The Weisfeiler-Leman Dimension of Conjunctive Queries*

ANDREAS GÖBEL, Hasso Plattner Institute, University of Potsdam, Germany
LESLIE A. GOLDBERG, Department of Computer Science, University of Oxford, United Kingdom
MARC ROTH, School of Electronic Engineering and Computer Science, Queen Mary University of London, United Kingdom

A graph parameter is a function f on graphs with the property that, for any pair of isomorphic graphs G_1 and G_2, $f(G_1) = f(G_2)$. The Weisfeiler–Leman (WL) dimension of f is the minimum k such that, if G_1 and G_2 are indistinguishable by the k-dimensional WL-algorithm then $f(G_1) = f(G_2)$. The WL-dimension of f is ∞ if no such k exists. We study the WL-dimension of graph parameters characterised by the number of answers from a fixed conjunctive query to the graph. Given a conjunctive query φ, we quantify the WL-dimension of the function that maps every graph G to the number of answers of φ in G.

The works of Dvorák (J. Graph Theory 2010), Dell, Grohe, and Rattan (ICALP 2018), and Neuen (ArXiv 2023) have answered this question for *full* conjunctive queries, which are conjunctive queries without existentially quantified variables. For such queries φ, the WL-dimension is equal to the treewidth of the Gaifman graph of φ.

In this work, we give a characterisation that applies to all conjunctive queries. Given any conjunctive query φ, we prove that its WL-dimension is equal to the *semantic extension width* $\mathrm{sew}(\varphi)$, a novel width measure that can be thought of as a combination of the treewidth of φ and its *quantified star size*, an invariant introduced by Durand and Mengel (ICDT 2013) describing how the existentially quantified variables of φ are connected with the free variables. Using the recently established equivalence between the WL-algorithm and higher-order Graph Neural Networks (GNNs) due to Morris et al. (AAAI 2019), we obtain as a consequence that the function counting answers to a conjunctive query φ cannot be computed by GNNs of order smaller than $\mathrm{sew}(\varphi)$.

The majority of the paper is concerned with establishing a lower bound of the WL-dimension of a query. Given any conjunctive query φ with semantic extension width k, we consider a graph F of treewidth k obtained from the Gaifman graph of φ by repeatedly cloning the vertices corresponding to existentially quantified variables. Using a modification due to Fürer (ICALP 2001) of the Cai-Fürer-Immerman construction (Combinatorica 1992), we then obtain a pair of graphs $\chi(F)$ and $\hat{\chi}(F)$ that are indistinguishable by the $(k-1)$-dimensional WL-algorithm since F has treewidth k. Finally, in the technical heart of the paper, we show that φ has a different number of answers in $\chi(F)$ and $\hat{\chi}(F)$. Thus, φ can distinguish two graphs that cannot be distinguished by the $(k-1)$-dimensional WL-algorithm, so the WL-dimension of φ is at least k.

CCS Concepts: • **Theory of computation → Logic and databases**; *Complexity theory and logic*; • **Mathematics of computing** → *Graph theory*.

Additional Key Words and Phrases: Weisfeiler Leman Algorithm, Conjunctive Queries, Counting Problems, Graph Homomorphisms, CFI Graphs

86

Authors' addresses: Andreas Göbel, Hasso Plattner Institute, University of Potsdam, Potsdam, Germany; Leslie A. Goldberg, Department of Computer Science, University of Oxford, Oxford, United Kingdom; Marc Roth, School of Electronic Engineering and Computer Science, Queen Mary University of London, London, United Kingdom.

XXXX-XXXX/2024/5-ART86 $15.00
https://doi.org/10.1145/3651587

ACM Reference Format:
Andreas Göbel, Leslie A. Goldberg, and Marc Roth. 2024. The Weisfeiler-Leman Dimension of Conjunctive Queries. 2, 2 (PODS), Article 86 (May 2024), 17 pages. https://doi.org/10.1145/3651587

1 INTRODUCTION

The Weisfeiler–Leman (WL) algorithm [38] and its higher dimensional generalisations [9] are amongst the most well-studied heuristics for graph isomorphism. This algorithm works as follows. For each positive integer k, the k-dimensional WL-algorithm iteratively maps k-tuples of vertices of a graph to multisets of colours. Two graphs G and G' are said to be k-WL-equivalent, denoted $G \cong_k G'$, if this algorithm returns the same vertex colouring for G and G', up to consistently renaming the colours. For the specific case $k = 1$ the WL-algorithm is equivalent to the colour-refinement algorithm, which is a widely used and efficiently-implementable heuristic for graph isomorphism (see e.g. [2, 21]).

In addition to applications to graph isomorphism, recent works have shown that the expressiveness of Graph Neural Networks (GNNs) and their higher order generalisations is precisely characterised by the WL-algorithm [30, 39]. This result has sparked a flurry of research with the objective of determining which graph parameters are invariant on graphs that are indistinguishable by the WL-algorithm [3, 4, 7, 12, 25, 29, 31]. We refer the reader to the survey by Grohe [20] for further reading.

Over the years, surprising alternative characterisations of k-WL-equivalence have been established.

(I) $G \cong_1 G'$ if and only if G and G' are fractionally isomorphic (see [36, 37]).
(II) For each positive integer k, $G \cong_k G'$ if and only if there is no first-order formula with counting quantifiers that uses at most $k + 1$ variables and that can distinguish G and G' [9, 23].
(III) For each positive integer k, $G \cong_k G'$ if and only if, for each graph H of treewidth at most k, the number of graph homomorphisms from H to G is equal to the number of graph homomorphisms from H to G' [15, 18]. This is the characterisation of k-WL-equivalence that will be used in this work (see Definition 19).

The characterisation in (III) has ignited interest in studying the *WL-dimension* of counting graph homomorphisms and of counting related patterns [3, 7, 12, 25, 31].

A graph parameter f is a function from graphs that is invariant under isomorphisms. The WL-dimension of a graph parameter f is the minimum positive integer k such that f cannot distinguish k-WL-equivalent graphs (see Definition 20). Building upon the works of Dvorák [18], Dell, Grohe and Rattan [15], Roberson [33], and Seppelt [35], it has very recently been shown by Neuen [31] that the WL-dimension of the graph parameter that counts homomorphisms from a fixed graph H is exactly the treewidth of H. It is well known that counting homomorphisms is equivalent to counting answers to conjunctive queries without existentially quantified variables (see e.g. [32]); such conjunctive queries are also called *full* conjunctive queries. In this work, we consider all conjunctive queries including those that have existentially quantified variables and we answer the fundamental question: *What is the WL-dimension of the graph parameter that counts answers to fixed conjunctive queries?*
To state our results, we first introduce some central concepts.

1.1 Conjunctive Queries and Semantic Extension Width

A *conjunctive query* φ consists of a set of free variables $X = \{x_1, \dots, x_k\}$ and a set of (existentially) quantified variables $Y = \{y_1, \dots, y_\ell\}$, and is of the form $\varphi(x_1, \dots, x_k) = \exists y_1, \dots, y_\ell : A_1 \wedge \cdots \wedge A_m$, such that each A_i is an atom $R(\vec{z})$ where R is a relation symbol and $\vec{z}$ is a vector of variables in

$X \cup Y$. Since we focus in this work on undirected graphs without self-loops, in our setting there is only one binary relation symbol E, so all atoms are of the form $E(z_1, z_2)$. An *answer* to φ in a graph G is an assignment a from the free variables X to $V(G)$ such there is an assignment h from $X \cup Y$ to $V(G)$ which agrees with a on X and has the property that, for each atom $E(z_1, z_2)$, the image $\{h(z_1), h(z_2)\}$ is an edge of G.

As is common in the literature (see e.g. [10, 11, 16, 32]), we can equivalently express the answers of φ in a graph G as partial homomorphisms to G. Let H be the graph with vertex set $X \cup Y$ that has as edges the pairs of variables in $X \cup Y$ that occur in a common atom. Then the answers of φ in G are the mappings $a\colon X \to V(G)$ that can be extended to a homomorphism from H to G. For this reason, following the notation of [16], we will from now an refer to a conjunctive query as a pair (H, X) where H is a graph and X is a subset of vertices of H corresponding to the free variables. We will say that (H, X) is *connected* if H is a connected graph. We will write $\mathrm{Ans}((H, X), G)$ for the set of answers of (H, X) in G; this is made formal in Section 2.1. The WL-dimension of a conjunctive query (H, X) is the WL-dimension of the graph parameter that maps every graph G to $|\mathrm{Ans}((H, X), G)|$.

