by $\pi_2^{-1}(m_1) \cup \pi_2^{-1}(m_2)$, and identifying the pairs of vertices $\langle \pi_1(x), m_1 \rangle$ and $\langle \pi_1(x), m_2 \rangle$ for each x. □

Theorems 13 and 14 together imply that any class $\mathcal{G}_{\leq M}$ for non-bipartite M is sufficient to distinguish all graphs.

6 Conclusions

There are many other classes of graphs that might be interesting to study. One natural example are graphs with bounded maximum degree. Lovász and Szegedy [2] proved that series-parallel connectors (2-labeled quantum graphs with labels on distinct vertices, equivalent to a single vertex with two labels) exist. 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, 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 only bipartite graphs. 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? One might also ask what classes are sufficient to recognize other graph properties.

Acknowledgements

I would like to thank Jaroslav Nešetřil for pointing my attention to the problem and for useful discusions about it.

References

[1] L. Lovász, *Operations with structures*, Acta Math. Hung. **18** (1967), 321-328.

[2] L. Lovász and B. Szegedy, *Contractors and connectors in graph algebras.*

[3] Jin-yi Cai, M. Fürer, N. Immerman, *An optimal lower bound on the number of variables for graph identification*, Combinatorica **12(4)** (1992),389-410.

[4] P. J. Cameron, *6-Transitive Graphs*, J. Combinat. Theory B **28** (1980), 168-179.

[5] Ja.Ju.Gol'fand, M.H.Klin, *On k-homogeneous graphs* (Russian), Algorithmic Investigations in Combinatorics, Moscow, Nauka (1978), 76-85.

[6] S. Fisk, *Distinguishing graphs by the number of homomorphisms*, Discussiones Mathematicae - graph theory **15** (1995), 73-75.

[7] J. Nešetřil, P. Ossona de Mendez, *The grad of a graph and classes with bounded expansion*, In 7th International Colloquium on Graph Theory. INRIA, 2005.