The Weisfeiler-Leman Dimension of Conjunctive Queries*

Andreas Göbel
Hasso Plattner Institute
University of Potsdam
Germany

Leslie Ann 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

29 October 2023

Abstract

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 qureies. Given any conjunctive query φ, we prove that its WL-dimension is equal to the *semantic extension width* $\mathsf{sew}(\varphi)$, 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 $\mathsf{sew}(\varphi)$.

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 recent modification due to Roberson (ArXiv 2022) 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.

*Andreas Göbel was funded by the project PAGES (project No. 467516565) of the German Research Foundation (DFG). All data is provided in full in the results section of this paper.

1 Introduction

The Weisfeiler–Leman (WL) algorithm [41] and its higher dimensional generalisations [10] 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, 25]).

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 [33, 42]. 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 [32, 13, 4, 3, 34, 28, 8]. We refer the reader to the survey by Grohe [24] 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 [39, 40].

(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' [26, 10].

(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' [21, 16]. 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 [13, 3, 34, 28, 8].

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 [21], Dell, Grohe and Rattan [16], Roberson [36], and Seppelt [38], it has very recently been shown by Neuen [34] 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. [35]); 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

$$\varphi(x_1, \dots, x_k) = \exists y_1, \dots, y_\ell \colon A_1 \wedge \dots \wedge A_m,$$

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. [35, 11, 12, 17]), 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 [17], 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 $\mathsf{sew}(H, X)$ 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 [11, 17]. 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 $\mathsf{sew}(S_k, X_k) = k$.

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* $\mathsf{sew}(H, X)$*.*

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 $\mathsf{sew}(H, X) \leq k$, $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$.*

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 [28]). 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 [31, 1, 27], researchers have studied the extent to which GNNs (and their higher order generalisations [33]) are able to count selected small patterns such as homomorphisms [28], subgraphs [8], and induced subgraphs [13].

Following [33] but simplifying the notation for our needs, we represent a t-layer order-k GNN N as a tuple $N = (G, W_1, \ldots, W_t, f_0, \ldots, f_t)$ 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 $G, W_1, \ldots, W_{i-1}, f_0, \ldots, f_{i-1}$. 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 it means for a GNN to "compute" a function on graphs. So far, this has been studied in a somewhat limited context. Namely, a GNN is said to "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 say that a GNN $N = (G, W_1, \ldots, W_t, f_0, \ldots, f_t)$ is "fully refined" if there is no GNN $N' = (G, W_1', \ldots, W_t', f_0, f_1', \ldots, f_{t'}')$ such that $P_{N'}(G)$ strictly refines $P_N(G)$.

In this setting, Morris et al. [33] 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 [33] 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 [21], Dell, Grohe and Rattan [16], and Lanzinger and Barcelo [28] 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 [21, 16]. It follows explicitly from [28, 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 [21, 16, 36, 28]. Specifically, Lanzinger and Barcelo [28] 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 $\mathsf{sew}(H, X) = k$ 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 [28, Lemma 7] and Proposition 3, for all k, all graphs H with treewidth k, and all graphs G there is a fully refined GNN $N \in \mathcal{N}_k$ with underlying graph G such that $|\mathrm{Hom}(H, G)|$ 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 $F_1, \ldots, F_n$ of treewidth at most k, such that, $|\mathrm{Ans}((H, X), G)|$ can be efficiently computed from the counts $|\mathrm{Hom}(F_i, G)|$.

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 $\mathsf{sew}(H, X)$. 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 $\mathsf{sew}(H, X) = k$. 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 $\mathsf{sew}(H, X)$. In particular, consider (H, X) with $\mathsf{sew}(H, X) = k$. 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. [22]). 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 $\#\mathrm{CQ}(\Psi)$ 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 63). 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 [12, 17] 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 [36].

(C) Barceló et al. [4], and Lanzinger and Barceló [28] 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 [7] and by Scheidt and Schweikardt [37] 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, and we prove the lower bound in Section 4. Those two sections can be read independently from each other and the majority of the conceptual and technical work is done in Section 4. 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 projection of a pair to its first component, that is, $\pi_1(a, b) = a$. Given a positive integer ℓ we set $[\ell] = \{1, \dots, \ell\}$.

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 [17].

[1] A characterisation for the special case of *constant* arity $r \geq 2$ was recently established independently by Butti and Dalmau [9], and by Dawar, Jakl, and Reggio [15].

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. [11, 12, 17]) 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 [12, 17]. 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 $\mathsf{tw}(H)$, is the width of an optimal tree decomposition of H. The treewidth of a conjunctive query (H, X), denoted by $\mathsf{tw}(H, X)$, 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 $\mathsf{ew}(H, X) := \mathsf{tw}(\Gamma(X, H))$.

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 $\mathsf{sew}(H, X)$, 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

$$\begin{aligned} E_X &= \{\{u, v\} \in E(H) \cap X^2\}, \\ E_{X,Y} &= \{\{u, (v, i)\} \mid u \in X, v \in Y, i \in [\ell], \{u, v\} \in E(H)\}, \text{ and} \\ E_Y &= \{\{(u, i), (v, i)\} \mid \{u, v\} \in E(H) \cap Y^2, i \in [\ell]\}. \end{aligned}$$

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 $\gamma[H, X, \ell] \colon V(F_\ell(H, X)) \to V(H)$ as follows:

$$\gamma[H, X, \ell](u) = \begin{cases} u & u \in X \\ \pi_1(u) & u \in Y \times [\ell] \end{cases}$$

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 $\mathsf{ew}(H, X)$.*

Proof. Let $\Gamma = \Gamma(H, X)$ and let $C_1, \ldots, C_m$ 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 $\mathsf{tw}(F_\ell(H, X)) \leq \mathsf{tw}(\Gamma)$. □

The following lemma follows implicitly from [5]. We include a proof for completeness.

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

Proof. Choose any $\ell > |V(H)| + 1$, and let $\gamma = \gamma[H, X, \ell]$. Let $(T, \mathcal{B})$, with $\mathcal{B} = \{B_t\}_{t \in V(T)}$ be an optimal tree decomposition of $F_\ell(H, X)$. We prove the lemma by constructing a tree decomposition $(T', \mathcal{B}')$ of $\Gamma(X, H)$ with with width at most the width of $(T, \mathcal{B})$. Let $T' = T$. For each $t \in V(T)$, define $B'_t = \{\gamma(v) \mid v \in B_t \text{ and } v \text{ is not of the form } (v, i) \text{ for } i > 1\}$.

We claim that $(T', \mathcal{B}')$ is a tree-decomposition of $\Gamma(H, X)$. This claim proves the lemma, since the width of $(T', \mathcal{B}')$ is clearly at most the width of $(T, \mathcal{B})$, and since the extension width of (H, X) is, by definition, the treewidth of $\Gamma(H, X)$. Hence it remains to prove our claim by establishing (T1), (T2), and (T3) from Definition 10. In each case, for each $v \in V(\Gamma(H, X))$, let $v' = v$ if $v \in X$ and let $v' = (v, 1)$ if $v \in Y$.

(T1) Consider $v \in V(\Gamma(H, X)) = V(H)$. Then $v' \in V(F_\ell(H, X))$ and thus there is a bag B_t with $v' \in B_t$. Since $v' \neq (v, i)$ for $i > 1$ and since $\gamma(v') = v$, $v \in B'_t$.

(T2) Consider $v \in V(\Gamma(H, X))$ and let s and t be any pair of vertices of T such that $v \in B'_s$ and $v \in B'_t$. We show that there is an s-t-path P in T such that $v \in B'_u$ for each $u \in P$. Since $v \in B'_s$ and $v \in B'_t$, we have $v' \in B_s$ and $v' \in B_t$. Using that $(T, \mathcal{B})$ is a tree-decomposition, there is a path P in T such that $v' \in B_u$, and thus $v \in B'_u$, for each $u \in P$.

(T3) Consider $e = \{v_1, v_2\} \in E(\Gamma(H, X))$. Recall from Definition 11 that $E(\Gamma(H, X)) = E(H) \cup E'$, where E' contains all $\{u, v\}$ such that there is a connected component C of $H[Y]$ that is adjacent to both u and v in H.

We distinguish between two cases. For the easy case, suppose that $e \in E(H)$. Then $e' = \{v'_1, v'_2\} \in E(F_\ell(H, X))$. Thus there is a bag $B_t \in \mathcal{B}$ with $e' \subseteq B_t$. Consequently, $e \in B'_t$.

For the more difficult case, suppose $e \in E'$. Then $v_1, v_2 \in X$ and thus $v_1 = v'_1$ and $v_2 = v'_2$. Moreover, there is a connected component C of $H[Y]$ that is adjacent to both v_1 and v_2. Since there are ℓ copies of C in $F_\ell(H, X)$, there are at least ℓ vertex disjoint paths from v_1 to v_2 in $F_\ell(H, X)$. Now, using known separation properties of tree decompositions (see for instance [19, Lemma 5]), there is either a bag B_t of $(T, \mathcal{B})$ that contains v_1 and v_2 — in this case, we are done — or there is an edge $e = \{s, t\}$ of T such that $S := B_s \cap B_t$ separates v_1 and v_2, that is, S does not contain v_1 and v_2, and every v_1-v_2-path of $F_\ell(H, X)$ intersects S. This requires $|S| \geq \ell$ and thus $|B_s| \geq \ell$ (and $|B_t| \geq \ell$). Consequently, using the fact that $(T, \mathcal{B})$ is optimal, the treewidth of $F_\ell(H, X)$ is at least $\ell > |V(H)| + 1$, which yields a contradiction, since by Lemma 16 we have

$$\mathsf{tw}(F_\ell(H, X)) \leq \mathsf{ew}(H, X) = \mathsf{tw}(\Gamma(H, X)) \leq |V(\Gamma(H, X))| = |V(H)| .$$

Thus v_1 and v_2 must both be contained in some bag B_t, and thus also in the bag B'_t.

With (T1)-(T3) established, the proof is concluded. □

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* $\mathsf{ew}(H,X) = \max\{\mathsf{tw}(F_\ell(H,X)) \mid \ell \in \mathbb{Z}_{>0}\}$.

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 [21] and Dell, Grohe and Rattan [16]. 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* $\mathsf{ew}(H,X)$.

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

$$H^G(\Upsilon) = \{h \in \mathrm{Ans}((H,X),G) \mid \mathrm{Ext}(h) = \Upsilon\}$$
$$\hat{H}^G_\ell(\Upsilon) = \{h \in \mathrm{Hom}(F_\ell(H,X),G) \mid \mathrm{Ext}(h|_X) = \Upsilon\}$$

First observe that for any $\Upsilon \subseteq \Omega$, $|\hat{H}_\ell^G(\Upsilon)| = |H^G(\Upsilon)| \cdot |\Upsilon|^\ell$. Moreover,

$$|\mathrm{Ans}((H,X),G)| = \sum_{\emptyset \neq \Upsilon \subseteq \Omega} |H^G(\Upsilon)|, \text{and}$$
$$|\mathrm{Hom}(F_\ell(H,X),G)| = \sum_{\emptyset \neq \Upsilon \subseteq \Omega} |\hat{H}_\ell^G(\Upsilon)|.$$

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 $i \in \{1, \ldots, n\}$. 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 $\mathsf{ew}(H,X)$ 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 $F_1, \ldots, F_n$ 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 $\ell \in \{1, \ldots, n\}$.*

Proof of Theorem 21. Let (H, X) be a conjunctive query. Let $k = \mathsf{ew}(H, X)$. 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 $\mathsf{ew}(H, X)$ and Corollary 18, for every positive integer ℓ, the treewidth of $F_\ell(H, X)$ is at most k. Thus, $|\mathrm{Hom}(F_\ell(H, X), G)| = |\mathrm{Hom}(F_\ell(H, X), G')|$. The claim then follows directly by Lemma 22. □

4 Lower Bound on the WL-Dimension

The goal of this section, which is the technical heart of the paper, is the proof of the following lower bound.

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 $\mathsf{ew}(H, X)$.*

In order to prove Theorem 24, we will find graphs G and G' such that $G \cong_{k-1} G'$, where $k = \mathsf{ew}(H, X)$, and $|\mathrm{Ans}((H, X), G)| \neq |\mathrm{Ans}((H, X), G')|$. As explained in the introduction, we will rely on a recently developed version of the CFI graphs of Cai, Fürer and Immerman [10]. The following subsection will provide a concise and self-contained explanation of the construction and properties of CFI graphs.

4.1 CFI Graphs

We start with a formal definition of a well-known version of CFI graphs [23] (see also [36]).