Semantic Extension Width. Let (H, X) be a conjunctive query and let $Y = V(H) \setminus X$. The graph $\Gamma(H, X)$ is obtained from H by adding an edge between each pair of vertices $u \neq v$ in X if and only if there is a connected component in $H[Y]$ that is adjacent to both u and v. We then define the *extension width* of (H, X) as the treewidth of $\Gamma(H, X)$; the definition of treewidth can be found in Section 2.2.

The *semantic extension width* of a conjunctive query (H, X), denoted by $\mathrm{sew}(H, X)$ is then the minimum extension width of any conjunctive query (H', X') that is *counting equivalent* to (H, X), i.e., any conjunctive query (H', X') such that, for every graph G, $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H', X'), G)|$. A discussion of counting equivalence and *counting minimal* conjunctive queries can be found in Section 2.1.

Before stating our main result, we provide an example of a conjunctive query and its semantic extension width: Let (S_k, X_k) be the *k-star query*: $X_k = \{x_1, \dots, x_k\}$ and S_k has vertices $X_k \cup \{y\}$ and edges $\{x_i, y\}$ for all $i \in [k]$. Note that the answers of (S_k, X_k) in a graph G are precisely the assignments from X_k to $V(G)$ such that the vertices all of the images of vertices in X_k have a common neighbour. The k-star query is acyclic (i.e., S_k has treewidth 1) and it has played an important role as a base case for complexity classifications regarding counting answers to conjunctive queries [10, 16]. The graph $\Gamma(S_k, X_k)$ is the $(k+1)$-clique which has treewidth k. Since it is also minimal with respect to counting equivalence, we have $\mathrm{sew}(S_k, X_k) = k$.

1.2 Our Contributions

We now state our main result.

Theorem 1. *Let (H, X) be a connected conjunctive query with $X \neq \emptyset$. Then the WL-dimension of (H, X) is equal to its semantic extension width* $\mathrm{sew}(H, X)$.

In Theorem 1, the WL-dimension of (H, x) is captured by its semantic extension width rather than by its extension width, which is the treewidth of $\Gamma(H, X)$. This is because $H[Y]$ may contain a high-treewidth subgraph that does not influence the number of answers.

As an immediate corollary of Theorem 1, we obtain the following alternative characterisation of WL-equivalence.

Corollary 2. *For each positive integer k, two graphs G and G' are k-WL-equivalent if and only if, for each connected conjunctive query (H, X) with $X \neq \emptyset$ and* $\mathrm{sew}(H, X) \leq k$, $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$.

As the following sections show, our classification of the WL-dimension of conjunctive queries has further strong consequences regarding the expressive power of graph neural networks (GNNs), the parameterised complexity of counting answers to conjunctive queries, and the WL-dimension of first-order formulas with universal quantifiers such as the formula corresponding to dominating sets.

GNNs and Conjunctive Queries. Over the last decade, GNNs have received increasing attention due to their application to computations involving graph structured data (see [25]). Motivated by the fact that the number of occurrences of small patterns can capture interesting global information about graphs, and can therefore be used to compare graphs [1, 24, 28], researchers have studied the extent to which GNNs (and their higher order generalisations [30]) are able to count selected small patterns such as homomorphisms [25], subgraphs [7], and induced subgraphs [12].

Following [30] but simplifying the notation for our needs, we represent a t-layer order-k GNN N as a tuple $N = (G, W_1, \dots, W_t, f_0, \dots, f_t)$ where G is a graph, each W_i is a set of weights, and each f_i assigns a feature vector to each k-tuple of nodes. The GNN specifies how f_i is computed from $G, W_1, \dots, W_{i-1}, f_0, \dots, f_{i-1}$. We use $f_N(G)$ to denote the final feature vector so $f_N(G) = f_t$. The feature vector $f_N(G)$ induces a partition on the k-tuples of vertices of G, which we call $P_N(G)$.

We next explain what we mean when we say that a GNN can "compute" a function on graphs. So far, this has been studied in a somewhat limited context. Namely, we say that a GNN can "compute" a function $A_N(G)$ if $A_N(G)$ can be computed in polynomial time from $P_N(G)$. Thus, when we say that a GNN can count small patterns, we mean that the number of such patterns can be efficiently computed from $P_N(G)$. We do not address the issue of whether the GNN can itself do the polynomial-computation that is needed to compute $A_N(G)$ from $P_N(G)$. Issues of dimension are also beyond the scope of this paper — in our setting the feature vector induces a partition on the k-tuples of vertices of G — for a brief discussion about how the dimension can be reduced see [30].

We say that a GNN $N = (G, W_1, \dots, W_t, f_0, \dots, f_t)$ is "fully refined" if there is no GNN $N' = (G, W'_1, \dots, W'_t, f_0, f'_1, \dots, f'_{t'})$ such that $P_{N'}(G)$ strictly refines $P_N(G)$.

In this setting, Morris et al. [30] established an equivalence between the expressive power of fully-refined order-k GNNs and the k-dimensional WL algorithm. For this, let $\mathcal{N}_k$ be the set of fully-refined order-k GNNs. Propositions 3 and 4 of [30] give the following proposition.

Proposition 3. *For all $N \in \mathcal{N}_k$, $P_N(G)$ is exactly the the same as the partition $P_{WL}(G)$ on k-tuples of vertices that is computed by k-WL when it is run with input G and the initial partition induced by the initial feature vector f_0 of N.*

Building upon Proposition 3, the works of Dvorák [18], Dell, Grohe and Rattan [15], and Lanzinger and Barcelo [25] determine the expressiveness of fully refined GNNs in the context of homomorphism counting. Essentially, order-k GNNs can count homomorphisms from a graph H if and only if the treewidth of H is at most k. The "if" direction has already been used implicitly in [15, 18]. It follows explicitly from [25, Theorem 6 and Lemma 7]. The "only if" direction follows by combining Proposition 3 with the upper and lower bounds on the WL dimension of counting homomorphisms [15, 18, 25, 33]. Specifically, Lanzinger and Barcelo [25] show that homomorphisms from H to G can be efficiently computed from the vertex refinement produced when WL-k is run on input G starting from the partition in which each k-tuple is assigned a part based on the subgraph that induces.

Our classification (Theorem 1) provides a similar picture in the context of counting answers to conjunctive queries. First, we will show that if (H, X) is a conjunctive query with $\mathrm{sew}(H, X) = k$ then for all graphs G there is a fully refined GNN $N \in \mathcal{N}_k$ with underlying graph G that computes $|\mathrm{Ans}((H, X), G)|$. This follows from the following two observations.

(1) From [25, Theorem 6] and Proposition 3, for all k, all sequences $F_1, \dots, F_n$ of graphs of treewidth at most k, all sequences $\mu_1, \dots, \mu_k$ of rational numbers, and all graphs G there is a fully refined GNN $N \in \mathcal{N}_k$ with underlying graph G such that $\sum_{i=1}^{n} \mu_i |\mathrm{Hom}(F_i, G)|$ can be efficiently computed from $P_N(G)$.
(2) From our work (see Observation 23), for all graphs G there is a finite sequence of graphs $F_1, \dots, F_n$ of treewidth at most k, such that $|\mathrm{Ans}((H, X), G)|$ can be written as such as sum.

For the other direction we will show that if a fully refined GNN can compute the number of answers from (H, X) then the order of this GNN is at least $\mathrm{sew}(H, X)$. The proof is based on the following two observations.

(1) From Proposition 3, for all graphs G' and G'' such that $G' \cong_k G''$ and all GNNs $N', N'' \in \mathcal{N}_k$ with underlying graphs G' and G'', and any function $A_N(G)$ that is efficiently computable from $P_N(G)$, $A_{N'}(G') = A_{N''}(G'')$.
(2) Let (H, X) be a conjunctive query with $\mathrm{sew}(H, X) = k$. From our Theorem 1, there are graphs G and G' such that $G \cong_{k-1} G'$ and $|\mathrm{Ans}((H, X), G)| \neq |\mathrm{Ans}((H, X), G')|$.

We can use these two facts to show that if a fully refined GNN can compute the number of answers from (H, X) then its order is at least $\mathrm{sew}(H, X)$. In particular, consider (H, X) with $\mathrm{sew}(H, X) = k$. Suppose for contradiction that, for some $j < k$, some GNN $N \in \mathcal{N}_j$ can compute $A_N(G) = |\mathrm{Ans}((H, X), G)|$. For all G and G' with $G \cong_{k-1} G'$ we have $G \cong_j G'$ so from (1), we have $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$, contradicting (2).

