Relational Colour Refinement for Non-Relational Signatures

Bachelor's Thesis

Theodor Jurij Teslia

August 28, 2025

RWTH Aachen University

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Ut purus elit, vestibulum ut, placerat ac, adipiscing vitae, felis. Curabitur dictum gravida mauris. Nam arcu libero, nonummy eget, consectetuer id, vulputate a, magna. Donec vehicula augue eu neque. Pellentesque habitant morbi tristique senectus et netus et malesuada fames ac turpis egestas. Mauris ut leo. Cras viverra metus rhoncus sem. Nulla et lectus vestibulum urna fringilla ultrices. Phasellus eu tellus sit amet tortor gravida placerat. Integer sapien est, iaculis in, pretium quis, viverra ac, nunc. Praesent eget sem vel leo ultrices bibendum. Aenean faucibus. Morbi dolor nulla, malesuada eu, pulvinar at, mollis ac, nulla. Curabitur auctor semper nulla. Donec varius orci eget risus. Duis nibh mi, congue eu, accumsan eleifend, sagittis quis, diam. Duis eget orci sit amet orci dignissim rutrum.

Nam dui ligula, fringilla a, euismod sodales, sollicitudin vel, wisi. Morbi auctor lorem non justo. Nam lacus libero, pretium at, lobortis vitae, ultricies et, tellus. Donec aliquet, tortor sed accumsan bibendum, erat ligula aliquet magna, vitae ornare odio metus a mi. Morbi ac orci et nisl hendrerit mollis. Suspendisse ut massa. Cras nec ante. Pellentesque a nulla. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Aliquam tincidunt urna. Nulla ullamcorper vestibulum turpis. Pellentesque cursus luctus mauris.

Other thesis template?

1 Introduction

The graph isomorphism problem is a very interesting and important problem in both theoretical and applied computer science. The question, whether two graphs are structurally identical comes up in many fields of research. For example when trying to recognize identical chemical structures for patent verification [14] or finding symmetries for optimizing problems such as the Boolean Satisfiability problem [11]. Furthermore, the graph isomorphism problem is interesting from a complexity theoretic perspective, as it is a prominent example for a problem, which has not been proven to be either NP-complete or solvable in polynomial time. [1]

One (incomplete) algorithm that is often used in practical isomorphism problem solvers [11] is *Colour Refinement*, short CR, also called the *1-dimensional Weisfeiler-Leman algorithm*. Given two graphs, it can prove in quasilinear time that they are not isomorphic [3]. Concretely, CR is an iterative algorithm that, in the beginning, assigns every vertex the same colour and in following iterations assigns each one a new colour, based on the colours of its neighbours. This procedure is repeated, until the partition of the vertices induced by the colouring stays the same. We then say that Colour Refinement distinguishes two graphs, if there is some colour that appears differently often in the two graphs. It is easy to see that two isomorphic graphs are not distinguished by Colour Refinement. This is equivalent to the fact that if two graphs get distinguished by Colour Refinement, then they cannot be isomorphic. Furthermore, while it is not possible to infer the opposite direction, that is two non-isomorphic graphs always get distinguished by CR, it was shown by Babai, Erdős and Selkow that almost all graphs get distinguished by it [2]. However, there exist some classes of graphs, that cannot be distinguished by Colour Refinement, for example the classes of regular graphs with the same number of vertices.

Aside from isomorphism testing, Colour Refinement has applications in different fields. Incidentally, the first recorded occurrence of this algorithm appeared in 1965 and dealt with the description of chemical structures [13]. Its significance for computer science has been recognised later by Weisfeiler and Leman in 1968 [18]. One interesting application of Colour Refinement is in the reduction of the dimension of linear programs. By defining a variant of Colour Refinement on matrices which finds a partition of the rows and columns, it is possible to reformulate a linear program with a considerably smaller dimension. This method of first reducing the problem and then solving the reduced instance has been shown to be more performant than the standard way of solving linear programs. [10] Another application can be found in the field of machine learning, more precisely for kernel methods. Generally, the aim of kernel methods is to assign a similarity value between two elements, which then can be used in more complex machine learning techniques such as support vector machines or regression. An emerging concept in this field are kernels for graphs, also called graph kernels, which are a method to compare two graphs, and represent how similar they are with a single value. Standard graph kernel methods consider certain subgraphs and count the (dis-)similarities. [17] One interesting application of Colour Refinement can be found in its usage as a graph kernel. When fixating an integer h and counting for each of the first h Colour Refinement steps how many vertices between two graphs share the same colour, we get a graph kernel. This Weisfeiler-Leman Graph Kernel has an adequate ability to to classify graphs, while having a significantly better runtime than classical graph kernels. [9]

The importance of CR in theoretical computer science can be seen, when we consider other characterisations of its distinguishing power. Thus, CR can be equivalently characterised by counting homomorphisms from trees and by considering a certain logic. Due to Dvořák [8] and Dell, Grohe and Rattan [7] we have the following characterisation. Given two graphs G and H, we have that Colour Refinement distinguishes them if, and only if, there is some tree T such that the number of homomorphisms from T to G is different than to H. Such a characterisation can also be done through logic. We define C_2 as the logic that extends first-order logic by counting quantifiers and only uses up to two variables. Then, it was the shown by Cai and Immerman [6] and Immerman and Lander [12] that the following holds: Colour Refinement distinguishes G and H if, and only if, there is a sentence

$\varphi \in \mathsf{C}_2$ such that $G \models \varphi$ and $H \not\models \varphi$.