Definition 25 (CFI graphs, $\chi(G, W)$)**.** Let G be a graph and let W be a subset of $V(G)$. For every vertex w of G, let $\delta_{w,W} = |\{w\} \cap W|$. The graph $\chi(G, W)$ is defined as follows. The vertex set is $V(\chi(G, W)) := \{(w, S) \mid w \in V(G), S \subseteq N_G(w), \delta_{w,W} \equiv |S| \pmod 2\}$. The edge set is

$$E(\chi(G, W)) := \{\{(w, S), (w', S')\} \mid \{w, w'\} \in E(G) \text{ and } w' \in S \iff w \in S'\}.$$

For any fixed G, the isomorphism class of $\chi(G, W)$ depends only on the parity of $|W|$:

Lemma 26 (Lemma 3.2 in [36])**.** *Let G be a connected graph and let $W, W' \subseteq V(G)$. Then $\chi(G, W) \cong \chi(G, W')$ if and only if $|W| \equiv |W'| \pmod 2$.*

Neuen [34] established the following WL-equivalence result for $\chi(G, \emptyset)$ and $\chi(G, \{w\})$.

Lemma 27 (Theorem 4.2, Lemma 4.4 and Theorem 5.1 in [34])**.** *Let G be a graph of treewidth t and let w be a vertex of G. Then for all $k < t$, $\chi(G, \emptyset) \cong_k \chi(G, \{w\})$.*

4.2 Cloning Vertices in CFI Graphs

We will introduce some notions and properties of coloured graphs and of CFI graphs due to Roberson [36]. First of all, since we will work with vertex colourings induced by homomorphisms throughout this section, we adopt the well-established notion of *H-colourings* of graphs.

$$\begin{tikzcd} H \arrow[r, "h"] \arrow[dr, "\tau"'] & G \arrow[d, "c"] \\ & F \end{tikzcd}$$

Figure 1: Each homomorphism h from H to G induces a homomorphism τ from H to F by composing h with the F-colouring c of G. By partitioning $\mathrm{Hom}(H, G)$ along the induced homomorphisms to F, we obtain Observation 31.

Definition 28. We refer to a homomorphism from a graph G to a graph H as an H-*colouring* of G.

Recall that π_1 is the projection that maps a pair (a, b) to the first component a.

Observation 29 ([36])**.** *Let F be a graph and let W be a subset of $V(F)$. The function π_1 is a homomorphism from $\chi(F, W)$ to F.*

Technically, to get the F-colouring of $\chi(F, W)$ in Observation 29, one should restrict π_1 to the domain $V(\chi(F, W))$, but it will not be important to capture this in our notation. The following is an extension of a notion introduced in [36, Section 3.1] from CFI graphs to coloured graphs.

Definition 30. Let H, G, and F be graphs, let c be a homomorphism from G to F, and let τ be a homomorphism from H to F. We define $\mathrm{Hom}_\tau(H, G, F, c) = \{h \in \mathrm{Hom}(H, G) \mid c(h(\cdot)) = \tau\}$.

Observation 31. *Let H, G, and F be graphs and let c be a homomorphism from G to F. Then*

$$|\mathrm{Hom}(H, G)| = \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, G, F, c)|.$$

Theorem 32 (Theorem 3.6 in [36])**.** *Let H be a graph, let F be a connected graph, let $W \subseteq V(F)$, and let $\tau \in \mathrm{Hom}(H, F)$. Then $|\mathrm{Hom}_\tau(H, \chi(F, W), F, \pi_1)| \leq |\mathrm{Hom}_\tau(H, \chi(F, \emptyset), F, \pi_1)|$.*

Definition 33 (Cloning Colour-Blocks)**.** Let G be a graph, let F be a connected graph, and let c be a homomorphism from G to F. Let k be a positive integer, let $\vec{v} = (v_1, \ldots, v_k)$ be a k-tuple of pairwise distinct vertices of F, and let $\vec{z} = (z_1, \ldots, z_k)$ be k-tuple of positive integers. The graph $\mathcal{G}(G, F, c, \vec{v}, \vec{z})$ is obtained from G by cloning, for each $i \in [k]$, the colour class of v_i under c precisely $z_i - 1$ times. Formally, for each $v \in V(F)$, let $B_v = c^{-1}(v)$. The vertices of $\mathcal{G}(G, F, c, \vec{v}, \vec{z})$ are

$$\bigcup_{u \in V(F) \setminus \vec{v}} B_u \cup \bigcup_{i \in [k]} (B_{v_i} \times \{1, \ldots, z_i\}).$$

The vertices contained in $\bigcup_{u \in V(F) \setminus \vec{v}} B_u$ are called *primal vertices* and the vertices contained in $\bigcup_{i \in [k]} (B_{v_i} \times \{1, \ldots, z_i\})$ are called *cloned vertices*. Two vertices x and y of $\mathcal{G}(G, F, c, \vec{v}, \vec{z})$ are adjacent if and only if

- x and y are primal vertices, and $\{x, y\} \in E(G)$, or
- x is primal, y is a clone, and $\{x, \pi_1(y)\} \in E(G)$, or
- x is a clone, y is primal, and $\{\pi_1(x), y\} \in E(G)$, or
- x and y are clones, and $\{\pi_1(x), \pi_1(y)\} \in E(G)$.

We define a function $\mathcal{C}(G, F, c, \vec{v}, \vec{z})\colon V(\mathcal{G}(G, F, c, \vec{v}, \vec{z}) \to V(F)$ by mapping primal vertices u to $c(u)$, and cloned vertices (u, i) to $c(u)$. It is easy to see that $\mathcal{C}(G, F, c, \vec{v}, \vec{z})$ is a homomorphism from $\mathcal{G}(G, F, c, \vec{v}, \vec{z})$ to F.

Lemma 34. *Let H and G be graphs and let F be a connected graph. Let c be a homomorphism from G to F and let τ be a homomorphism from H to F. Let $\vec{v} = (v_1, \dots, v_k)$ be a k-tuple of distinct vertices of F and let $\vec{z} = (z_1, \dots, z_k)$ be a k-tuple of positive integers. For all $i \in [k]$, let d_i be the number of vertices of H that are mapped by τ to v_i, i.e., $d_i = |\{u \in V(H) \mid \tau(u) = v_i\}|$. Let $G' = \mathcal{G}(G, F, c, \vec{v}, \vec{z})$ and let $c' = \mathcal{C}(G, F, c, \vec{v}, \vec{z})$. Then*

$$|\mathrm{Hom}_\tau(H, G', F, c')| = |\mathrm{Hom}_\tau(H, G, F, c)| \cdot \prod_{i=1}^{k} z_i^{d_i}.$$

Proof. Let $\rho\colon V(G') \to V(G)$ be the function that maps cloned vertices to their original counterparts, that is

$$\rho(x) = \begin{cases} x & \text{if } x \text{ is a primal vertex} \\ \pi_1(x) & \text{otherwise} \end{cases}$$

Observe that ρ is a homomorphism from G' to G. We define an equivalence relation on the set $\mathrm{Hom}_\tau(H, G', F, c')$ by setting $h \sim h'$ if and only if $\rho(h(\cdot)) = \rho(h'(\cdot))$. For every $h \in \mathrm{Hom}_\tau(H, G', F, c')$, $\rho(h(\cdot))$ is a homomorphism from H to G since it is the composition of homomorphisms from H to G' and from G' to G. Moreover, since $h \in \mathrm{Hom}_\tau(H, G', F, c')$, $c'(h(\cdot)) = \tau$. From the definitions of ρ and c' it is immediate that $c(\rho(\cdot)) = c'$. Thus $c(\rho(h(\cdot)) = c'(h(\cdot)) = \tau$, proving that $\rho(h(\cdot)) \in \mathrm{Hom}_\tau(H, G, F, c)$; consider Figure 2 for an illustration. Consequently, we can represent each equivalence class of $\sim$ by a homomorphism $\hat{h} \in \mathrm{Hom}_\tau(H, G, F, C)$.

Finally, each equivalence class has size $\prod_{i=1}^{k} z_i^{d_i}$, since for each $i \in [k]$, there are z_i possibilities for each of the d_i vertices $u \in V(H)$ with $\tau(u) = v_i$. □

Next we show that cloning colour-blocks in CFI graphs preserves WL-equivalence:

Lemma 35. *Let F be a connected graph of treewidth $t+1$, let $W \subseteq V(F)$, let $\vec{v} = (v_1, \dots, v_k)$ be a k-tuple of distinct vertices of F, and let $\vec{z} = (z_1, \dots, z_k)$ be a k-tuple of positive integers. Then $\mathcal{G}(\chi(F, \emptyset), F, \pi_1, \vec{v}, \vec{z}) \cong_t \mathcal{G}(\chi(F, W), F, \pi_1, \vec{v}, \vec{z})$.*

Proof. Let $G_\emptyset = \mathcal{G}(\chi(F, \emptyset), F, \pi_1, \vec{v}, \vec{z})$ and $G_W = \mathcal{G}(\chi(F, W), F, \pi_1, \vec{v}, \vec{z})$. The lemma will follow immediately by Definition of WL-equivalence (Definition 19) once we show that, for each graph H of treewidth at most t, $|\mathrm{Hom}(H, G_\emptyset)| = |\mathrm{Hom}(H, G_W)|$. To this end, fix any graph H with treewidth at most t. By Lemma 27, $\chi(F, \emptyset) \cong_t \chi(F, W)$ since F has treewidth $t + 1$. Again, by Definition 19, we have $|\mathrm{Hom}(H, \chi(F, \emptyset))| = |\mathrm{Hom}(H, \chi(F, W))|$.

$$
\begin{tikzcd}
H \arrow[r, "h"] \arrow[dr, "\tau"'] & G' \arrow[r, "\rho"] \arrow[d, "c'"] & G \arrow[dl, "c"] \\
& F &
\end{tikzcd}
$$

Figure 2: Illustration for the proof of Lemma 34: $G' = \mathcal{G}(G, F, c, \vec{v}\,\vec{z})$ is the graph obtained from G by cloning vertices (Definition 33), and ρ is the homomorphism from G' to G that maps each cloned vertex in G' to its primal vertex in G. Moreover, c is the F-colouring of G and $c' = \mathcal{C}(G, F, c, \vec{v}, vz)$ is, by Definition 33, the composition of c and ρ, i.e., each cloned vertex is mapped by c' to the colour of its primal vertex.

By Observation 29, π_1 is a homomorphism from $\chi(F, \emptyset)$ to F and from $\chi(F, W)$ to F. By Observation 31,

$$\sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, \chi(F, \emptyset), F, \pi_1)| = \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, \chi(F, W), F, \pi_1).$$

Combining this with Theorem 32, we find that for all $\tau \in \mathrm{Hom}(H, F)$,

$$|\mathrm{Hom}_\tau(H, \chi(F, \emptyset), F, \pi_1))| = |\mathrm{Hom}_\tau(H, \chi(F, W), F, \pi_1)|. \tag{1}$$

Therefore

$$\begin{aligned}
|\mathrm{Hom}(H, G_\emptyset)| &= \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, G_\emptyset, F, \pi_1)| && \text{(Observation 31)} \\
&= \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, \chi(F, \emptyset), F, \pi_1)| \cdot \prod_{i=1}^{k} z_i^{d_i} && \text{(Lemma 34)} \\
&= \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, \chi(F, W), F, \pi_1)| \cdot \prod_{i=1}^{k} z_i^{d_i} && \text{(Equation (1))} \\
&= \sum_{\tau \in \mathrm{Hom}(H,F)} |\mathrm{Hom}_\tau(H, G_W, F, \pi_1)| && \text{(Lemma 34)} \\
&= |\mathrm{Hom}(H, G_W)|, && \text{(Observation 31)}
\end{aligned}$$

where the quantities $d_1, \dots, d_k$ depend on τ as in Lemma 34. This concludes the proof. $\square$

4.3 Reduction to the Colourful Case

Throughout this section, we fix a conjunctive query (H, X) such that H is connected, and with $X = \{x_1, \dots, x_k\}$. Fix also a positive integer ℓ. Let $Y = V(H) \setminus X$ and $F = F_\ell(H, X)$.

Recall that the vertices of F are $X \cup (Y \times [\ell])$. We start by extending the partitioning result from Section 4.2 from graphs to conjunctive queries. Moreover, recall the definition of the mapping $\gamma\colon V(F) \to V(H)$ (Definition 14):

$$\gamma(u) = \begin{cases} u & u \in X \\ \pi_1(u) & u \in Y \times [\ell] \end{cases}$$

$$
\begin{tikzcd}
V(F) \arrow[r, "\gamma"] & V(H) & X \arrow[l, "\tau"'] \arrow[dl, "h"] \\
& V(G) \arrow[ul, "\hat{c}"] \arrow[u, "c"] &
\end{tikzcd}
$$

Figure 3: Illustrations of the mappings used in Definition 36.

Recall also that by Observation 15, the function γ is a homomorphism from F to H.