Parameterised counting of answers to conjunctive queries. The next consequence of our main result reveals a surprising connection between the complexity of counting answers to conjunctive queries and their WL-dimension. Given a class of conjunctive queries Ψ, the counting problem $\#\mathrm{CQ}(\Psi)$ takes as input a pair consisting of a conjunctive query $(H, X) \in \Psi$ and a graph G and outputs $|\mathrm{Ans}((H, X), G)|$. We say that a class of conjunctive queries has *bounded* WL-dimension if there is a constant B that upper bounds the WL-dimension of all queries in the class. The assumption $\mathrm{FPT} \neq W[1]$ is the central (and widely accepted) hardness assumption in parameterised complexity theory (see e.g. [19]). We say that a conjunctive query is *counting minimal* if it is a minimal representative with respect to counting equivalence (see Definition 9). Theorem 1 implies Corollary 4.

Corollary 4. *Let Ψ be a recursively enumerable class of counting minimal and connected conjunctive queries with at least one free variable. The problem $\#CQ(\Psi)$ is solvable in polynomial time if and only if the WL-dimension of Ψ is bounded; the "only if" is conditioned under the assumption* $\mathrm{FPT} \neq W[1]$.

Quantum Queries and the WL dimension of counting dominating sets. Our main result also enables us to classify the WL-dimension of more complex queries including unions of conjunctive queries and conjunctive queries with disequalities and negations over the free variables. The statement of this classification requires the consideration of finite linear combinations of conjunctive queries (also known as *quantum queries*; see Definition 29). A quantum query is of the form $Q = \sum_{i=1}^{\ell} c_i \cdot (H_i, X_i)$ where, for all $i \in [\ell]$, $c_i \in \mathbb{Q} \setminus \{0\}$. The (H_i, X_i) are connected and pairwise non-isomorphic conjunctive queries where each (H_i, X_i) is counting minimal and $X_i \neq \emptyset$.

It is well known [11, 16] that unions of conjunctive queries, existential positive queries, and conjunctive queries with disequalities and negations over the free variables all have (unique) expressions as quantum queries, that is, the number of answers to those more complex queries can be computed by evaluation the respective quantum query according to the definition $|\mathrm{Ans}(Q, G)| := \sum_{i=1}^{\ell} c_i \cdot |\mathrm{Ans}((H_i, X_i), G)|$. For this reason, understanding the WL-dimension of linear combination of conjunctive queries allows us to also understand the WL-dimension of more complex queries.

Defining the *hereditary semantic extension width* of a quantum query Q, denoted by $\mathsf{hsew}(Q)$, as the maximum semantic extension width of its terms, we obtain the following.

Corollary 5. *The WL-dimension of a quantum query Q is equal to* $\mathsf{hsew}(Q)$.

As a final corollary of our main result we take a look at a concrete graph parameter, the WL-dimension of which was not known so far: the parameter that maps each graph G to the number of size-k dominating sets in G. Here, a dominating set of a graph G is a subset of vertices D of G such that each vertex of G is either contained in D or is adjacent to a vertex in D. With an easy argument, we show that counting dominating sets of size k can be expressed as a linear combination of the k-star queries (S_k, X_k). Using Theorem 1 and Corollary 5, we obtain the following corollary.

Corollary 6. *For each positive integer k, the WL-dimension of the graph parameter that maps each graph G to the number of size-k dominating sets in G is equal to k.*

1.3 Discussion and Outlook

We stated and proved our result for the case of connected conjunctive queries with at least one free variable over graphs. However, our result can easily be extended to the following.

(A) For disconnected queries, the WL-dimension will just be the maximum of the semantic extension widths of the connected components.
(B) If no variable is free, then counting answers of a conjunctive query becomes equivalent to *deciding* the existence of a homomorphism. The WL-dimension of the corresponding graph parameter is equal to the treewidth of the query modulo homomorphic equivalence, which for queries without free variables is the same as semantic extension width. This can be proved along the lines of the analysis of Roberson [33].
(C) Barceló et al. [4], and Lanzinger and Barceló [25] have shown very recently that the WL-algorithm (and the notions of WL-equivalence and WL-dimension) readily extend from graphs to *knowledge graphs*, i.e., directed graphs with vertex labels and edge labels; parallel edges with distinct labels are allowed, but self-loops are not allowed. It is not hard to see that our analysis applies to knowledge graphs as well.

Since the technical content of this paper is already quite extensive, we decided to defer the formal statement and proofs of (A)–(C) to a future journal version.

Finally, extending our results from graphs to relational structures is more tricky, since it is not known yet whether and how WL-equivalence can be characterised via homomorphism indistinguishability from structures of higher arity.[1] However, recent works by Böker [6] and by Scheidt and Schweikardt [34] provide first evidence that homomorphism counts from hypergraphs of bounded generalised hypertreewidth might be the right answer. We leave this for future work.

1.4 Organisation of the Paper

We start by introducing further necessary notation and concepts in Section 2. Afterwards, we prove the upper bound of the WL-dimension in Section 3; the proof of the lower bound is deferred to the full version [22]. Finally, we prove Theorem 1, as well as its consequences, in Section 5.

2 PRELIMINARIES

Given a set S, we write $\mathsf{Bij}(S)$ for the set of all bijections from S to itself. Given a function $f : A \to B$ and a subset $X \subseteq A$, we write $f|_X : X \to B$ for the restriction of f on X. We write π_1 for the

[1] A characterisation for the special case of *constant* arity $r \geq 2$ was recently established independently by Butti and Dalmau [8], and by Dawar, Jakl, and Reggio [14].

projection of a pair to its first component, that is, $\pi_1(a,b) = a$. Given a positive integer ℓ we set $[\ell] = \{1,\ldots,\ell\}$.

All graphs in this paper are undirected and simple (without self-loops and without parallel edges). Given a graph $G = (V,E)$, a vertex $u \in V$ and a subset U of V, $N(u) = \{v \in V \mid \{u,v\} \in E\}$ and $N(U) = \cup_{u \in U} N(u)$. We say that a connected component C of a graph H is *adjacent* to a vertex v of H if there is a vertex $u \in C$ that is adjacent to v. Given a subset S of vertices of a graph G, we write $G[S]$ for the graph induced by the vertices in S.

A *homomorphism* from a graph H to a graph G is a function $h : V(H) \to V(G)$ such that, for all edges $\{u,v\} \in E(H)$, $\{h(u),h(v)\}$ is an edge of G. We write $\mathrm{Hom}(H,G)$ for the set of all homomorphisms from H to G. An *isomorphism* from H to G is a bijection $b : V(H) \to V(G)$ such that, for all $u,v \in V(H)$, $\{u,v\} \in E(H)$ if and only if $\{h(u),h(v)\} \in E(G)$. We say that H and G are *isomorphic*, denoted by $H \cong G$, if there is an isomorphism from H to G. An *automorphism* of a graph H is an isomorphism from H to itself, and we write $\mathrm{Aut}(H)$ for the set of all automorphisms of H.

2.1 Conjunctive Queries

As stated in the introduction, we focus on conjunctive queries on graphs. This allows us to follow the notation of [16].

Definition 7. A *conjunctive query* is a pair (H,X) where H is the underlying graph and X is the set of free variables. When H and X are clear from context we will use Y to denote $V(H) \setminus X$. We say that a conjunctive query (H,X) is *connected* if H is connected.

It is well known (see e.g. [10, 11, 16]) that the set of answers of a conjunctive query in a graph G is the set of assignments from the free variables to the vertices of G that can be extended to a homomorphism.

Definition 8. Let (H,X) be a conjunctive query and let G be a graph. The set of answers of (H,X) in G is given by $\mathrm{Ans}((H,X),G) = \{a : X \to V(G) \mid \exists h \in \mathrm{Hom}(H,G) : h|_X = a\}$.

We say that two conjunctive queries (H_1,X_1) and (H_2,X_2) are *isomorphic*, denoted by $(H_1,X_1) \cong (H_2,X_2)$ if there is an isomorphism from H_1 to H_2 that maps X_1 to X_2.

Throughout this work, we will focus on *counting minimal* conjunctive queries.

Definition 9 (Counting Equivalence and Counting Minimality)**.** We say that two conjunctive queries (H_1,X_1) and (H_2,X_2) are *counting equivalent*, denoted by $(H_1,X_1) \sim (H_2,X_2)$, if for each graph G, $|\mathrm{Ans}((H_1,X_1),G)| = |\mathrm{Ans}((H_2,X_2),G)|$. A conjunctive query is said to be *counting minimal* if it it is minimal (with respect to taking subgraphs) in its counting equivalence class.