Definition 36. Let G be a graph, let c be an H-colouring of G, and let τ be a function from X to $V(H)$. Define

$$\mathrm{Ans}^{\tau}((H,X),(G,c)) = \{h \in \mathrm{Ans}((H,X),G) \mid c(h(\cdot)) = \tau(\cdot)\}\}$$

Let $\hat{c}$ be an F-colouring of G. Define

$$\mathrm{Ans}^{\tau}((H,X),(G,\hat{c})) = \{h \in \mathrm{Ans}((H,X),G) \mid \gamma(\hat{c}(h(\cdot)) = \tau(\cdot)\}\}$$

Observation 37. *Let (G,c) be an H-coloured graph. Then*

$$|\mathrm{Ans}((H,X),G)| = \sum_{\tau : X \to V(H)} |\mathrm{Ans}^{\tau}((H,X),(G,c))|.$$

We will adapt Lemma 34 to conjunctive queries in the following way.

Lemma 38. *Let G be a graph and let c be an F-colouring of G, and assume that F is connected. Let $\tau\colon X \to V(H)$. Let $\vec{v} = (x_1,\dots,x_k)$. Let $\vec{z} = (z_1,\dots,z_k)$ be a k-tuple of positive integers. For all $i \in [k]$ let d_i be the number of vertices of H that are mapped by τ to x_i, i.e., $d_i := |\{u \in X \mid \tau(u) = x_i\}|$. Let $G' = \mathcal{G}(G,F,c,\vec{v},\vec{z})$ and $c' = \mathcal{C}(G,F,c,\vec{v},\vec{z})$. Then*

$$|\mathrm{Ans}^{\tau}((H,X),(G',c')| = |\mathrm{Ans}^{\tau}((H,X),(G,c))| \cdot \prod_{i=1}^{k} z_i^{d_i}.$$

Proof. We encourage the reader to consider Figure 4 for keeping track of the mappings and homomorphisms used in the proof. Let $\rho\colon V(G') \to V(G)$ be the function that maps cloned vertices to their original counterparts, that is

$$\rho(w) = \begin{cases} w & \text{if } w \in Y \\ \pi_1(w) & \text{otherwise} \end{cases}$$

Observe that ρ is a homomorphism from G' to G. We define an equivalence relation on the set $\mathrm{Ans}^{\tau}((H,X),(G',c'))$ by setting $h \sim h'$ if and only if $\rho(h(\cdot)) = \rho(h'(\cdot))$. For every $h \in \mathrm{Ans}^{\tau}((H,X),(G',c'))$, we have $\rho(h(\cdot)) \in \mathrm{Ans}((H,X),G)$.

Moreover, since $h \in \mathrm{Ans}^{\tau}((H,X),(G',c'))$, $\gamma(c'(h(\cdot))) = \tau$. From the definitions of ρ and c' it is immediate that $\gamma(c(\rho(\cdot))) = \gamma(c'(\cdot))$. Thus $\gamma(c(\rho(h(\cdot))) = \gamma(c'(h(\cdot))) = \tau$, proving that $\rho(h(\cdot)) \in \mathrm{Ans}^{\tau}((H,X),(G,c))$. Consequently, we can represent each equivalence class of $\sim$ by a homomorphism $\hat{h} \in \mathrm{Ans}^{\tau}((H,X),(G,c))$.

Finally, each equivalence class has size $\prod_{i=1}^{k} z_i^{d_i}$, since for each $i \in [k]$, there are z_i possibilities for each of the d_i vertices $u \in V(H)$ with $\tau(u) = v_i$. $\square$

$$
\begin{tikzcd}
 & X \arrow[d, "\tau"] \arrow[dddl, "\hat{h}"'] \arrow[dddr, "h"] & \\
 & V(H) & \\
 & V(F) \arrow[u, "\gamma"] & \\
V(G) \arrow[ur, "c"] & & V(G') \arrow[ul, "c'"'] \arrow[ll, "\rho"']
\end{tikzcd}
$$

Figure 4: Overview of the mappings and homomorphisms used in the proof of Lemma 38.

Lemma 40 is the main technical lemma in this section, which is concerned with $\chi(F, \emptyset)$ and $\chi(F, \{x_1\})$. Recall that the projection π_1 is an F-colouring of $\chi(F, \emptyset)$ and $\chi(F, \{x_1\})$. Moreover γ is a homomorphism from F to H. Thus:

Observation 39. *$\gamma(\pi_1(\cdot))$ is an H-colouring of both $\chi(F, \emptyset)$ and $\chi(F, \{x_1\})$.*

Recall that $\mathsf{Bij}(X)$ is the set of all bijections from X to X.

Lemma 40. *Let $c = \gamma(\pi_1(\cdot))$. Let $\vec{v} = (x_1, \ldots, x_k)$. Suppose that*

$$\sum_{\tau \in \mathsf{Bij}(X)} |\mathrm{Ans}^\tau((H, X), (\chi(F, \emptyset), c))| - |\mathrm{Ans}^\tau((H, X), (\chi(F, \{x_1\}), c))| \neq 0.$$

Then there is a k-tuple $\vec{z} = (z_1, \ldots, z_k)$ of positive integers such that

$$|\mathrm{Ans}((H, X), \mathcal{G}(\chi(F, \emptyset), F, c, \vec{v}, \vec{z}))| \neq |\mathrm{Ans}((H, X), \mathcal{G}(\chi(F, \{x_1\}), F, c, \vec{v}, \vec{z}))|.$$

Proof. Let $G_0 = \mathcal{G}(\chi(F, \emptyset), c, \vec{v}, \vec{z}))$, $c_0 = \mathcal{C}(\chi(F, \emptyset), F, c, \vec{v}, \vec{z}))$, $G_1 = \mathcal{G}(\chi(F, \{x_1\}), F, c, \vec{v}, \vec{z}))$, and $c_1 = \mathcal{C}(\chi(F, \{x_1\}), F, c, \vec{v}, \vec{z}))$,

Suppose for contradiction that, for every k-tuple $\vec{z}$ of positive integers, $|\mathrm{Ans}((H, X), G_0)| = |\mathrm{Ans}((H, X), G_1)|$. By Observation 37

$$\sum_{\tau \colon X \to V(H)} \Big(|\mathrm{Ans}^\tau((H, X), (G_0, c_0))| - |\mathrm{Ans}^\tau((H, X), (G_1, c_1))\Big)| = 0.$$

Let $d_i^\tau := |\{u \in X \mid \tau(u) = x_i\}|$. Applying Lemma 38,

$$\sum_{\tau \colon X \to V(H)} \Big(|\mathrm{Ans}^\tau((H, X), (\chi(F, \emptyset), c))| - |\mathrm{Ans}^\tau((H, X), (\chi(F, \{x_1\}), c))| \Big) \prod_{i=1}^{k} z_i^{d_i^\tau} = 0. \quad (2)$$

Define $b_\tau := |\mathrm{Ans}^\tau((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^\tau((H,X),(\chi(F,\{x_1\}),c))|$. Then, treating the z_i as variables, we can define a k-variate polynomial

$$P(z_1,\dots,z_k) := \sum_{\tau: X\to V(H)} b_\tau \cdot \prod_{i=1}^{k} z_i^{d_i^\tau}.$$

By (2), $P(z_1,\dots,z_k)=0$ for all k-tuples of positive integers $(z_1,\dots,z_k)$. We wish to apply polynomial interpolation, which requires us to collect coefficients for all monomials: To this end, observe first that all d_i^τ are bounded from above by k since $|X|=k$. Let $\mathcal{A}_k$ be the set of all k-tuples of integers $(a_1,\dots,a_k)$ with $0\le a_i\le k$ for each $i\in[k]$. Given a k-tuple $\vec{a}\in\mathcal{A}_k$, we write $\mathsf{coeff}(\vec{a})$ for the coefficient of the monomial $\prod_{i=1}^k z_i^{a_i}$, that is,

$$P(z_1,\dots,z_k) = \sum_{\vec{a}\in\mathcal{A}_k} \mathsf{coeff}(\vec{a}) \cdot \prod_{i=1}^{k} z_i^{a_i}.$$

By (2), $P(z_1,\dots,z_k)=0$ for all k-tuples of positive integers $(z_1,\dots,z_k)$. By multivariate polynomial interpolation, only the constant 0 polynomial can satisfy this conditions. Concretely, we obtain $\mathsf{coeff}(1,1,\dots,1)=0$. This yields the desired result since $\mathsf{Bij}(X)$ is the set of functions $\tau\colon X\to V(H)$ with $d_i^\tau=1$ for all $i\in[k]$. Thus,

$$\sum_{\tau\in\mathsf{Bij}(X)} |\mathrm{Ans}^\tau((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^\tau((H,X),(\chi(F,\{x_1\}),c))| = \mathsf{coeff}(1,1,\dots,1)=0,$$

contradicting the assumption in the statement of the lemma, and therefore completing the proof. □

Corollary 41. *Suppose that the treewidth of F is t and that F is connected. Let $c=\gamma(\pi_1(\cdot))$. Suppose that*

$$\sum_{\tau\in\mathsf{Bij}(X)} |\mathrm{Ans}^\tau((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^\tau((H,X),(\chi(F,\{x_1\}),c))| \neq 0.$$

Then the WL-dimension of (H,X) is at least t.

Proof. Let $\vec{v}=(x_1,\dots,x_k)$. By Lemma 35, for every k-tuple of positive integers $\vec{z}$,

$$\mathcal{G}(\chi(F,\emptyset),F,\pi_1,\vec{v},\vec{z}) \cong_{(t-1)} \mathcal{G}(\chi(F,\{x_1\}),F,\pi_1,\vec{v},\vec{z}).$$

From the definition of the cloning operation (Definition 33) and the fact that γ is the identity on X (see Definition 14), $\mathcal{G}(\chi(F,\emptyset),F,\pi_1,\vec{v},\vec{z}) = \mathcal{G}(\chi(F,\emptyset),F,c,\vec{v},\vec{z})$. Similarly, $\mathcal{G}(\chi(F,\{x_1\}),F,\pi_1,\vec{v},\vec{z}) = \mathcal{G}(\chi(F,\{x_1\}),F,c,\vec{v},\vec{z})$. So for every k-tuple of positive integers $\vec{z}$, $\mathcal{G}(\chi(F,\emptyset),F,c,\vec{v},\vec{z}) \cong_{(t-1)} \mathcal{G}(\chi(F,\{x_1\}),F,c,\vec{v},\vec{z})$.

However, by Lemma 40, there is a k-tuple $\vec{z}=(z_1,\dots,z_k)$ of positive integers such that

$$|\mathrm{Ans}((H,X),\mathcal{G}(\chi(F,\emptyset),F,c,\vec{v},\vec{z}))| \neq |\mathrm{Ans}((H,X),\mathcal{G}(\chi(F,\{x_1\}),F,c,\vec{v},\vec{z}))|.$$

This shows that the function $G\mapsto|\mathrm{Ans}((H,X),G)|$ can distinguish $(t-1)$-WL invariant graphs and thus the WL-dimension of (H,X) at least t. □

We will consider the set of partial automorphisms of a conjunctive query, defined as follows:

Definition 42 ($\mathsf{Aut}(H,X)$)**.** $\mathsf{Aut}(H,X) := \{\tau : X \to X \mid \exists a \in \mathsf{Aut}(H) : a|_X = \tau\}$.

Note that a mapping $\tau : X \to X$ can only be extended to an automorphism of H if it is bijective. Thus:

Observation 43. $\mathsf{Aut}(H,X) \subseteq \mathsf{Bij}(X)$

The following result will be very useful for the remainder of this section; it can be found in Corollary 54.4 in the full version [18] of [17].

Lemma 44 ([18, 17])**.** *Let h be a homomosphism from H to H that surjectively maps X onto X. If (H,X) is counting minimal then h is an automorphism of H.*

Lemma 45. *Let G be a graph, let c be a homomorphism from G to H, and let $\tau \in \mathsf{Bij}(X) \setminus \mathsf{Aut}(H,X)$. If (H,X) is counting minimal, then $|\mathrm{Ans}^{\tau}((H,X),(G,c))| = 0$.*

Proof. Assume for contradiction that there is a homomorphism $h \in \mathrm{Ans}^{\tau}((H,X),(G,c))$. Then there is $\varphi \in \mathrm{Hom}(H,G)$ with $\varphi|_X = h$. Moreover, $c(h(\cdot)) = \tau(\cdot)$. Thus $c(\varphi(\cdot))$ is a homomorphism from H to H such that, for all $x \in X$, $c(\varphi(x)) = \tau(x)$. Since $\tau \in \mathsf{Bij}(X)$, $c(\varphi(\cdot))$ maps X surjectively to itself. By Lemma 44, $c(\varphi(\cdot))$ is an automorphism of H, and thus $\tau \in \mathsf{Aut}(H,X)$, contradicting the assumption in the statement of the lemma that $\tau \notin \mathsf{Aut}(H,X)$. $\square$

Lemma 46. *Let G be a graph, let c be a homomorphism from G to H, and let $\tau \in \mathsf{Aut}(H,X)$. Then $|\mathrm{Ans}^{\tau}((H,X),(G,c))| = |\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))|$.*

Proof. We will construct a bijection between $\mathrm{Ans}^{\tau}((H,X),(G,c))$ and $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$.