It is known that all counting minimal conjunctive queries within a counting equivalence class are isomorphic [11, 16]. If a query has no existential variables so that $X = V(H)$ then counting equivalence is the same as isomorphism. If all variables are quantified so that $X = \emptyset$ then counting equivalence is the same as homomorphic equivalence (also called semantic equivalence).

2.2 Treewidth and Extension Width

We start by introducing tree decompositions and treewidth.

Definition 10. Let H be a graph. A *tree decomposition* of H is a pair consisting of a tree T and a collection of sets, called *bags*, $\mathcal{B} = \{B_t\}_{t \in V(T)}$, such that the following conditions are satisfied:

(T1) For all $v \in V(H)$ there is a bag B_t with $v \in B_t$.
(T2) For all $v \in V(H)$ the subgraph of T induced by the vertex set $\{t \in V(T) \mid v \in B_t\}$ is connected.

(T3) For all $e \in E(H)$, there is a bag B_t with $e \subseteq B_t$.

The *width* of $(T, \mathcal{B})$ is $\max_{t \in V(T)} |B_t| - 1$ and a tree decomposition of minimum width is called *optimal.* The *treewidth* of H, denoted by $\mathrm{tw}(H)$, is the width of an optimal tree decomposition of H. The treewidth of a conjunctive query (H, X), denoted by $\mathrm{tw}(H, X)$, is the treewidth of H.

Next we introduce the extension width of a conjunctive query.

Definition 11 ($\Gamma(H, X)$ and Extension Width)**.** Let (H, X) be a conjunctive query. The *extension* $\Gamma(H, X)$ of (H, X) is a graph with vertex set $V(H)$ and edge set $E(H) \cup E'$, where E' is the set of all $\{u, v\}$ such that $u, v \in X$, $u \neq v$, and there is a connected component of $H[Y]$ which is adjacent to both u and v in H. The *extension width* of of a conjunctive query (H, X) is defined by $\mathrm{ew}(H, X) := \mathrm{tw}(\Gamma(X, H))$.

We will often restrict our analysis to counting minimal conjunctive queries. This requires us to lift the notion of extension width as follows.

Definition 12 (Semantic Extension Width)**.** The *semantic extension width* of a conjunctive query (H, X), denoted by $\mathrm{sew}(H, X)$, is the extension width of a counting minimal conjunctive query (H', X') with $(H, X) \sim (H', X')$.

Note that the semantic extension width is well-defined since all counting minimal (H', X') with $(H, X) \sim (H', X')$ are isomorphic.

2.3 The ℓ-copy $F_\ell(H, X)$

One of the most central operations on conjunctive queries invoked in this work is a cloning operation on existentially quantified variables, defined as follows.

Definition 13 ($F_\ell(H, X)$)**.** Let (H, X) be a conjunctive query and let ℓ be a positive integer. The *ℓ-copy* $F_\ell(H, X)$ is defined as follows. The vertex set of $F_\ell(H, X)$ is $X \cup (Y \times [\ell])$. Let $E_X = \{\{u, v\} \in E(H) \cap X^2\}$, $E_{X,Y} = \{\{u, (v, i)\} \mid u \in X, v \in Y, i \in [\ell], \{u, v\} \in E(H)\}$, and $E_Y = \{\{(u, i), (v, i)\} \mid \{u, v\} \in E(H) \cap Y^2, i \in [\ell]\}$. The edge set of $F_\ell(H, X)$ is $E_X \cup E_{X,Y} \cup E_Y$.

There is a natural homomorphism from $F_\ell(H, X)$ to H which we denote by $\gamma[H, X, \ell]$.

Definition 14. Let (H, X) be a conjunctive query and let ℓ be a positive integer. Define the map $\gamma[H, X, \ell] \colon V(F_\ell(H, X)) \to V(H)$ as follows. $\gamma[H, X, \ell](u) = u$ if $u \in X$ and $\gamma[H, X, \ell](u) = \pi_1(u)$ if $u \in Y \times [\ell]$. We will just write $\gamma = \gamma[H, X, \ell]$ if (H, X) and ℓ are clear from the context.

Observation 15. *The function γ is a homomorphism from $F_\ell(H, X)$ to H.*

Next, we relate the treewidth of the graph $F_\ell(H, X)$ to the extension width of (H, X).

Lemma 16. *Let (H, X) be a conjunctive query and let ℓ be a positive integer. The treewidth of $F_\ell(H, X)$ is at most* $\mathrm{ew}(H, X)$.

Proof. Let $\Gamma = \Gamma(H, X)$ and let $C_1, \dots, C_m$ be the connected components of $H[Y]$. For each $i \in [m]$, let $\delta_i = N(C_i) \cap X$ and let $\hat{C}_i = C_i \cup \delta_i$. Since δ_i is a clique in Γ, there is an optimal tree decomposition $(\mathcal{T}_i, \mathcal{B}_i)$ of $\Gamma[\hat{C}_i]$ with δ_i as a bag. For $j \in [\ell]$, let $(\mathcal{T}_i^j, \mathcal{B}_i^j)$ be a copy of $(\mathcal{T}_i, \mathcal{B}_i)$ where B_i^j is the bag corresponding to δ_i.

Let $(\mathcal{T}_X, \mathcal{B}_X)$ be an optimal tree decomposition of $\Gamma[X]$. Choose $(\mathcal{T}_X, \mathcal{B}_X)$ such that there is a bag $B_{X,i}$ corresponding to each δ_i.

Finally, construct a tree decomposition $(\mathcal{T}, \mathcal{B})$ of $F_\ell(H, X)$ by identifying $B_{X,i}$ and B_i^j for each $i \in [m]$ and $j \in [\ell]$. This tree decomposition shows that $\mathrm{tw}(F_\ell(H, X)) \leq \mathrm{tw}(\Gamma)$. ◻

The following lemma follows implicitly from [5]. For completeness, we included a proof in the full version [22].

Lemma 17. *Let (H, X) be a conjunctive query. There exists a positive integer ℓ such that* $\mathrm{ew}(H, X) \leq \mathrm{tw}(F_\ell(H, X))$.

In combination, Lemmas 16 and 17 provide an alternative characterisation of the extension width, which we will be using for the remainder of the paper.

Corollary 18. *Let (H, X) be a conjunctive query. Then* $\mathrm{ew}(H, X) = \max\{\mathrm{tw}(F_\ell(H, X)) \mid \ell \in \mathbb{Z}_{>0}\}$.

Proof. The corollary follows immediately from Lemmas 16 and 17. □

2.4 Weisfeiler-Leman Equivalence, Invariance and Dimension

In order to make this work self-contained, we will use the characterisation of Weifeiler-Leman (from now on just "WL") equivalence via homomorphism indistinguishability due to Dvorák [18] and Dell, Grohe and Rattan [15]. We recommend the survey of Arvind for a short but comprehensive introduction to the classical definition using the WL-algorithm [2].

Definition 19 (WL-Equivalence)**.** Let k be a positive integer. Two graphs G and G' are *k-WL-equivalent*, denoted by $G \cong_k G'$, if for every graph H of treewidth at most k we have $|\mathrm{Hom}(H, G)| = |\mathrm{Hom}(H, G')|$.

Note that WL-equivalence is monotone in the sense that $G \cong_k G'$ implies that for every $k' \leq k$, $G \cong_{k'} G'$. A graph parameter f is called *k-WL-invariant* if, for every pair of graphs G, G' with $G \cong_k G'$, $f(G) = f(G')$. Observe that, for $k \geq k'$, every k' -WL-invariant graph parameter is also k-WL-invariant. Thus, following the definition of Arvind et al. [3], we define the *WL-dimension* of a graph parameter f as the minimum k for which f is k-WL-invariant, if such a k exists, and ∞ otherwise.

Definition 20 (WL-dimension of conjunctive queries)**.** Let (H, X) be a conjunctive query. The *WL-dimension* of (H, X) is the WL-dimension of the function $G \mapsto |\mathrm{Ans}((H, X), G)|$.

3 UPPER BOUND ON THE WL-DIMENSION

The goal of this section is to prove the following upper bound.

Theorem 21. *Let (H, X) be a conjunctive query. Then the WL-dimension of (H, X) is at most* $\mathrm{ew}(H, X)$.

For the proof of Theorem 21 we will use the following interpolation argument.