Since $\tau \in \mathsf{Aut}(H,X)$ there is an automorphism a of H such that $a|_X = \tau$. Let a^{-1} be the inverse of a in the group $\mathsf{Aut}(H)$ and observe that $\tau^{-1} = a^{-1}|_X$.

We first show that, for every function $h \in \mathrm{Ans}^{\tau}((H,X),(G,c))$, the function $h(\tau^{-1}(\cdot))$ is in $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$. To see this, consider $h \in \mathrm{Ans}^{\tau}((H,X),(G,c))$. From the definition of $\mathrm{Ans}^{\tau}((H,X),(G,c))$, there is a homomorphism $\varphi \in \mathrm{Hom}(H,G)$ with $\varphi|_X = h$ and that $c(h(\cdot)) = \tau$. Since a^{-1} is an automorphism of H, $\varphi(a^{-1}(\cdot)) \in \mathrm{Hom}(H,G)$. Also, for every $x \in X$, $\varphi(a^{-1}(x)) = \varphi(\tau^{-1}(x))$. Since $\tau^{-1}(x) \in X$ and $\varphi|_X = h$, $\varphi(\tau^{-1}(x)) = h(\tau^{-1}(x))$. Putting these equalities together, $\varphi(a^{-1}(x)) = h(\tau^{-1}(x))$. Thus $h(\tau^{-1}(\cdot)) \in \mathrm{Ans}((H,X),G)$. Finally, since $h \in \mathrm{Ans}^{\tau}((H,X),(G,c))$, $c(h(\tau^{-1}(\cdot))) = \tau(\tau^{-1}(\cdot)) = \mathsf{id}(\cdot)$ so $h(\tau^{-1}(\cdot)) \in \mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$, as required.

Let b be the function that maps every $h \in \mathrm{Ans}^{\tau}((H,X),(G,c))$ to the function $b(h) := h(\tau^{-1}(\cdot))$. We have shown that b is a map from $\mathrm{Ans}^{\tau}((H,X),(G,c))$ to $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$. By a symmetric argument, the function $\hat{b}$ that maps every $h \in \mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$ to $\hat{b}(h) := h(\tau(\cdot))$ is a map from $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$ to $\mathrm{Ans}^{\tau}((H,X),(G,c))$.

Since $b(\hat{b}(\cdot)) = \mathsf{id}$ and $\hat{b}(b(\cdot)) = \mathsf{id}$, b and $\hat{b}$ are bijections, completing the proof. $\square$

Corollary 47 is the main tool that we will use to lower bound the WL-dimension of conjunctive queries.

Corollary 47. *Suppose that $\mathsf{tw}(F) = t > 1$, that F is connected, and that (H,X) is counting minimal. Let $c = \gamma(\pi_1(\cdot))$. Suppose that $|\mathrm{Ans}^{\mathsf{id}}((H,X),(\chi(F,\emptyset),c))|$ is not equal to $|\mathrm{Ans}^{\mathsf{id}}((H,X),(\chi(F,\{x_1\}),c))|$. Then the WL-dimension of (H,X) is at least t.*

Proof. By Corollary 41 it suffices to establish the following.

$$\sum_{\tau \in \mathsf{Bij}(X)} |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\{x_1\}),c))| \neq 0.$$

By Observation 39, c is an H-colouring of both $\chi(F,\emptyset)$ and $\chi(F,\{x_1\})$. Thus we can use Lemmas 45 and 46, and obtain:

$$\begin{aligned}
&\sum_{\tau \in \mathsf{Bij}(X)} |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\{x_1\}),c))| \\
&= \sum_{\tau \in \mathsf{Aut}(H,X)} |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^{\tau}((H,X),(\chi(F,\{x_1\}),c))| \\
&= |\mathsf{Aut}(H,X)| \cdot \Big(|\mathrm{Ans}^{\mathsf{id}}((H,X),(\chi(F,\emptyset),c))| - |\mathrm{Ans}^{\mathsf{id}}((H,X),(\chi(F,\{x_1\}),c))| \Big).
\end{aligned}$$

The first factor is non-zero because $\mathsf{Aut}(H,X)$ is non-empty. The second factor is non-zero by the assumption in the statement of the corollary. □

4.4 Proving the Lower Bound

We first set up some notation, following [17].

Definition 48 (Colour-prescribed Homomorphisms)**.** Let H and G be graphs, let (H,X) be a conjunctive query, and let c be an H-colouring of G. Define

$$\begin{aligned}
\mathsf{cpHom}(H,(G,c)) &= \{h \in \mathrm{Hom}(H,G) \mid \forall v \in V(H), c(h(v)) = v\}, \text{and} \\
\mathsf{cpAns}((H,X),(G,c)) &= \{a\colon X \to V(G) \mid \ \exists h \in \mathsf{cpHom}(H,(G,c)) \text{ such that } h|_X = a \ \}.
\end{aligned}$$

Homomorphisms in $\mathsf{cpHom}(H,(G,c))$ are said to be "colour-prescribed" with respect to c.

Observation 49. $\mathsf{cpAns}((H,X),(G,c)) \subseteq \mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$

Observation 49 follows directly from the definitions. Recall that $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c)) = \{h \in \mathrm{Ans}((H,X),G) \mid \forall v \in X, c(h(v)) = v\}\}$. So $c(h(v)) = v$ is required for all $v \in X$ in the definition of $\mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$, while for all $v \in V(H)$ in the definition of $\mathsf{cpAns}((H,X),(G,c))$. Lemma 50 shows equivalence for counting minimal conjunctive queries.

Lemma 50. *Let (H,X) be a counting minimal conjunctive query, let G be a graph, and let c be an H-colouring of G. Then $\mathsf{cpAns}((H,X),(G,c)) = \mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$.*

Proof. Observation 49 proves one direction. For the other direction, consider a map $a \in \mathrm{Ans}^{\mathsf{id}}((H,X),(G,c))$. From the definition, there is a homomorphism $h \in \mathrm{Hom}(H,G)$ such that $h|_X = a$ and $c(a(x)) = x$ for all $x \in X$. From the definitions of c and h, the function $\varphi := c(h(\cdot))$ is a homomorphism from H to itself. Since φ maps X to X and (H,X) is counting minimal, Lemma 44 guarantees that φ is an automorphism of H. Let $h' = h(\varphi^{-1}(\cdot))$. Clearly, $h' \in \mathrm{Hom}(H,G)$. Moreover, for each $v \in V(H)$, $c(h'(v)) = c(h(\varphi^{-1}(v))) = \varphi(\varphi^{-1})(v) = v$. Thus $h' \in \mathsf{cpHom}(H,(G,c))$. Since φ is the identity on X, the same is true of φ^{-1}. Thus, for all $x \in X$, $h'(x) = h(\varphi^{-1}(x)) = h(x) = a(x)$, implying that $a \in \mathsf{cpAns}((H,X),(G,c))$. □

The following definition will be useful; it will provide a parity condition which is both sufficient and necessary for containment in the relevant set of answers.

Definition 51 (Extendable Assignments, $\mathcal{E}(X,F,W)$)**.** Let (H,X) be a conjunctive query with $\emptyset \subsetneq X = \{x_1,\ldots,x_k\} \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $Y = V(H) \setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Let $C_1,\ldots,C_m$ be the connected components of $H[Y]$. For each $i \in [m]$, denote the vertex sets of the ℓ copies of C_i in F by $V_i^1,\ldots,V_i^\ell$. Let W be a subset of X and let φ be an assignment from X to $V(\chi(F,W))$ such that, for all $p \in [k]$, $c(\varphi(x_p)) = x_p$. Define $S_1,\ldots,S_k$ such that $\varphi(x_p) = (x_p, S_p)$ for all $p \in [k]$. We say that φ is *extendable* if the following two conditions hold.

(E1) For every $\{x_a,x_b\} \in E(H[X])$, $x_a \in S_b \iff x_b \in S_a$.

(E2) For every $i \in [m]$ there is a $j \in [\ell]$ such that $\sum_{p=1}^{k} |S_p \cap V_i^j|$ is even.

Define $\mathcal{E}(X,F,W) := \{\varphi : X \to \chi(F,W) \mid \varphi \text{ is extendable and } \forall x \in X,\, c(\varphi(x)) = x\}$.

We will prove Lemmas 52, 53 and 54 in Section 4.4.1.

Lemma 52. *Let (H,X) be a conjunctive query with $\emptyset \subsetneq X \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $V(H)\setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Then $|\mathcal{E}(X,F,\emptyset)| > |\mathcal{E}(X,F,\{x_1\})|$.*

Lemma 53. *Let (H,X) be a conjunctive query with $\emptyset \subsetneq X = \{x_1,\ldots,x_k\} \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $V(H)\setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Let W be a subset of X and let φ be an assignment from X to $V(\chi(F,W))$ such that, for all $p \in [k]$, $c(\varphi(x_p)) = x_p$. If φ is not extendable, then $\varphi \notin \mathsf{cpAns}((H,X),(\chi(F,W),c))$.*

Lemma 54. *Let (H,X) be a conjunctive query with $\emptyset \subsetneq X = \{x_1,\ldots,x_k\} \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $V(H)\setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Let W be a subset of X and let φ be an assignment from X to $V(\chi(F,W))$ such that, for all $p \in [k]$, $c(\varphi(x_p)) = x_p$. If φ is extendable, then $\varphi \in \mathsf{cpAns}((H,X),(\chi(F,W),c))$*

Lemma 55. *Let (H,X) be a conjunctive query with $\emptyset \subsetneq X \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $V(H)\setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Let W be a subset of X. Then $\mathsf{cpAns}((H,X),(\chi(F,W),c)) = \mathcal{E}(X,F,W)$.*

Proof. Let φ be an assignment from X to $V(\chi(F,W))$. If there is an $x \in X$ such that $c(\varphi(x)) \neq x$ then $\varphi \notin \mathcal{E}(X,F,W)$ and $\varphi \notin \mathsf{cpAns}((H,X),(\chi(F,W),c))$. Otherwise, if φ is not extendable then it not in $\mathcal{E}(X,F,W)$ by definition and it is not in $\mathsf{cpAns}((H,X),(\chi(F,W),c))$ by Lemma 53. If φ is extendable then it is in $\mathcal{E}(X,F,W)$ by definition, Also, by Lemma 54, φ is in $\mathsf{cpAns}((H,X),(\chi(F,W),c))$. □

Lemma 56. *Let (H,X) be a conjunctive query with $\emptyset \subsetneq X \subsetneq V(H)$ and suppose that x_1 is adjacent to at least one vertex in $V(H)\setminus X$. Suppose that H is connected and let ℓ be an odd positive integer. Let $F = F_\ell(H,X)$ and let $c = \gamma(\pi_1(\cdot))$. Then $|\mathsf{cpAns}((H,X),(\chi(F,\emptyset),c))| > |\mathsf{cpAns}((H,X),(\chi(F,\{x_1\}),c))|$.*

Proof. The lemma follows immediately from Lemmas 52 and 55. □

We can now state and prove Lemma 57, the main goal of this subsection, which will enable us to immediately infer Theorem 24.

Lemma 57. *Let (H, X) be a counting minimal conjunctive query such that H is connected, and $\emptyset \subsetneq X \subsetneq V(H)$. Without loss of generality, suppose that x_1 is adjacent to at least one vertex in $V(H) \setminus X$. Let ℓ be an odd positive integer. Let $F = F_\ell(H, X)$ and let $c = \gamma(\pi_1(\cdot))$. Then*

$$|\mathrm{Ans}^{\mathsf{id}}((H, X), (\chi(F, \emptyset), c))| > |\mathrm{Ans}^{\mathsf{id}}((H, X), (\chi(F, \{x_1\}), c))|$$

Proof. The lemma follows directly from Lemma 50 and lemma 56. □

We can now prove Theorem 24.

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 $\mathsf{ew}(H, X)$.*

Proof. By Corollary 18, $\mathsf{ew}(H, X) = \max\{\mathsf{tw}(F_\ell(H, X)) \mid \ell > 0\}$. Choose ℓ large enough such that $\mathsf{ew}(H, X) = \mathsf{tw}(F_\ell(H, X))$. We can assume, without loss of generality, that ℓ is odd: If ℓ is even, choose $\ell + 1$ instead and note that $\mathsf{tw}(F_{\ell+1}(H, X)) \geq \mathsf{tw}(F_\ell(H, X))$ since $F_\ell(H, X)$ is a subgraph of $F_{\ell+1}(H, X)$ and since treewidth is monotone under taking subgraphs.