Lemma 22. *Let (H, X) be a conjunctive query. Let G_1 and G_2 be graphs. Suppose that, for all positive integers ℓ,* $|\mathrm{Hom}(F_\ell(H, X), G_1)| = |\mathrm{Hom}(F_\ell(H, X), G_2)|$. *Then* $|\mathrm{Ans}((H, X), G_1)| = |\mathrm{Ans}((H, X), G_2)|$.

Proof. Let G be a graph and let $\sigma\colon X \to V(G)$. Define $\mathrm{Ext}(\sigma) = \{\rho\colon Y \to V(G) \mid \sigma \cup \rho \in \mathrm{Hom}(H, G)\}$. Let Ω be the set of functions from Y to $V(G)$ and consider any $\Upsilon \subseteq \Omega$. Define $H^G(\Upsilon) = \{h \in \mathrm{Ans}((H, X), G) \mid \mathrm{Ext}(h) = \Upsilon\}$ and $\hat{H}^G_\ell(\Upsilon) = \{h \in \mathrm{Hom}(F_\ell(H, X), G) \mid \mathrm{Ext}(h|_X) = \Upsilon\}$. First observe that for any $\Upsilon \subseteq \Omega$, $|\hat{H}^G_\ell(\Upsilon)| = |H^G(\Upsilon)| \cdot |\Upsilon|^\ell$. Moreover, $|\mathrm{Ans}((H, X), G)| = \sum_{\emptyset \neq \Upsilon \subseteq \Omega} |H^G(\Upsilon)|$, and $|\mathrm{Hom}(F_\ell(H, X), G)| = \sum_{\emptyset \neq \Upsilon \subseteq \Omega} |\hat{H}^G_\ell(\Upsilon)|$. Now let G_1 and G_2 be graphs with $|\mathrm{Hom}(F_\ell(H, X), G_1)| = |\mathrm{Hom}(F_\ell(H, X), G_2)|$ for all positive integers ℓ.

Let Ω_1 be the set of functions from Y to $V(G_1)$ and let Ω_2 be the set of functions from Y to $V(G_2)$. Let $\hat{n} = \max\{|\Omega_1|, |\Omega_2|\}$. For every positive integer ℓ,

$$\begin{aligned}
&|\mathrm{Hom}(F_\ell(H,X),G_1)| = |\mathrm{Hom}(F_\ell(H,X),G_2)| \\
\Leftrightarrow &\sum_{\emptyset\neq\Upsilon\subseteq\Omega_1} |\hat{H}_\ell^{G_1}(\Upsilon)| - \sum_{\emptyset\neq\Upsilon\subseteq\Omega_2} |\hat{H}_\ell^{G_2}(\Upsilon)| = 0 \\
\Leftrightarrow &\sum_{\emptyset\neq\Upsilon\subseteq\Omega_1} |H^{G_1}(\Upsilon)|\cdot|\Upsilon|^\ell - \sum_{\emptyset\neq\Upsilon\subseteq\Omega_2)} |H^{G_2}(\Upsilon)|\cdot|\Upsilon|^\ell = 0 \\
\Leftrightarrow &\sum_{i=1}^{\hat{n}} i^\ell \cdot \left(\sum_{\substack{\Upsilon\subseteq\Omega_1\\|\Upsilon|=i}} |H^{G_1}(\Upsilon)| - \sum_{\substack{\Upsilon\subseteq\Omega_2\\|\Upsilon|=i}} |H^{G_2}(\Upsilon)| \right) = 0
\end{aligned}$$

Note that this yields a system of linear equations. For each positive integer ℓ, we have the equation $\sum_{i=1}^{\hat{n}} c_i \cdot i^\ell = 0$ where

$$c_i = \left(\sum_{\substack{\Upsilon\subseteq\Omega_1\\|\Upsilon|=i}} |H^{G_1}(\Upsilon)| - \sum_{\substack{\Upsilon\subseteq\Omega_2\\|\Upsilon|=i}} |H^{G_2}(\Upsilon)| \right).$$

The matrix corresponding to this system of equations is a Vandermonde matrix, so it is invertible. Thus $c_i = 0$ for all $i \in \{1,\dots,n\}$. Therefore

$$\begin{aligned}
|\mathrm{Ans}((H,X),G_1)| &= \sum_{\emptyset\neq\Upsilon\subseteq\Omega_1} |H^{G_1}(\Upsilon)| = \sum_{i=1}^{\hat{n}} \sum_{\substack{\Upsilon\subseteq\Omega_1\\|\Upsilon|=i}} |H^{G_1}(\Upsilon)| \\
&= \sum_{i=1}^{\hat{n}} \sum_{\substack{\Upsilon\subseteq\Omega_2\\|\Upsilon|=i}} |H^{G_2}(\Upsilon)| = \sum_{\emptyset\neq\Upsilon\subseteq\Omega_2} |H^{G_2}(\Upsilon)| \\
&= |\mathrm{Ans}((H,X),G_2)|.
\end{aligned}$$

□

The proof of Lemma 22 immediately implies the following observation, Observation 23. Note that the graphs $F_\ell(H,X)$ that are referred to in Lemma 22 have treewidth at most $\mathrm{ew}(H,X)$ by Lemma 16. In Observation 23 there are two possibilities. If we start with a query (H,X) that is counting minimal, we can apply directly the proof of Lemma 22. Otherwise, we apply the proof of Lemma 22 to a counting-equivalent counting-minimal query.

Observation 23. *Let (H,X) be a conjunctive query of semantic extension width k and let G be a graph. There is a finite sequence of graphs $F_1,\dots,F_n$ of treewidth at most k, such that $|\mathrm{Ans}((H,X),G)|$ can be computed via Gaussian elimination from the homomorphism counts $|\mathrm{Hom}(F_\ell,G)|$ for $\ell \in \{1,\dots,n\}$.*

Proof of Theorem 21. Let (H,X) be a conjunctive query. Let $k = \mathrm{ew}(H,X)$. We wish to show that the WL-dimension of (H,X) is at most k which is equivalent to showing that the function $G \mapsto |\mathrm{Ans}((H,X),G)|$ is k-WL invariant. To do this, we show that, for any pair of graphs G and G' with $G \cong_k G'$, $|\mathrm{Ans}((H,X),G)| = |\mathrm{Ans}((H,X),G')|$.

Consider G and G' with $G \cong_k G'$. This implies that for every graph H with treewidth at most k, $|\mathrm{Hom}(H,G)| = |\mathrm{Hom}(H,G')|$. From the definition of $\mathrm{ew}(H,X)$ and Corollary 18, for every positive integer ℓ, the treewidth of $F_\ell(H,X)$ is at most k. Therefore, $|\mathrm{Hom}(F_\ell(H,X),G)| = |\mathrm{Hom}(F_\ell(H,X),G')|$. The claim then follows directly by Lemma 22. □

4 LOWER BOUND ON THE WL-DIMENSION

The lower bound for the WL-dimension is significantly more challenging; it reads as follows.

Theorem 24. *Let (H, X) be a counting minimal conjunctive query such that H is connected, and $\emptyset \subsetneq X \subsetneq V(H)$. Then the WL-dimension of (H, X) is at least $\mathrm{ew}(H, X)$.*

In order to prove Theorem 24, we will find so-called *CFI graphs*[2] G and G' such that $G \cong_{k-1} G'$, where $k = \mathrm{ew}(H, X)$, and $|\mathrm{Ans}((H, X), G)|$ is different from $|\mathrm{Ans}((H, X), G')|$. Due to the space constraints and the extensive technical requirements necessary for our construction, we defer the proof of Theorem 24 to the full version [22].

5 MAIN RESULT AND CONSEQUENCES

With upper and lower bounds established, we are now able to proof Theorem 1, which we restate for convenience.

Theorem 1. *Let (H, X) be a connected conjunctive query with $X \neq \emptyset$. Then the WL-dimension of (H, X) is equal to its semantic extension width* $\mathrm{sew}(H, X)$.

Proof. We first consider the special case where (H, X) is a *full conjunctive query*, that is, no variable of (H, X) is existentially quantified so $X = V(H)$. In this case, (H, X) is counting minimal, since counting equivalence is the same as isomorphism in this case [16]. Moreover, $\Gamma(H, X) = H$. Thus $\mathrm{sew}(H, X) = \mathrm{tw}(H)$. Since $\mathrm{Ans}((H, X), G) = \mathrm{Hom}(H, G)$ for $X = V(H)$, counting answers to (H, X) is the same as counting homomorphisms from H, and the WL-dimension of counting homomorphisms is $\mathrm{tw}(H)$ as shown by Neuen [31].

Now consider the case where $X \neq V(H)$ and let (H', X') be a counting minimal conjunctive query with $(H', X') \sim (H, X)$. Then, $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H', X'), G)|$ for every graph G and thus (H, X) and (H', X') have the same WL-dimension. Furthermore, since (H, X) is connected, so is (H', X') — see [17, Section 6]. Theorems 21 and 24 now state that the WL-dimension of (H', X') is equal to $\mathrm{ew}(H', X')$. Finally, by definition of semantic extension width, we have $\mathrm{sew}(H, X) = \mathrm{ew}(H', X')$, concluding the proof. □

For the remainder of the paper, we will discuss the most important consequences of our main result; due to the space constraints, the proofs of this section, except for the proof of Corollary 6 are deferred to the full version [22].

5.1 Homomorphism Indistinguishability and Conjunctive Queries

Given a class of graphs $\mathcal{F}$, two graphs G and G' are called $\mathcal{F}$*-indistinguishable*, denoted by $G \cong_{\mathcal{F}} G'$ if for all $F \in \mathcal{F}$ we have $|\mathrm{Hom}(F, G)| = |\mathrm{Hom}(F, G')|$. If $\mathcal{F}$ is the class of all graphs, then a classical result of Lovász states that $\cong_{\mathcal{F}}$ coincides with isomorphism (see e.g. Theorem 5.29 in [26]). Recent years have seen numerous exciting results on the structure of $\mathcal{F}$-indistinguishability, depending on the class $\mathcal{F}$: For example, Dvorák [18], and Dell, Grohe and Rattan [15] have shown that $\cong_{\mathcal{F}}$ coincides with $\cong_k$, i.e., with k-WL-equivalence, if $\mathcal{F}$ is the class of all graphs of treewidth at most k, and Mancinska and Roberson have shown that $\cong_{\mathcal{F}}$ coincides with what is called quantum-isomorphism if $\mathcal{F}$ is the class of all planar graphs [27].

To state our first corollary, we extend the notion of homomorphism indistinguishability to conjunctive queries.

Definition 25. Let Ψ be a class of conjunctive queries. Two graphs G and G' are Ψ-*indistinguishable*, denoted by $G \cong_{\Psi} G'$, if for all queries $(H, X) \in \Psi$ we have $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$.

[2]Named after Cai, Fürer, and Immerman [9]

Then, using the notion of conjunctive query indistinguishability, we obtain a new characterisation of k-WL-equivalence.

Corollary 26 (Corollary 2, restated)**.** *Let k be a positive integer and let Ψ_k be the set of all connected conjunctive queries with at least one free variable and with semantic extension width at most k. Then for any pair of graphs G and G', $G \cong_k G'$ if and only if $G \cong_{\Psi_k} G'$.*

In the following corollary, we show that the treewidth of a conjunctive query alone is insufficient for describing the WL-dimension. This is even the case for treewidth 1, i.e., for acyclic queries.

Corollary 27. *The class of acyclic conjunctive queries has unbounded WL-dimension, that is, there is no k such that $G \cong_k G'$ if and only if $G \cong_{\mathcal{T}} G'$, where $\mathcal{T}$ is the class of all acyclic conjunctive queries.*

Corollary 27 is in stark contrast to the quantifier-free case. The WL-dimension of any acyclic conjunctive query $(H, V(H))$ is equal to 1 since this case is equivalent to counting homomorphisms from acyclic graphs [15, 18]. Given Corollary 27 one might ask how powerful indistinguishability by acyclic conjunctive queries is: Is there any $k > 1$ such that $\mathcal{T}$-indistinguishability is at least as powerful as k-WL-equivalence? We provide a negative answer.

Observation 28. *Let $2K_3$ denote the graph consisting of two disjoint triangles and let C_6 denote the 6-cycle. Let (H, X) be a connected and acyclic conjunctive query. Then $|\mathrm{Ans}((H, X), 2K_3)| = |\mathrm{Ans}((H, X), C_6|$.*

In other words, acyclic conjunctive queries cannot even distinguish $2K_3$ and C_6, which are the most common examples of graphs which are 1-WL-equivalent, but which are not 2-WL-equivalent.

5.2 WL-Dimension and the Complexity of Counting

In this section, we give a connection between WL-dimension and the parameterised complexity of counting answers to conjunctive queries. Recall that, given a class of conjunctive queries Ψ, the counting problem #CQ(Ψ) takes as input a pair consisting of a conjunctive query $(H, X) \in \Psi$ and a graph G and outputs $|\mathrm{Ans}((H, X), G)|$.

Recall that a class of conjunctive queries is said to have *bounded* WL-dimension if there is a constant B that upper bounds the WL-dimension of all queries in the class. We first restate the classification of #CQ(Ψ).

Corollary 4. *Let Ψ be a recursively enumerable class of counting minimal and connected conjunctive queries with at least one free variable. The problem #CQ(Ψ) is solvable in polynomial time if and only if the WL-dimension of Ψ is bounded; the "only if" is conditioned under the assumption* $\mathrm{FPT} \neq W[1]$.

5.3 Linear Combinations of Conjunctive Queries

The study of linear combinations of homomorphism counts dates back to the work of Lovász (see the textbook [26]). Recently, staring with the work of Chen and Mengel [11] and of Curticapean, Dell and Marx [13], the study of these linear combinations has re-arisen in the context of parameterised counting complexity theory. Moreover, the works of Seppelt [35], Neuen [31], and Lanzinger and Barceló [25] have shown that the WL-dimension of a function evaluating a finite linear combination of homomorphism counts is equal to the maximum WL-dimension of any term in the combination. Using Theorem 1, we establish a similar result (Corollary 5) for linear combinations of conjunctive queries. This gives a precise quantification of the WL-dimension of unions of conjunctive queries and of conjunctive queries with disequalities and negations over the free variables.

Following Lovász's notion of a "quantum graph"[26, Chapter 6.1], we formalise our linear combinations as follows.

Definition 29 (Quantum Query)**.** A *quantum query* Q is a formal finite linear combination of conjunctive queries $Q = \sum_{i=1}^{\ell} c_i \cdot (H_i, X_i)$ such that, for all $i \in [\ell]$, $c_i \in \mathbb{Q} \setminus \{0\}$ and (H_i, X_i) is a connected and counting minimal conjunctive query with $X_i \neq 0$. Moreover, the conjunctive queries (H_i, X_i) are pairwise non-isomorphic. We call the queries (H_i, X_i) the *constituents* of Q. The number of answers of Q in a graph G is defined as $|\mathrm{Ans}(Q, G)| := \sum_{i=1}^{\ell} c_i \cdot |\mathrm{Ans}((H_i, X_i), G)|$.

Chen and Mengel [11], and Dell, Roth, and Wellnitz [16] have shown that for every union φ of (connected) conjunctive queries with at least one free variable there is a quantum query $Q[\varphi]$ such that, for all graphs G, the number of answers of φ in G is equal to $|\mathrm{Ans}(Q[\varphi], G)|$. Moreover, $Q[\varphi]$ is unique up to reordering terms (and up to isomorphim of the constituents). They have also shown a similar result when φ is a conjunctive query with disequalities and negations.

Definition 30 (Hereditary Semantic Extension Width)**.** The *hereditary semantic extension width* of a quantum query $Q = \sum_{i=1}^{\ell} c_i \cdot (H_i, X_i)$ is $\mathsf{hsew}(Q) = \max\{\mathsf{sew}(H_i, X_i) \mid i \in [\ell]\}$.

We define the WL-dimension of a quantum query Q as the WL-dimension of the graph parameter $G \mapsto |\mathrm{Ans}(Q, G)|$. A classification of the WL-dimension of quantum queries was shown by Seppelt [35] in *the special case of homomorphisms*, i.e., the case in which each constituent (H_i, X_i) satisfies $X_i = V(H_i)$. We provide the full picture; our classification is restated below for convenience:

Corollary 5. *The WL-dimension of a quantum query Q is equal to* $\mathsf{hsew}(Q)$.

5.4 Star Queries and Dominating Sets

In this final section we use Theorem 1 to determine the WL-dimension of counting dominating sets (proving Corollary 6).

Definition 31 (Dominating Set)**.** Let G be a graph and let k be a positive integer. A *dominating set* of G is a subset $D \subseteq V(G)$ such that each vertex of G is either contained in D or adjacent to a vertex in D. The set $\Delta_k(G)$ contains all size-k dominating sets of G.

For analysing the WL-dimension of counting size-k dominating sets we will consider, as an intermediate step, the k-star-query.