Let $F = F_\ell(H, X)$ (note that F is connected as (H, X) is connected) and let $c = \gamma(\pi_1(\cdot))$. By Corollary 47, it suffices to show that $|\mathrm{Ans}^{\mathsf{id}}((H, X), (\chi(F, \emptyset), c))|$ is not equal to $|\mathrm{Ans}^{\mathsf{id}}((H, X), (\chi(F, \{x_1\}), c))|$, which holds by Lemma 57. This concludes the proof. □

4.4.1 Proofs of Lemmas 52, 53 and 54

This section will use the following combinatorial lemma, Lemma 58, which seems to be folklore.

Lemma 58. *Let G be a connected graph and let $S \subseteq V(G)$ be a vertex-subset of even cardinality. Then there is an assignment $\beta : E(G) \to \{0, 1\}$ such that, for all $v \in V(G)$,*

$$\sum_{u \in N(v)} \beta(\{u, v\}) \equiv \begin{cases} 1 \mod 2 & \text{if } v \in S, \\ 0 \mod 2 & \text{if } v \notin S. \end{cases} \tag{3}$$

Proof. The proof is by induction on $n = |V(G)|$. If $n = 1$, then $E(G) = \emptyset$ so necessarily $S = \emptyset$. By convention, the (empty) sum is 0 which is as desired since $v \notin S$.

Assume for the induction hypothesis that the claim holds for n and let $G = (V, E)$ be a graph with $n + 1$ vertices. Let v be a vertex of G such that $G' := G \setminus \{v\}$ is connected. Let $v_1, \ldots, v_d$ be the neighbours of v. Note that $d > 0$ since G is connected. We consider two cases.

Case 1. If $v \in S$: Set $\beta^*(\{v, v_1\}) = 1$ and $\beta^*(\{v, v_i\}) = 0$ for all $1 < i \leq d$. Set $S' = (S \setminus v) \oplus v_1$. Note that S' has even cardinality, since $S \setminus v$ has odd cardinality. We can thus apply the induction hypothesis to G' and S' and obtain a function $\beta' : E(G') \to \{0, 1\}$ that satisfies (3) for G'. It is easy to see that $\beta^* \cup \beta'$ satisfies (3) for G.

Case 2. If $v \notin S$: Set $\beta^*(\{v, v_i\}) = 0$ for all $i \in [d]$. We apply the induction hypothesis to G' and S and obtain a function $\beta' : E(G') \to \{0, 1\}$ that satisfies (3) in G'. It is easy to see that $\beta^* \cup \beta'$ satisfies (3) in G.

□

We finish by proving Lemmas 53, 54 and 52.

Proof of Lemma 53. Let $Y = V(H) \setminus X$. Let $G = \chi(F, W)$. For all $p \in [k]$, $i \in [m]$ and $j \in [\ell]$, define S_p, C_i, and V_i^j as in Definition 51. If (E1) is not satisfied, then φ cannot be extended to a homomorphism from H to G. If (E2) is not satisfied then there is an $i \in [m]$ such that, for all $j \in [\ell]$, $\sum_{p=1}^{k} |S_p \cap V_i^j|$ is odd. Fix this i. Assume for contradiction that $\varphi \in \mathsf{cpAns}((H, X), (G, c))$. Then there is a homomorphism $h \in \mathsf{cpHom}(H, (G, c))$ with $h|_X = \varphi$.

Let $y_1, \ldots, y_t$ be the vertices of C_i. Since $h \in \mathsf{cpHom}(H, (G, c))$, the definition of c implies that for all $v \in V(H)$, $\gamma(\pi_1(h(v))) = v$. Recall that the vertices of $F = F_\ell(H, X)$ are $X \cup (Y \times [\ell])$ and that the vertices of $G = \chi(F, W)$ are pairs (w, T) where w is a vertex of F and $T \subseteq N(w)$. So for a vertex $v = y_s$ of C_i, $\gamma(\pi_1(h(y_s))) = y_s$ implies that $h(y_s)$ is of the form $((y_s, j), T_s)$ for some $j \in [\ell]$. Moreover, since C_i is a connected component of $H[Y]$ there is a single j such that, for all $s \in [t]$, $h(y_s) = ((y_s, j), T_s)$.

We next define a t-by-$(k + t)$ matrix M, indexed by vertices of F. The rows of M are indexed by $(y_1, j), \ldots, (y_t, j)$, and the columns of M are indexed $x_1, \ldots, x_k, (y_1, j), \ldots, (y_t, j)$. The entries of M are defined as follows. For $s \in [t]$,

$$M((y_s, j), v) = \begin{cases} 1 & \text{if } v \in T_s, \\ 0 & \text{if } v \notin T_s. \end{cases}$$

From the definition of $G = \chi(F, W)$, the size of T_s is odd if $(y_s, j) \in W$ and even otherwise. Since W is a subset of X, every set T_s has even cardinality. Thus every row of M has an even number of 1s, and therefore M has an even number of 1s. Let M_X be the submatrix of M containing only the columns indexed by $x_1, \ldots, x_k$. Let M_Y be the submatrix of M containing only the columns indexed by $(y_1, j), \ldots, (y_t, j)$.

Note that M_Y is a square matrix. We next show that it is symmetric. If $M((y_s, j), (y_{s'}, j)) = 1$, then $(y_{s'}, j) \in T_s$, which implies that (y_s, j) and $(y_{s'}, j)$ are adjacent in F and thus y_s and $y_{s'}$ are adjacent in H. Suppose for contradiction that $M((y_{s'}, j), (y_s, j)) = 0$. Then $(y_s, j) \notin T_{s'}$. So $((y_s, j), T_s)$ and $((y_{s'}, j), T_{s'})$ are not adjacent in $G = \chi(F, W)$, contradicting the fact that h is a homomorphism from H to G. Thus, our assumption was wrong, and M_Y is symmetric. Since the diagonal of M_Y contains only 0s (since H and F do not have self-loops), the number of 1s in M_Y is even.

To finish the proof we will show that, for every $p \in [k]$, the number of 1s in the column of M indexed by x_p is $|S_p \cap V_i^j|$ so the number of 1s in M_X is $\sum_{p=1}^{k} |S_p \cap V_i^j|$, which is odd by the choice of i. Thus, M_X has an odd number of 1s and M_Y has an even number of 1s, contradicting the fact that M has an even number of 1s. We conclude that our initial assumption for contradiction, that $\varphi \in \mathsf{cpAns}((H, X), (G, c))$, is false, proving the lemma.

So to finish, fix $p \in [k]$. We will show that the number of 1s in the column of M indexed by x_p is $|S_p \cap V_i^j|$.

Before considering the entries of this column of M, we establish a fact which we will use twice — if x_p and (y_s, j) are adjacent in F then, from the definition of F, x_p and y_s are adjacent in H. Since h is a homomorphism from H to G that extends φ, from the definition of S_p, $h(x_p) = \varphi(x_p) = (x_p, S_p)$, so it follows that (x_p, S_p) is adjacent to $h(y_s) = ((y_s, j), T_s)$ in G. Using this fact, we will show that the number of 1s in the column of M indexed by x_k is $|S_p \cap V_i^j|$.

First, consider any s such that that $M((y_s, j), x_p) = 1$. From the definition of M, $x_p \in T_s$ so, from the definition of $G = \chi(F, W)$, x_p and (y_s, j) are adjacent in F. From the fact, (x_p, S_p)

is adjacent to $((y_s, j), T_s)$ in G. From the definition of G, $(y_s, j) \in S_p$. By construction, $(y_s, j) \in V_i^j$, so $(y_s, j) \in S_p \cap V_i^j$.

Finally, consider any s such that that $M((y_s, j), x_p) = 0$. From the definition of M, $x_p \notin T_s$. There are two cases.

Case 1: If x_p and (y_s, j) are not adjacent in F then, from the definition of S_p, (y_s, j) is not in S_p so it is clearly not in $S_p \cap V_i^j$.

Case 2. If x_p and (y_s, j) are adjacent in F but $x_p \notin T_s$ then from the fact (x_p, S_p) is adjacent to $((y_s, j), T_s)$ in G. So from the definition of G, $(y_s, j) \notin S_p$ so it is clearly not in $S_p \cap V_i^j$.

As required, we thus obtain that the number of 1s in the column of M indexed by x_k is $|S_p \cap V_i^j|$.

$\square$

Proof of Lemma 54. Let $Y = V(H) \setminus X$. Let $G = \chi(F, W)$. For all $p \in [k]$, $i \in [m]$ and $j \in [\ell]$, define S_p, C_i, and V_i^j as in Definition 51. Condition (E1) ensures that, for every $\{x_a, x_b\} \in E(H[X])$, (x_a, S_a) and (x_b, S_b) are adjacent in G. By (E2), for every $i \in [m]$ there is a $j_i \in [\ell]$ such that $\sum_{p=1}^{k} |S_p \cap V_i^{j_i}|$ is even.

To show that $\varphi \in \mathsf{cpAns}((H, X), (G, c))$, we will construct an $h \in \mathsf{cpHom}(H, (G, c))$ with $h|_X = \varphi$. To do this we will choose a value for $h(y)$ for each $y \in Y$. For every $i \in [m]$, let $V(C_i) := \{y_{i,1}, \ldots, y_{i,t_i}\}$ be the vertices of the connected component C_i of $H[Y]$. Given a vertex $y_{i,s} \in V(C_i)$, let $N_i(y_{i,s})$ denote the set of its neighbours in C_i. For every $y_{i,s} \in V(C_i)$, let $T_{i,s,X} = \{x_p \in X \mid (y_{i,s}, j_i) \in S_p\}$.

Let $\Omega_i = \{y_{i,s} \ : \ |T_{i,s,X}| \text{ is odd}\}$. We will show that $|\Omega_i|$ is even. To see this, let M_X be a matrix with rows indexed by $(y_{i,1}, j_i), \ldots, (y_{i,t_i}, j_i)$ and columns indexed by X. Define $M(y_{i,s}, j_i), x_p)$ to be 1 if $x_p \in T_{i,s,X}$ and 0 otherwise. By construction, the number of 1s in the column of M indexed by x_p is the number of vertices $(y_{i,s}, j_i) \in S_p$ so this is $|S_p \cap V_i^{j_i}|$. By the choice of j_i, the number of 1s in M_X is even. So Ω_i, which contains the indices of rows with an odd number of 1s, has even cardinality. Apply Lemma 58 with graph C_i and $S = \Omega_i$ to obtain an assignment $\beta\colon E(C_i) \to \{0, 1\}$ such that, for all $y_{i,s} \in \Omega_i$, $\sum_{y_{i,s'} \in N_i(y_{i,s})} \beta(\{y_{i,s'}, y_{i,s}\})$ is odd and for all $y_{i,s} \in V(C_i) \setminus \Omega_i$, this sum is even. Finally, let $T_{i,s,Y} = \{(y_{i,s'}, j_i) \mid y_{i,s'} \in N_i(y_{i,s}) \text{ and } \beta(\{y_{i,s'}, y_{i,s}\}) = 1\}$ and let $T_{i,s} = T_{i,s,X} \cup T_{i,s,Y}$. We then define $h(y_{i,s}) = ((y_{i,s}, j_i), T_{i,s})$. Our goal is to show that $h \in \mathsf{cpHom}(H, (G, c))$. For this we require

- Property 1: Each $h(y_{i,s})$ is a vertex of G.
- Property 2: For all $y_{i,s} \in Y$, $c(h(y_{i,s})) = y_{i,s}$. (This follows immediately from the definition of h.)
- Property 3: h is a homomorphism from H to G. That is, all three of the following hold.
  - Property 3a: For every edge $\{x_a, x_b\}$ of $E(H[X])$, $\{h(x_a), h(x_b)\}$ is an edge of G (this follows immediately from (E1) and the definition of G).
  - Property 3b: For every edge $\{x_p, y_{i,s}\}$ of H with $x_p \in X$ and $y_{i,s} \in C_i$, $\{h(x_p), h(y_{i,s})\}$ is an edge of G.
  - Property 3c: For every edge $\{y_{i,s}, y_{i,s'}\}$ of C_i, $\{h(y_{i,s}), h(y_{i,s'})\}$ is an edge of G.

We start by showing Property 1 — that each $h(y_{i,s})$ is a vertex of G. For this, we need two constraints to be satisfied as follows.