Definition 32. Let k be a positive integer. The *k-star* is the conjunctive query (S_k, X_k) where $X_k = \{x_1, \ldots, x_k\}$, $V(S_k) = X \cup \{y\}$, and $E(S_k) = \{\{x_i, y\} \mid i \in [k]\}$.

The k-star is often written in the more prominent form $\varphi(x_1, \ldots, x_k) = \exists y : \bigwedge_{i=1}^{k} E(x_i, y)$. It is well known that (S_k, X_k) is counting minimal (see e.g. [16]). Moreover, $\Gamma(S_k, X_k)$ is the $(k+1)$-clique, which has treewidth k. Thus $\mathsf{sew}(S_k, X_k) = k$. Corollary 33 follows immediately from Theorem 1.

Corollary 33. *The WL-dimension of (S_k, X_k) is k.*

We can now provide the proof of Corollary 6, which we decided not to defer to the full version, since it illustrates the combined power of the machinery developed in this paper.

Corollary 34 (Corollary 6, restated)**.** *The WL-dimension of the function $G \mapsto |\Delta_k(G)|$ is k.*

Proof. We start with the lower bound. To this end, given a graph G, and a conjunctive query (H, X), we set

$$\mathsf{Inj}((H, X), G) = \{a \in \mathrm{Ans}((H, X), G) \mid a \text{ is injective}\}.$$

Let $I = \{(i, j) \in [k]^2 \mid i < j\}$ and consider a subset $J \subseteq I$. The query $(S_k, X_k)/J$ is obtained by identifying x_i and x_j if and only if $(i, j) \in J$. Observe that $(S_k, X_k)/J \cong (S_\ell, X_\ell)$ for some $\ell \leq k$. By the principle of inclusion and exclusion, for each graph G,

$$
\begin{aligned}
|\mathsf{Inj}((S_k, X_k), G)| &= \sum_{J \subseteq I} (-1)^{|J|} \cdot |\mathrm{Ans}((S_k, X_k)/J, G)| \\
&= \sum_{i=1}^{k} c_i \cdot |\mathrm{Ans}((S_i, X_i), G)|,
\end{aligned}
$$

where $c_i = \{J \subseteq I \mid (S_k, X_k)/J \cong (S_i, X_i)\}$. Thus, $|\mathsf{Inj}((S_k, X_k), G)|$ computes the number of answers to the quantum query with constituents (S_i, X_i) and coefficients c_i. Moreover, $c_k = 1$ since $(S_k, X_k)/J \cong (S_k, X_k)$ if and only if $J = \emptyset$.[3] By Corollary 5 and the fact that $\mathrm{sew}(S_\ell, X_\ell) = \ell$ for all $\ell \in [k]$, the WL-dimension of the function $G \mapsto |\mathsf{Inj}((S_k, X_k), G)|$ is equal to k.

Next observe that $|\mathsf{Inj}((S_k, X_k), G)|/k!$ is the number of k-vertex subsets A of G such that there is a vertex $y \in V(G)$ that is adjacent to all $a \in A$. Thus $\binom{|V(G)|}{k} - |\mathsf{Inj}((S_k, X_k), G)|/k!$ is equal to the size of the set

$$D_k(G) := \{A \subseteq V(G) \mid |A| = k \ \wedge \ \forall y \in V(G) : \exists a \in A : \{a, y\} \notin E(G)\}.$$

Let $\overline{G}$ be the self-loop-free complement of G, that is, two *distinct* vertices u and v in $V(\overline{G}) = V(G)$ are adjacent in $\overline{G}$ if and only if they are not adjacent in G. Observe that $\{a, y\} \notin E(G)$ if and only if $a = y$ or $\{a, y\} \in E(\overline{G})$. Therefore $|D_k(G)| = |\Delta_k(\overline{G})|$.

We are now ready to prove that the WL-dimension of the function $G \mapsto |\Delta_k(G)|$ is at least k. Suppose for contradiction that its WL-dimension is k' for some $1 \leq k' < k$. Then, for all G and G' with $G \cong_{k'} G'$, $|\Delta_k(G)| = |\Delta_k(G')|$. However, we know that the WL-dimension of $G \mapsto |\mathsf{Inj}((S_k, X_k), G)|$ is equal to k. Thus there are graphs F and F' with $F \cong_{k'} F'$ and $|\mathsf{Inj}((S_k, X_k), F)| \neq |\mathsf{Inj}((S_k, X_k), F')|$. It is well known (see e.g. Seppelt [35]) that $F \cong_{k'} F'$ implies $\overline{F} \cong_{k'} \overline{F'}$. Therefore $|\Delta_k(\overline{F})| = |\Delta_k(\overline{F'})|$. Let K_1 be the (treewidth 0) graph containing one isolated vertex. The number of homomomorphisms from K_1 to F determines the number of vertices of F so $F \cong_{k'} F'$ implies $|V(F)| = |V(F')|$. Let $n = |V(F)| = |V(F')|$. In summary,

$$
\begin{aligned}
|\mathsf{Inj}((S_k, X_k), F)| &= k! \left(\binom{n}{k} - |\Delta_k(\overline{F})| \right) \\
&= k! \left(\binom{n}{k} - |\Delta_k(\overline{F'})| \right) = |\mathsf{Inj}((S_k, X_k), F')|,
\end{aligned}
$$

which contradicts the choice of F and F' and concludes the proof of the lower bound.

For the upper bound, we have to show that $F \cong_k F'$ implies $\Delta_k(F) = \Delta_k(F')$, which is an immediate consequence of our previous analysis. Since the WL-dimension of $G \mapsto |\mathsf{Inj}((S_k, X_k), G)|$ is equal to k, we have

$$
\begin{aligned}
|\Delta_k(F)| &= \binom{n}{k} - |\mathsf{Inj}((S_k, X_k), \overline{F})|/k! \\
&= \binom{n}{k} - |\mathsf{Inj}((S_k, X_k), \overline{F'})|/k! = |\Delta_k(F')|.
\end{aligned}
$$

This concludes the proof. □

[3] To obtain a quantum query, we need to remove all terms (S_i, X_i) with $c_i = 0$. In fact, following the analysis in [13], it can be shown that none of the c_i is 0. However, since we only need $c_k \neq 0$, we omit going into further details.

ACKNOWLEDGMENTS

We are very grateful to Matthias Lanzinger for helpful advice during the early stages of this work.

Andreas Göbel was funded by the project PAGES (project No. 467516565) of the German Research Foundation (DFG).

REFERENCES

[1] Noga Alon, Phuong Dao, Iman Hajirasouliha, Fereydoun Hormozdiari, and S. Cenk Sahinalp. 2008. Biomolecular network motif counting and discovery by color coding. *Bioinformatics* 24, 13 (07 2008), i241–i249.

[2] Vikraman Arvind. 2016. The Weisfeiler-Lehman Procedure. *Bull. EATCS* 120 (2016). http://eatcs.org/beatcs/index.php/beatcs/article/view/442

[3] Vikraman Arvind, Frank Fuhlbrück, Johannes Köbler, and Oleg Verbitsky. 2022. On the Weisfeiler-Leman dimension of fractional packing. *Inf. Comput.* 288 (2022), 104803. https://doi.org/10.1016/j.ic.2021.104803

[4] Pablo Barceló, Floris Geerts, Juan L. Reutter, and Maksimilian Ryschkov. 2021. Graph Neural Networks with Local Graph Parameters. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, Marc'Aurelio Ranzato, Alina Beygelzimer, Yann N. Dauphin, Percy Liang, and Jennifer Wortman Vaughan (Eds.). 25280–25293.