- Constraint 1: Every vertex of $T_{i,s}$ must be a neighbour of $(y_{i,s}, j_i)$ in F.
- Constraint 2: $|T_{i,s}|$ must be even (this constraint comes from the definition of G since $y_{i,s}$ is not in W, which is a subset of X).

Constraint 1 follows directly from the definition of $T_{i,s}$. To see that every $x_p \in T_{i,s,X}$ is a neighbour of $(y_{i,s}, j_i)$ note that $(y_{i,s}, j_i) \in S_p$ which implies that $\{x_p, (y_{i,s}, j_i)\}$ is an edge of F since (x_p, S_p) is a vertex of G. It is immediate from the definition of $T_{i,s,Y}$ and F that every vertex in $T_{i,s,Y}$ is a neighbour of $(y_{i,s}, j_i)$ in F.

Constraint 2 is by construction. Consider any pair (i, s). If $|T_{i,s,X}|$ is even then $y_{i,s} \notin \Omega_i$ so $|T_{i,s,Y}|$ is even. On the other hand, if $|T_{i,s,X}|$ is odd then $Y_{i,s} \in \Omega_i$, so $|T_{i,s,Y}|$ is odd.

We next consider Property 3b. Consider an edge $\{x_p, y_{i,s}\}$ of H with $x_p \in X$ and $y_{i,s} \in C_i$. Then $h(x_p) = (x_p, S_p)$. If $(y_{i,s}, j_i) \in S_p$ then $x_p \in T_{i,s,X}$ so $x_p \in T_{i,s}$ and $h(y_{i,s}) = ((y_{i,s}, j_i), T_{i,s})$ is connected to (x_p, S_p) in G by the definitions of F and G. Similarly, if $(y_{i,s}, j_i) \notin S_p$ then $x_p \notin T_{i,s}$ so again $((y_{i,s}, j_i), T_{i,s})$ is connected to (x_p, S_p) in G.

To finish the proof, we establish Property 3c. Consider an edge $\{y_{i,s}, y_{i,s'}\}$ of C_i. Then $h(y_{i,s}) = ((y_{i,s}, j_i), T_{i,s})$ and $h(y_{i,s'}) = ((y_{i,s'}, j_i), T_{i,s'})$ By construction, $(y_{i,s'}, j_i) \in T_{i,s}$ iff $(y_{i,s}, j_i) \in T_{i,s'}$. Hence $h(y_{i,s})$ and $h(y_{i,s'})$ are connected in G. □

Proof of Lemma 52. Let $Y = V(H) \setminus X$. For $i \in [m]$ and $j \in [\ell]$, define C_i and V_i^j as in Definition 51. For every subset W of X and every map $\varphi\colon X \to V(\chi(F, W))$ that satisfies $c(\varphi(x_p)) = x_p$ for all $p \in [k]$, let $S_1(\varphi), \ldots, S_k(\varphi)$ be the sets defined in Definition 51. For every $\varphi \in \mathcal{E}(X, F, W)$, (E2) guarantees that for every $i \in [m]$ there is a $j_i \in [\ell]$ such that $\sum_{p=1}^k |S_p(\varphi) \cap V_i^{j_i}|$ is even. We will partition $\mathcal{E}(X, F, W)$ into disjoint sets in terms of i and j_i as follows. First we define $\mathcal{E}(X, F, W, 1)$ by fixing $i = 1$ and $j_i > 1$.

$$\mathcal{E}(X, F, W, 1) := \{\varphi \in \mathcal{E}(X, F, W) \mid \exists j_1 > 1 : \textstyle\sum_{p=1}^k |S_p(\varphi) \cap V_1^{j_1}| \text{ is even}\}.$$

For all $i \in \{2, \ldots, m\}$, define

$$\mathcal{E}(X, F, W, i) := \{\varphi \in \mathcal{E}(X, F, W) \backslash (\cup_{q=1}^{i-1} \mathcal{E}(X, F, W, q)) \mid \exists j_i > 1 : \textstyle\sum_{p=1}^k |S_p(\varphi) \cap V_i^{j_i}| \text{ is even}\}.$$

Finally, define $\mathcal{E}(X, F, W, 0) := \mathcal{E}(X, F, W) \setminus (\bigcup_{i=1}^m \mathcal{E}(X, F, W, i))$.

Since the sets $\mathcal{E}(X, F, W, 0), \ldots, \mathcal{E}(X, F, W, m)$ are a disjoint partition of $\mathcal{E}(X, F, W)$ for every subset W of X, the lemma follows immediately from the following three claims, which clearly imply $|\mathcal{E}(X, F, \emptyset)| > |\mathcal{E}(X, F, \{x_1\}|$, as required.

Claim 1. For all $i \in [m]$, $|\mathcal{E}(X, F, \emptyset, i)| = |\mathcal{E}(X, F, \{x_1\}, i)|$.

Fix $i \in [m]$. To prove Claim 1, we construct a bijection b from $\mathcal{E}(X, F, \emptyset, i)$ to $\mathcal{E}(X, F, \{x_1\}, i)$. Since H is connected, there is path from C_i to x_1 in H that starts at some vertex $y \in C_i$, takes an edge from y to X and does not re-visit C_i before reaching x_1. So there is a path from $(y, 1)$ to x_1 in F whose vertices are in $X \cup \{(y', 1) \mid y' \in Y\}$ and is of the form $P = (y, 1)x_{t_1}P_1x_{t_2}P_2x_{t_3}P_3 \ldots x_{t_{s-1}}P_{s-1}x_{t_s}$ where $x_{t_1}, \ldots, x_{t_s}$ are distinct vertices in X with $t_s = 1$ and each P_j is either empty or for some $i(j) \in [m] \setminus \{i\}$, it is a non-empty simple path in F whose vertices are in $V^1_{i(j)}$. Without loss of generality, the path visits every connected component of $H[Y]$ at most once so for any distinct j and j' such that P_j and $P_{j'}$ are both non-empty, $i(j) \neq i(j')$.

Our goal is to define the bijection b. For each $\varphi \in \mathcal{E}(X, F, \emptyset, i)$ and each $p \in [k]$, we first define a subset $S'_p(\varphi)$ of vertices of F. Let $N_P(x_p)$ be the neighbours of x_p in the path P (if x_p is not in the path P, then $N_P(x_p) = \emptyset$). Then define $S'_p(\varphi) = S_p(\varphi) \oplus N_P(x_p)$. Finally, we define $b(\varphi)$ to be the map which maps every $p \in [k]$ to $(x_p, S'_p(\varphi))$.

We wish to show that $(x_p, S'_p(\varphi))$ is a vertex of $\chi(F, \{x_1\})$. Since $(x_p, S_p(\varphi))$ is a vertex of $\chi(F, \emptyset)$, $S_p(\varphi)$ is a subset of $N_F(x_p)$. So, by construction, $S'_p(\varphi)$ is a subset of $N_F(x_p)$. Each set $S_p(\varphi)$ has even cardinality. Note that $N_P(x_p)$ has even cardinality unless $p = 1$, in which case it has odd cardinality. Thus, $S'_p(\varphi)$ has even cardinality unless $p = 1$, in which case it has odd cardinality.

The map b is a bijection since it can be inverted using $S_p(\varphi) = S'_p(\varphi) \oplus N_P(p)$. The map $b(\varphi)$ satisfies (E1) since φ satisfies (E1) and $x_a \in N_P(x_b)$ iff $x_b \in N_P(x_a)$. The map $b(\varphi)$ It satisfies (E2) since the definition of $\mathcal{E}(X, F, \emptyset, i)$ guarantees $j_i > 1$ so $S'_p(\varphi) \cap V_i^{j_i} = S_p(\varphi) \cap V_i^{j_i}$. It satisfies $c(\varphi(x_p)) = x_p$ for all $x_p \in X$ by construction. Thus, $b(\varphi) \in \mathcal{E}(X, F, \{x_1\})$. Finally, the same $j_i > 1$ that shows $\varphi \in \mathcal{E}(X, F, \emptyset, i)$ shows that $b(\varphi)$ is in $\mathcal{E}(X, F, \{x_1\}, i)$.

Claim 2. $|\mathcal{E}(X, F, \emptyset, 0)| > 0$.

To prove Claim 2, we exhibit a $\varphi \in \mathcal{E}(X, F, \emptyset, 0)$. Let $Z \subseteq Y$ be a set containing exactly one vertex from each component $C_1, \ldots, C_m$ such that every vertex in Z is adjacent to X in H. For each $z \in Z$, let $p(z) = \min\{p \in [k] \mid (x_p, z) \in E(H)\}$. For each $p \in [k]$ let $S_p = \{z \in Z \mid p(z) = p\} \times \{2, \ldots, \ell\}$. Since $\ell > 1$ is odd, $|S_p|$ is even. Let φ be the map from X to $V(\chi(F, \emptyset))$ such that, for all $p \in [k]$, $\varphi(x_p) = (x_p, S_p)$. The map φ satisfies (E1) since $S_p \cap X = \emptyset$ for all $p \in [k]$. It satisfies (E2) for any $i \in [m]$ by taking $j = 1$ since for all $p \in [k]$, $|S_p \cap V_i^1| = 0$. It satisfies $c(\varphi(x_p)) = x_p$ for all $p \in [k]$ by construction so $\varphi \in \mathcal{E}(X, F, \emptyset)$. To see that $\varphi \in \mathcal{E}(X, F, \emptyset, 0)$, fix any $i \in [m]$ and any $j \in \{2, \ldots, \ell\}$. Let z be the unique vertex in $Z \cap V(C_i)$. Then (z, j) is the unique vertex in $S_{p(z)} \cap V_i^j$ Furthermore, for $p' \neq p(z)$ with $p' \in [k]$, $S_{p'} \cap V_i^j$ is empty. Thus, $\sum_{p=1}^k |S_p \cap V_i^j|$ is odd.

Claim 3. $|\mathcal{E}(X, F, \{x_1\}, 0)| = 0$.

We conclude by proving Claim 3. Assume for contradiction that $\varphi \in \mathcal{E}(X, F, \{x_1\}, 0)$. Let $G = \chi(F, \{x_1\})$. As in Definition 51, define $S_1, \ldots, S_k$ such that, for all $p \in [k]$, $\varphi(x_p) = (x_p, S_p)$. Since $\varphi(x_p)$ is in $V(G)$, $|S_1|$ is odd and for every $p \in \{2, \ldots, k\}$, $|S_p|$ is even. We will analyse a matrix M. The rows of M are indexed by $x_1, \ldots, x_k$ and the columns of M are indexed by the vertices of F. For every vertex v of F, $M(x_p, v)$ is defined to be 1 if $v \in S_p$ and 0 otherwise. Since $|S_1|$ is odd and $|S_2|, \ldots, |S_k|$ are even, the total number of 1s in M is odd.

We split M into two submatrices: M_X consists of the columns indexed by vertices in X and M_Y consists of the columns indexed by the remaining vertices of F. For every $p \in [k]$, $x_p \notin S_p$ since (x_p, S_p) is a vertex of G and F has no self-loops. Also, M_X is symmetric by (E1). So the number of 1s in M_X is even.

Now consider M_Y. Since $\varphi \in \mathcal{E}(X, F, \{x_1\})$, (E2) implies that for every $i \in [m]$ there is a $j \in [\ell]$ such that $\sum_{p=1}^k |S_p \cap V_i^j|$ is even. Since $\varphi \in \mathcal{E}(X, F, \{x_1\}, 0)$, for each $i \in [m]$, the following conditions hold.

(C1) $\sum_{p=1}^k |S_p \cap V_i^1|$ is even, and

(C2) for all $j > 1$, $\sum_{p=1}^k |S_p \cap V_i^j|$ is odd.

For each $i \in [m]$ and $j \in [\ell]$, let M_i^j be the submatrix of M_Y containing only the columns indexed by the vertices in V_i^j. Then for all $i \in [m]$, condition (C1) implies that M_i^1 contains an even number of 1s and condition (C2) implies that for all $j \in \{2, \dots, \ell\}$, M_i^j contains an odd number of 1s. Since ℓ is odd, the total number of 1s in the matrices $M_i^1, \dots, M_i^\ell$ is even. Consequently, the total number of 1s in M_Y is even, contradicting the fact that M has an odd number of 1s and M_X has an even number of 1s. $\square$

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 $\mathsf{sew}(H, X)$.*

Proof. We first consider the special case where (H, X) is a *full conjunctive queries*, 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 [17]. Moreover, $\Gamma(H, X) = H$. Thus $\mathsf{sew}(H, X) = \mathsf{tw}(H)$. 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 $\mathsf{tw}(H)$ as shown by Neuen [34].

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 [18, Section 6]. Theorems 24 and 21 now state that the WL-dimension of (H', X') is equal to $\mathsf{ew}(H', X')$. Finally, by definition of semantic extension width, we have $\mathsf{sew}(H, X) = \mathsf{ew}(H', X')$, concluding the proof. $\square$

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 $|\mathrm{Hom}(F, G)| = |\mathrm{Hom}(F, G')|$ for all $F \in \mathcal{F}$. 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 [29]). Recent years have seen numerous exciting results on the structure of $\mathcal{F}$-indistinguishability, depending on the class $\mathcal{F}$: For example, Dvorák [21], and Dell, Grohe and Rattan [16] 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 [30].

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

Definition 59. Let Ψ be a class of conjunctive queries. Two graphs G and G' are Ψ-*indistinguishable*, denoted by $G \cong_{\Psi} G'$, if $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$ for all queries $(H, X) \in \Psi$.

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

Corollary 60 (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 $G \cong_{\Psi_k} G'$.*

Proof. For the first direction, suppose that $G \cong_k G'$ and consider $(H, X) \in \Psi_k$. Then $\mathsf{sew}(H, X) \le k$ and thus the WL-dimension of (H, X) is at most k by Theorem 1. Consequently, $|\mathrm{Ans}((H, X), G)| = |\mathrm{Ans}((H, X), G')|$. This shows that $G \cong_{\Psi_k} G'$.

For the other direction, suppose that $G \cong_{\Psi_k} G'$. Recall that $\mathsf{sew}(H, V(H)) = \mathsf{tw}(H)$ for all h. Thus Ψ_k contains all conjunctive queries $(H, V(H))$ with $\mathsf{tw}(H) \le k$. Since $\mathrm{Ans}((H, V(H)), F) = \mathrm{Hom}(H, F)$ for all H and F, $G \cong_{\Psi_k} G'$ implies $G \cong_{\mathcal{F}'} G'$ where $\mathcal{F}'$ is the class of all conjunctive queries $(H, V(H))$ where H is a *connected* graph with treewidth at most k. Finally, we can remove the connectivity constraint as follows: Let $F = F_1 \cup F_2$ be the disjoint union of two graphs F_1 and F_2. If F has treewidth at most k, then both F_1 and F_2 also have treewidth at most k, since treewidth is monotone under taking subgraphs. Thus, $G \cong_{\mathcal{F}'} G'$ implies $\mathrm{Hom}(F, G) = \mathrm{Hom}(F_1, G) \cdot \mathrm{Hom}(F_2, G) = \mathrm{Hom}(F_1, G') \cdot \mathrm{Hom}(F_2, G') = \mathrm{Hom}(F, G')$. Consequently $G \cong_{\mathcal{F}} G'$ where $\mathcal{F}$ is the class of all queries $(H, V(H))$ such that H is a graph with treewidth at most k and thus $G \cong_k G'$. □

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 61. *The class of acyclic conjunctive queries have 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.*

Proof. The corollary follows immediately from Theorem 1 and the fact that acyclic conjunctive queries can have arbitrary high semantic extension width. Recall the k-star query (S_k, X_k), where $X_k = \{x_1, \dots, x_k\}$, $V(S_k) = X_k \cup \{y\}$ and $E(S_k) = \{\{x_i, y\} \mid i \in [k]\}$. Clearly, (S_k, X_k) is acyclic. Moreover, it is well-known that (s_k, X_k) is counting minimal (see e.g. [17]). Finally, $\Gamma(S_k, X_k)$ is the $k+1$-clique, and thus $\mathsf{sew}(S_k, X_k) = \mathsf{tw}(K_{k+1}) = k$. □

Corollary 61 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 [21, 16]. Given Corollary 61 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 62. *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* $|\mathrm{Ans}((H, X), 2K_3)| = |\mathrm{Ans}((H, X), C_6|$.

Proof. For the disjoint union of two conjunctive queries $(H_1, X_1) \cup (H_2, X_2)$ and any graph G, $|\mathrm{Ans}((H_1, X_1) \cup (H_2, X_2), G)| = |\mathrm{Ans}((H_1, X_1), G)| \cdot |\mathrm{Ans}((H_2, X_2), G)|$. Thus, it suffices to prove the observation for connected queries.

Let (H, X) be any connected acyclic conjunctive query. If $X = \emptyset$ then the observation is trivial since computing $G \mapsto |\mathrm{Ans}((H, X), G)|$ is equivalent to deciding whether there is a homomorphism from H to G. As H is acyclic, we thus have $|\mathrm{Ans}((H, X), 2K_3)| = |\mathrm{Ans}((H, X), C_6| = 1$.

Now assume that $X \neq \emptyset$. For the proof, we associate the query (H, X) with an edge-weighted tree T as follows: The vertices of T are X, and two (distinct) vertices x_1, x_2 of T are adjacent if and only if there is a path from x_1 to x_2 in H, the intermediate vertices of which are all existentially quantified variables, i.e., contained in $V(H) \setminus X$. Note that there can be at most one such path since (H, X) is acyclic. The number of intermediate vertices on this path will become the weight of the edge $\{x_1, x_2\}$ in T; note that the weight is 0 if and only if x_1 and x_2 are adjacent in H. Let us write $w : E(T) \to \mathbb{N}$ for the weight function of the edges of T.

Now it is easy to see that, for any graph G without isolated vertices, the elements of $\mathrm{Ans}((H, X), G)$ are precisely the mappings $\varphi : V(T) \to V(G)$ such that for every $e = \{x_1, x_2\} \in E(T)$ there is a (not necessarily simple) walk from $\varphi(x_1)$ to $\varphi(x_2)$ with $w(e)$ internal vertices.

Let us write $\#\mathrm{Ans}((T, w), G)$ for the set of such mappings. It remains to show that $\#\mathrm{Ans}((T, w), C_6) = \#\mathrm{Ans}((T, w), 2K_3)$. We prove this claim by induction on $n = |V(T)|$. Since $V(T) = X$ and $X \neq \emptyset$, the induction base is $n = 1$, for which we have that $\#\mathrm{Ans}((T, w), C_6) = \#\mathrm{Ans}((T, w), 2K_3) = 6$.

For the induction step, fix a leaf x of T, let x' be its neighbour, and let $\hat{w} = w(\{x, x'\})$. Moreover, let $T' = T \setminus \{x\}$ and $w' = w|_{E(T')}$. By the induction hypothesis we have $\#\mathrm{Ans}((T', w'), C_6) = \#\mathrm{Ans}((T', w'), 2K_3)$. Now fix *any* pair of queries $\varphi_1 \in \mathrm{Ans}((T', w'), C_6)$ and $\varphi_2 \in \mathrm{Ans}((T', w'), 2K_3)$. Let s_1 and s_2 be the number of extensions of φ_1 and φ_2 that yield an element in $\mathrm{Ans}((T, w), C_6)$ and $\mathrm{Ans}((T, w), 2K3)$, respectively, that is, $s_1 = |V_1|$ and $s_2 = |V_2|$ where

$$V_1 = \{v \in V(C_6) \mid \varphi_1 \cup \{x \mapsto v\} \in \mathrm{Ans}((T, w), C_6)\}$$
$$V_2 = \{v \in V(2K_3) \mid \varphi_1 \cup \{x \mapsto v\} \in \mathrm{Ans}((T, w), 2K_3)\}$$

Next we claim

$$s_1 = s_2 = \begin{cases} 2 & \hat{w} = 0 \\ 3 & \hat{w} > 0 \end{cases}$$

Note that proving this claim concludes the proof since it implies that

$$\#\mathrm{Ans}((T, w), C_6) = s \cdot \#\mathrm{Ans}((T', w'), C_6) = s \cdot \#\mathrm{Ans}((T', w'), 2K_3) = \#\mathrm{Ans}((T, w), 2K3),$$

where $s \in \{2, 3\}$ depends only on $\hat{w}$.

Finally, to prove the claim, assume C_6 has vertices $\{0, \dots, 5\}$ and edges $\{i, i + 1 \mod 6\}$. Moreover, assume that $2K3$ has vertices $\{a, b, c, a', b', c'\}$ and that the triangles are $\{a, b, c\}$ and $\{a', b', c'\}$. Now let $u_1 = \varphi_1(x')$ and $u_2 = \varphi_2(x')$. W.l.o.g. assume that $u_1 = 0$ and $u_2 = a$.

- If $\hat{w} = 0$ then x must be mapped to a neighbour of the image of x'. Since both $2K3$ and C_6 are 2-regular, there are precisely 2 options for each case.
- If $\hat{w}$ is odd, then $V_1 = \{0, 2, 4\}$ and $V_2 = \{a, b, c\}$. Thus $n_1 = n_2 = 3$.
- Otherwise $\hat{w}$ is positive and even. Then $V_1 = \{1, 3, 5\}$ and $V_2 = \{a, b, c\}$. Thus $n_1 = n_2 = 3$.

This concludes the proof. □

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 $\#\mathrm{CQ}(\Psi)$ takes as input a pair consisting of a conjunctive query $(H,X) \in \Psi$ and a graph G and outputs $|\mathrm{Ans}((H,X),G)|$.

Recall 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.

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

Proof. Given a conjunctive query (H,X), recall the definition of the graph $\Gamma(H,X)$ from Definition 11. The *contract* of (H,X) is the induced graph subgraph $\Gamma[X]$ (see [17, Definition 8]).

Using the notion of contract, Chen, Durand, and Mengel [20, 11] and Dell, Roth and Wellnitz [17] have established an exhaustive complexity classification for $\#\mathrm{CQ}(\Psi)$. For an explicit statement, Theorem 10 in [17] states that (conditioned on $\mathrm{FPT} \neq W[1]$), $\#\mathrm{CQ}(\Psi)$ is solvable in polynomial time if and only if both the treewidth of Ψ and the treewidth of the *contracts* of Ψ are bounded.

By Theorem 1, the WL-dimension of a connected counting minimal query with $X \neq \emptyset$ is equal to its extension width. It remains to show that the condition that both the treewidth of Ψ and the treewidth of the contracts of Ψ are bounded is the same as the condition that Ψ has bounded extension width.

First, suppose that Ψ has bounded extension width. Then for some quantity B, every query (H,X) in Ψ has extension width at most B. So $\mathsf{tw}(\Gamma(X,H)) \leq B$. Since H and $\Gamma[X]$ are both subgraphs of $\Gamma(X,H)$, they both have treewidth at most B, as required.

For the other direction, suppose that the treewidth of every query in Ψ and every contract of every query in Ψ is at most B. Consider a query $(H,X) \in \Psi$. Let $t_1 \leq B$ be the treewidth of H and let $t_2 \leq B$ be the treewidth of $\Gamma[X]$. Let $\mathcal{T} = (T, \mathcal{B})$ be an optimal tree-decomposition of $\Gamma[X]$. Let $C_1, \dots, C_d$ be the connected components of $H[Y]$. For each $i \in [d]$ let δ_i be the subset of vertices in X that are adjacent to C_i in H. By the definition of Γ, each δ_i induces a clique in Γ, and thus in $\Gamma[X]$. Therefore, there is a bag B_i that contains δ_i (see e.g. [6, Lemma 3]). Since the treewidth of $\Gamma[X]$ is t_2 it follows that $|\delta_i| \leq t_2 + 1$.

Next, for each $i \in [d]$, let $\mathcal{T}_i$ be an optimal tree decomposition of C_i and note that the width of all $\mathcal{T}_i$ is at most t_1 since all components C_i are subgraphs of H. It is now straightforward to construct a tree-decomposition of width at most $t_1 + t_2$ of $\Gamma(H,X)$. For every $i \in [d]$, we add δ_i to each bag of $\mathcal{T}_i$. Finally, we fix an arbitrary node v_i of the tree T_i of $\mathcal{T}$ and connect it to the node of T with bag B_i. Clearly (T1)-(T3) are satisfied.

□

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 [29]). Recently, staring with the work of Chen and Mengel [12] and of Curticapean, Dell and Marx [14], the study of these linear combinations has re-arisen in the context of parameterised counting complexity theory. Moreover, the works of Seppelt [38], Neuen [34], and Lanzinger and Barceló [28] 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" [29, Chapter 6.1], we formalise our linear combinations as follows.

Definition 63 (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 [12], and Dell, Roth, and Wellnitz [17] 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 64 (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 $G \mapsto |\mathrm{Ans}(Q, G)|$. The following lemma was shown by Seppelt [38] in the special case of homomorphisms, i.e., the case in which each constituent (H_i, X_i) satisfies $X_i = V(H_i)$. The proof of the generalised version follows the same idea, combining Seppelt's approach and Theorem 1.

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

Proof. Let $Q = \sum_{i=1}^{\ell} c_i \cdot (H_i, X_i)$ and $k = \mathsf{hsew}(Q)$. We first show that the WL-dimension of Q is upper bounded by k; this is the easy direction. For each $i \in [\ell]$, let $k_i = \mathsf{sew}(H_i, X_i)$. By the definition of $\mathsf{hsew}(Q)$, $k = \max_{i \in [\ell]} k_i$.