[5] Hans L. Bodlaender. 2003. Necessary Edges in k-Chordalisations of Graphs. *J. Comb. Optim.* 7, 3 (2003), 283–290. https://doi.org/10.1023/A:1027320705349

[6] Jan Böker. 2019. Color Refinement, Homomorphisms, and Hypergraphs. In *Graph-Theoretic Concepts in Computer Science - 45th International Workshop, WG 2019, Vall de Núria, Spain, June 19-21, 2019, Revised Papers (Lecture Notes in Computer Science, Vol. 11789)*, Ignasi Sau and Dimitrios M. Thilikos (Eds.). Springer, 338–350. https://doi.org/10.1007/978-3-030-30786-8_26

[7] Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. 2023. Improving Graph Neural Network Expressivity via Subgraph Isomorphism Counting. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 1 (2023), 657–668. https://doi.org/10.1109/TPAMI.2022.3154319

[8] Silvia Butti and Víctor Dalmau. 2021. Fractional Homomorphism, Weisfeiler-Leman Invariance, and the Sherali-Adams Hierarchy for the Constraint Satisfaction Problem. In *46th International Symposium on Mathematical Foundations of Computer Science, MFCS 2021, August 23-27, 2021, Tallinn, Estonia (LIPIcs, Vol. 202)*, Filippo Bonchi and Simon J. Puglisi (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 27:1–27:19. https://doi.org/10.4230/LIPIcs.MFCS.2021.27

[9] Jin-yi Cai, Martin Fürer, and Neil Immerman. 1992. An optimal lower bound on the number of variables for graph identification. *Comb.* 12, 4 (1992), 389–410. https://doi.org/10.1007/BF01305232

[10] Hubie Chen and Stefan Mengel. 2015. A Trichotomy in the Complexity of Counting Answers to Conjunctive Queries. In *18th International Conference on Database Theory, ICDT 2015, March 23-27, 2015, Brussels, Belgium (LIPIcs, Vol. 31)*, Marcelo Arenas and Martín Ugarte (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 110–126. https://doi.org/10.4230/LIPIcs.ICDT.2015.110

[11] Hubie Chen and Stefan Mengel. 2016. Counting Answers to Existential Positive Queries: A Complexity Classification. In *Proceedings of the 35th ACM SIGMOD-SIGACT-SIGAI Symposium on Principles of Database Systems, PODS 2016, San Francisco, CA, USA, June 26 - July 01, 2016*, Tova Milo and Wang-Chiew Tan (Eds.). ACM, 315–326. https://doi.org/10.1145/2902251.2902279

[12] Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. 2020. Can Graph Neural Networks Count Substructures?. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.). https://proceedings.neurips.cc/paper/2020/hash/75877cb75154206c4e65e76b88a12712-Abstract.html

[13] Radu Curticapean, Holger Dell, and Dániel Marx. 2017. Homomorphisms are a good basis for counting small subgraphs. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing, STOC 2017, Montreal, QC, Canada, June 19-23, 2017*, Hamed Hatami, Pierre McKenzie, and Valerie King (Eds.). ACM, 210–223. https://doi.org/10.1145/3055399.3055502

[14] Anuj Dawar, Tomás Jakl, and Luca Reggio. 2021. Lovász-Type Theorems and Game Comonads. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 1–13. https://doi.org/10.1109/LICS52264.2021.9470609

[15] Holger Dell, Martin Grohe, and Gaurav Rattan. 2018. Lovász Meets Weisfeiler and Leman. In *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic (LIPIcs, Vol. 107)*, Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 40:1–40:14. https://doi.org/10.4230/LIPIcs.ICALP.2018.40

[16] Holger Dell, Marc Roth, and Philip Wellnitz. 2019. Counting Answers to Existential Questions. In *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece (LIPIcs, Vol. 132)*, Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 113:1–113:15. https://doi.org/10.4230/LIPIcs.ICALP.2019.113

[17] Holger Dell, Marc Roth, and Philip Wellnitz. 2019. Counting Answers to Existential Questions. *CoRR* abs/1902.04960 (2019). arXiv:1902.04960 http://arxiv.org/abs/1902.04960

[18] Zdenek Dvorák. 2010. On recognizing graphs by numbers of homomorphisms. *J. Graph Theory* 64, 4 (2010), 330–342. https://doi.org/10.1002/jgt.20461

[19] Jörg Flum and Martin Grohe. 2006. *Parameterized Complexity Theory*. Springer. https://doi.org/10.1007/3-540-29953-X

[20] Martin Grohe. 2021. The Logic of Graph Neural Networks. In *36th Annual ACM/IEEE Symposium on Logic in Computer Science, LICS 2021, Rome, Italy, June 29 - July 2, 2021*. IEEE, 1–17. https://doi.org/10.1109/LICS52264.2021.9470677

[21] Martin Grohe, Kristian Kersting, Martin Mladenov, and Pascal Schweitzer. 2021. Color Refinement and Its Applications. In *An Introduction to Lifted Probabilistic Inference*. The MIT Press. https://doi.org/10.7551/mitpress/10548.003.0023 arXiv:https://direct.mit.edu/book/chapter-pdf/2101088/c025000_9780262365598.pdf

[22] Andreas Göbel, Leslie Ann Goldberg, and Marc Roth. 2023. The Weisfeiler-Leman Dimension of Conjunctive Queries. arXiv:2310.19006 [cs.DM]

[23] Neil Immerman and Eric Lander. 1990. *Describing Graphs: A First-Order Approach to Graph Canonization*. Springer New York, New York, NY, 59–81. https://doi.org/10.1007/978-1-4612-4478-3_5

[24] Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. 2018. Network Global Testing by Counting Graphlets. In *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018 (Proceedings of Machine Learning Research, Vol. 80)*, Jennifer G. Dy and Andreas Krause (Eds.). PMLR, 2338–2346. http://proceedings.mlr.press/v80/jin18b.html

[25] Matthias Lanzinger and Pablo Barceló. 2023. On the Power of the Weisfeiler-Leman Test for Graph Motif Parameters. *CoRR* abs/2309.17053 (2023). https://doi.org/10.48550/arXiv.2309.17053 arXiv:2309.17053

[26] László Lovász. 2012. *Large Networks and Graph Limits*. Colloquium Publications, Vol. 60. American Mathematical Society.

[27] Laura Mancinska and David E. Roberson. 2020. Quantum isomorphism is equivalent to equality of homomorphism counts from planar graphs. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS 2020, Durham, NC, USA, November 16-19, 2020*, Sandy Irani (Ed.). IEEE, 661–672. https://doi.org/10.1109/FOCS46700.2020.00067

[28] R. Milo, S. Shen-Orr, S. Itzkovitz, N. Kashtan, D. Chklovskii, and U. Alon. 2002. Network Motifs: Simple Building Blocks of Complex Networks. *Science* 298, 5594 (2002), 824–827.

[29] Christopher Morris, Gaurav Rattan, and Petra Mutzel. 2020. Weisfeiler and Leman go sparse: Towards scalable higher-order graph embeddings. In *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin (Eds.).

[30] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 2019. Weisfeiler and Leman Go Neural: Higher-Order Graph Neural Networks. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 4602–4609. https://doi.org/10.1609/aaai.v33i01.33014602

[31] Daniel Neuen. 2023. Homomorphism-Distinguishing Closedness for Graphs of Bounded Tree-Width. *arXiv preprint arXiv:2304.07011* (2023).

[32] Reinhard Pichler and Sebastian Skritek. 2013. Tractable counting of the answers to conjunctive queries. *J. Comput. Syst. Sci.* 79, 6 (2013), 984–1001. https://doi.org/10.1016/j.jcss.2013.01.012

[33] David E Roberson. 2022. Oddomorphisms and homomorphism indistinguishability over graphs of bounded degree. *arXiv preprint arXiv:2206.10321* (2022).

[34] Benjamin Scheidt and Nicole Schweikardt. 2023. Counting Homomorphisms from Hypergraphs of Bounded Generalised Hypertree Width: A Logical Characterisation. In *48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France (LIPIcs, Vol. 272)*, Jérôme Leroux, Sylvain Lombardy, and David Peleg (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 79:1–79:15. https://doi.org/10.4230/LIPIcs.MFCS.2023.79

[35] Tim Seppelt. 2023. Logical Equivalences, Homomorphism Indistinguishability, and Forbidden Minors. In *48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France (LIPIcs, Vol. 272)*, Jérôme Leroux, Sylvain Lombardy, and David Peleg (Eds.). Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 82:1–82:15. https://doi.org/10.4230/LIPIcs.MFCS.2023.82

[36] Gottfried Tinhofer. 1986. Graph isomorphism and theorems of Birkhoff type. *Computing* 36, 4 (1986), 285–300. https://doi.org/10.1007/BF02240204

[37] Gottfried Tinhofer. 1991. A note on compact graphs. *Discret. Appl. Math.* 30, 2-3 (1991), 253–264. https://doi.org/10.1016/0166-218X(91)90049-3

[38] Boris Weisfeiler and Andrei Leman. 1968. The reduction of a graph to canonical form and the algebra which appears therein. *NTI, Series* 2, 9 (1968), 12–16. English translation by G. Ryabov available at https://www.iti.zcu.cz/wl2018/pdf/wl_paper_translation.pdf.

[39] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 2019. How Powerful are Graph Neural Networks?. In *7th International Conference on Learning Representations, ICLR 2019, New Orleans, LA, USA, May 6-9, 2019.* OpenReview.net. https://openreview.net/forum?id=ryGs6iA5Km

Received December 2023; revised February 2024; accepted March 2024