By Theorem 1, the WL-dimension of the function $G \mapsto |\mathrm{Ans}(H_i, X_i)|$ is k_i. Let G and G' be graphs such that $G \cong_k G'$ (which means that for every $i \in [\ell]$, $G \cong_{k_i} G'$). Then for every $i \in [\ell]$, $|\mathrm{Ans}((H_i, X_i), G)| = |\mathrm{Ans}((H_i, X_i), G')|$ so

$$|\mathrm{Ans}(Q, G)| = \sum_{i=1}^{\ell} c_i \cdot |\mathrm{Ans}((H_i, X_i), G)| = \sum_{i=1}^{\ell} c_i \cdot |\mathrm{Ans}((H_i, X_i), G')| = |\mathrm{Ans}(Q, G')|.$$

This means that the function $G \mapsto |\mathrm{Ans}(Q, G)|$ cannot distinguish k-WL-equivalent graphs so the WL-dimension of Q is at most k.

The more difficult direction is the lower bound. To this end, we will construct graphs F and F' such that $F \cong_{k-1} F'$ and $|\mathrm{Ans}(Q, F)| \neq |\mathrm{Ans}(Q, F')|$. This implies that the WL-dimension of the graph parameter $G \mapsto |\mathrm{Ans}(Q, G)|$ is greater than $k-1$ so the WL-dimension of Q is at least k.

Assume without loss of generality that $\mathsf{sew}(H_1, X_1) = k$. By Theorem 1 the WL-dimension of (H_1, X_1) is k so k is the minimum positive integer such that counting answers from (H_1, X_1) cannot distinguish k-equivalent graphs. Thus there are graphs G and G' with $G \cong_{k-1} G'$ and $|\mathrm{Ans}((H_1, X_1), G)| \neq |\mathrm{Ans}((H_1, X_1), G')|$.

The *tensor product* of two graphs A and B, denoted by $A \otimes B$, has the property that, for every graph H, $|\mathrm{Hom}(H, A \otimes B)| = |\mathrm{Hom}(H, A)| \cdot |\mathrm{Hom}(H, B)|$. See, for example, [29, Chapters 3.3 and 5.2.3]. For each graph T of treewidth at most $k-1$, $|\mathrm{Hom}(T, G \otimes H)| = |\mathrm{Hom}(T, G)| \cdot |\mathrm{Hom}(T, H)| = |\mathrm{Hom}(T, G')| \cdot |\mathrm{Hom}(T, H)| = |\mathrm{Hom}(T, G' \otimes H)|$ Thus, for all graphs H, $G \otimes H \cong_{k-1} G' \otimes H$.

Suppose for contradiction that, for all H, $|\mathrm{Ans}(Q, G \otimes H)| = |\mathrm{Ans}(Q, G' \otimes H)|$. This rewrites to

$$\sum_{i=1}^{\ell} c_i(|\mathrm{Ans}((H_i, X_i), G)| - |\mathrm{Ans}((H_i, X_i), G')|) \cdot |\mathrm{Ans}((H_i, X_i), H)| = 0.$$

Let $d_i := c_i(|\mathrm{Ans}((H_i, X_i), G)| - |\mathrm{Ans}((H_i, X_i), G')|)$ and $m = \max\{|V(H_i)| \mid i \in [\ell]\}$. Let $\mathcal{H}$ be the set of all graphs with at most m^m vertices. Then we obtain a system of linear equations containing the following equation for each $H \in \mathcal{H}$.

$$\sum_{i=1}^{\ell} d_i \cdot |\mathrm{Ans}((H_i, X_i), H)| = 0\,.$$

It was shown in [18, Lemma 34(iii)] that the matrix corresponding to this system has full rank. Therefore, for all $i \in [\ell]$, $d_i = 0$. This implies $|\mathrm{Ans}((H_1, X_1), G)| = |\mathrm{Ans}((H_1, X_1), G')|$, contradicting the choice of G and G'. Thus our assumption was wrong, and there is a graph H (in fact $H \in \mathcal{H}$) such that $|\mathrm{Ans}(Q, G \otimes H)| \neq |\mathrm{Ans}(Q, G' \otimes H)|$. Since $G \otimes H \cong_{k_1} G' \otimes H$, the proof is concluded. ◻

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

The k-star is often written in the more prominent form $\varphi(x_1, \dots, x_k) = \exists y : \bigwedge_{i=1}^{k} E(x_i, y)$. It is well-known that (S_k, X_k) is counting minimal (see e.g. [17]). Moreover, $\Gamma(S_k, X_k)$ is the $(k+1)$-clique, which has treewidth k. Thus $\mathsf{sew}(S_k, X_k) = k$. Corollary 67 follows immediately from Theorem 1.

Corollary 67. *The WL-dimension of (S_k, X_k) is k.*

We can now prove Corollary 6.

Corollary 68 (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 $\ell \le k$. By the principle of inclusion and exclusion, for each graph G,

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

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 $J = \emptyset$.[2] By Corollary 5 and the fact that $\mathsf{sew}(S_\ell, X_\ell) = \ell$ for all $\ell \in [k]$, the WL-dimension of $G \mapsto |\mathsf{Inj}((S_k,X_k),G)|$ 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 $1 \le k' < k$. 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 [38]) 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,

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

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

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

This concludes the proof. ◻

[2] 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 [14], 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.

References

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

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

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

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

[5] Hans L. Bodlaender. Necessary edges in k-chordalisations of graphs. *J. Comb. Optim.*, 7(3):283–290, 2003. doi:10.1023/A:1027320705349.

[6] Hans L. Bodlaender and Arie M. C. A. Koster. Safe separators for treewidth. *Discret. Math.*, 306(3):337–350, 2006. doi:10.1016/j.disc.2005.12.017.

[7] Jan Böker. Color refinement, homomorphisms, and hypergraphs. In Ignasi Sau and Dimitrios M. Thilikos, editors, *Graph-Theoretic Concepts in Computer Science - 45th International Workshop, WG 2019, Vall de Núria, Spain, June 19-21, 2019, Revised Papers*, volume 11789 of *Lecture Notes in Computer Science*, pages 338–350. Springer, 2019. doi:10.1007/978-3-030-30786-8_26.

[8] Giorgos Bouritsas, Fabrizio Frasca, Stefanos Zafeiriou, and Michael M. Bronstein. Improving graph neural network expressivity via subgraph isomorphism counting. *IEEE Trans. Pattern Anal. Mach. Intell.*, 45(1):657–668, 2023. doi:10.1109/TPAMI.2022.3154319.

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

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

[11] Hubie Chen and Stefan Mengel. A trichotomy in the complexity of counting answers to conjunctive queries. In Marcelo Arenas and Martín Ugarte, editors, *18th International*

Conference on Database Theory, ICDT 2015, March 23-27, 2015, Brussels, Belgium, volume 31 of *LIPIcs*, pages 110–126. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2015. doi:10.4230/LIPIcs.ICDT.2015.110.

[12] Hubie Chen and Stefan Mengel. Counting answers to existential positive queries: A complexity classification. In Tova Milo and Wang-Chiew Tan, editors, *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*, pages 315–326. ACM, 2016. doi:10.1145/2902251.2902279.

[13] Zhengdao Chen, Lei Chen, Soledad Villar, and Joan Bruna. Can graph neural networks count substructures? In Hugo Larochelle, Marc'Aurelio Ranzato, Raia Hadsell, Maria-Florina Balcan, and Hsuan-Tien Lin, editors, *Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, 2020, virtual*, 2020. URL: https://proceedings.neurips.cc/paper/2020/hash/75877cb75154206c4e65e76b88a12712-Abstract.html.

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

[15] Anuj Dawar, Tomás Jakl, and Luca Reggio. 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*, pages 1–13. IEEE, 2021. doi:10.1109/LICS52264.2021.9470609.

[16] Holger Dell, Martin Grohe, and Gaurav Rattan. Lovász meets weisfeiler and leman. In Ioannis Chatzigiannakis, Christos Kaklamanis, Dániel Marx, and Donald Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming, ICALP 2018, July 9-13, 2018, Prague, Czech Republic*, volume 107 of *LIPIcs*, pages 40:1–40:14. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2018. doi:10.4230/LIPIcs.ICALP.2018.40.

[17] Holger Dell, Marc Roth, and Philip Wellnitz. Counting answers to existential questions. In Christel Baier, Ioannis Chatzigiannakis, Paola Flocchini, and Stefano Leonardi, editors, *46th International Colloquium on Automata, Languages, and Programming, ICALP 2019, July 9-12, 2019, Patras, Greece*, volume 132 of *LIPIcs*, pages 113:1–113:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ICALP.2019.113.

[18] Holger Dell, Marc Roth, and Philip Wellnitz. Counting answers to existential questions. *CoRR*, abs/1902.04960, 2019. URL: http://arxiv.org/abs/1902.04960, arXiv:1902.04960.

[19] Yon Dourisboure and Cyril Gavoille. Tree-decompositions with bags of small diameter. *Discret. Math.*, 307(16):2008–2029, 2007. doi:10.1016/j.disc.2005.12.060.

[20] Arnaud Durand and Stefan Mengel. Structural tractability of counting of solutions to conjunctive queries. *Theory Comput. Syst.*, 57(4):1202–1249, 2015. doi:10.1007/s00224-014-9543-y.

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

[22] Jörg Flum and Martin Grohe. *Parameterized Complexity Theory*. Texts in Theoretical Computer Science. An EATCS Series. Springer, 2006. doi:10.1007/3-540-29953-X.

[23] Martin Fürer. Weisfeiler-lehman refinement requires at least a linear number of iterations. In Fernando Orejas, Paul G. Spirakis, and Jan van Leeuwen, editors, *Automata, Languages and Programming, 28th International Colloquium, ICALP 2001, Crete, Greece, July 8-12, 2001, Proceedings*, volume 2076 of *Lecture Notes in Computer Science*, pages 322–333. Springer, 2001. doi:10.1007/3-540-48224-5_27.

[24] Martin Grohe. 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*, pages 1–17. IEEE, 2021. doi:10.1109/LICS52264.2021.9470677.

[25] Martin Grohe, Kristian Kersting, Martin Mladenov, and Pascal Schweitzer. Color Refinement and Its Applications. In *An Introduction to Lifted Probabilistic Inference*. The MIT Press, 08 2021. arXiv:https://direct.mit.edu/book/chapter-pdf/2101088/c025000_9780262365598.pdf, doi:10.7551/mitpress/10548.003.0023.

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

[27] Jiashun Jin, Zheng Tracy Ke, and Shengming Luo. Network global testing by counting graphlets. In Jennifer G. Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning, ICML 2018, Stockholmsmässan, Stockholm, Sweden, July 10-15, 2018*, volume 80 of *Proceedings of Machine Learning Research*, pages 2338–2346. PMLR, 2018. URL: http://proceedings.mlr.press/v80/jin18b.html.

[28] Matthias Lanzinger and Pablo Barceló. On the power of the weisfeiler-leman test for graph motif parameters. *CoRR*, abs/2309.17053, 2023. arXiv:2309.17053, doi:10.48550/arXiv.2309.17053.

[29] László Lovász. *Large Networks and Graph Limits*, volume 60 of *Colloquium Publications*. American Mathematical Society, 2012.

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

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

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

[33] Christopher Morris, Martin Ritzert, Matthias Fey, William L. Hamilton, Jan Eric Lenssen, Gaurav Rattan, and Martin Grohe. 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*, pages 4602–4609. AAAI Press, 2019. doi:10.1609/aaai.v33i01.33014602.

[34] Daniel Neuen. Homomorphism-distinguishing closedness for graphs of bounded tree-width. *arXiv preprint arXiv:2304.07011*, 2023.

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

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

[37] Benjamin Scheidt and Nicole Schweikardt. Counting homomorphisms from hypergraphs of bounded generalised hypertree width: A logical characterisation. In Jérôme Leroux, Sylvain Lombardy, and David Peleg, editors, *48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France*, volume 272 of *LIPIcs*, pages 79:1–79:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.MFCS.2023.79.

[38] Tim Seppelt. Logical equivalences, homomorphism indistinguishability, and forbidden minors. In Jérôme Leroux, Sylvain Lombardy, and David Peleg, editors, *48th International Symposium on Mathematical Foundations of Computer Science, MFCS 2023, August 28 to September 1, 2023, Bordeaux, France*, volume 272 of *LIPIcs*, pages 82:1–82:15. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. doi:10.4230/LIPIcs.MFCS.2023.82.

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

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

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

[42] Keyulu Xu, Weihua Hu, Jure Leskovec, and Stefanie Jegelka. 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, 2019. URL: https://openreview.net/forum?id=ryGs6iA5Km.