From the above examples it can be seen that while Colour Refinement is a simple procedure, it can be applied in various situations. This versatility has been one of the reasons for its success and poses the question, whether it could be possible to formulate an analogous procedure for more than graphs. One obvious extension of graphs are hypergraphs. These are structures with a set of vertices and edges between those. However, while the edges of classical graphs connect two vertices, edges of hypergraphs can include an arbitrary number of them. One interesting result due to Böker [4] is an extension of Colour Refinement to hypergraphs, which gives rise to an analogous characterisation using homomorphism counting. Concretely, Colour Refinement for a hypergraph G is defined like classical CR on a coloured variant of the incidence graph of G. We then consider connected Berge-acyclic hypergraphs, that is hypergraphs whose incidence graphs are trees. When counting homomorphisms from those to hypergraphs, we get that Colour Refinement distinguishes two hypergraphs G and H if, and only if, there is some connected Berge-acyclic hypergraph B, such that B has a different number of homomorphisms to G than to H. Another result with respect to hypergraphs has been achieved by Scheidt and Schweickardt in [15]. They devised a 2-sorted counting logic called GC^k and proved that two hypergraphs G and H satisfy exactly the same GC^k sentences if, and only if, all hypergraphs with generalised hypertree width k have the same number of homomorphisms to G as to H. For the case where $k = 1$, we then get indistinguishability over the class of all α-acyclic hypergraphs, which is a strictly stronger measure of acyclicity than Berge-acyclicity. Interestingly, we will encounter α-acyclicity and GC^1 in section 3 for characterising relational structures, instead of hypergraphs. A first effort to extend Colour Refinement to relational structures has been made by Butti and Dalmau in [5]. They also defined Colour Refinement on the incidence graph of a relational structure and proved that this distinguishes two relational structures if, and only if, there is a connected Berge-acyclic relational structure with a different number of homomorphisms to the structures. A more recent result with respect to relational structures has been made by Scheidt and Schweikardt [16]. They defined an extension of Colour Refinement for relational structures, called Relational Colour Refinement, short RCR, which is stronger than the version in [5]. Thus it can distinguish structures that the variant in [5] cannot. Furthermore, they were able to define the logic $\mathsf{GF}(\mathsf{C})$, which characterises RCR in the same way as C_2 characterises classical CR. Additionally, the aforementioned α-acyclic structures characterise RCR as well. Concretely, we have that two relational structures of the same signature get distinguished by RCR if, and only if, there is an α-acyclic structure of the same signature such that is has a different amount of homomorphisms to the structures. A deeper discussion of the results from [16] can be found in section 3.

It can be seen that Colour Refinement for graphs, as for relational structures and hypergraphs have been studied. Especially the results from [16] seem like a very robust and usable extension for relational structures. Furthermore, we notice that functional, as well as non-relational signatures have not yet been investigated. As many practical and useful structures use non-relational signatures, for example all algebraic structures, we pose the question how robust the results from [16] are when using them with (unary) functions. Concretely, we consider two possible ways of how Relational Colour Refinement can be adapted to signatures with functions. For both approaches we investigate, whether they can be characterised by $\mathsf{GF}(\mathsf{C})$ over signatures with functions and by counting homomorphisms from acyclic structures with non-relational signatures. We will see that such a characterisation through logic is in fact possible, while counting homomorphisms from acyclic structures is not. Additionally, we prove a similar result to the (not existing) characterisations by counting homomorphisms. We show that while it is not possible to require that the acyclic structure is functional (given a signature with functions), it is possible to require the acyclic structure to be symmetric (given that the two distinguished structures are symmetric as well). Therefore, we get that two symmetric structures are distinguished by RCR if, and only if, there is some symmetric, acyclic structure with different numbers of homomorphisms to the structures.

The results of this thesis are as follows. We define two variants of RCR for non-relational signatures, called naive RCR and RCR_k for a $k \in \mathbb{N}$, where the latter is only defined for unary functions. Naive RCR encodes n-ary functions f directly as a $(n+1)$-ary relations R_f, such that $(x_1, \dots, x_n, y) \in R_f$ if, and only if, $f(x_1, \dots, x_n) = y$ and then applies normal RCR to the relational structure. RCR_k encodes unary functions f as a family of relations R_{f^1}, R_{f^2}, ..., where $(x, y) \in R_{f^i}$ if, and only if, $f^i(x) = y$, that is the n-times application of f on x. Furthermore we define $\mathsf{nfGF(C)}$ as the logic that is defined as $\mathsf{GF(C)}$ over non-relational signatures but does not allow the nesting of terms. Additionally we define $\mathsf{GF(C)}_k$ as the logic that is the normal extension of $\mathsf{GF(C)}$ to non-relational signatures but terms are only allowed to have an alternation-depth of function-applications of k. We then achieve the following results.

Theorem 2: For a non-relational signature σ and two σ-structures $\mathfrak{A}$ and $\mathfrak{B}$, we have that naive RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ if, and only if, there is a sentence $\varphi \in \mathsf{nfGF(C)}$ of signature σ such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$.

Theorem 9: For a non-relational signature σ that only contains relation-symbols and unary function-symbols, two σ-structures $\mathfrak{A}$ and $\mathfrak{B}$ and a $k \in \mathbb{N}$, we have that RCR_k distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ if, and only if, there is a sentence $\varphi \in \mathsf{GF(C)}_k$ of signature σ such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$.

Theorem ??: There exist two families of structures $(\mathfrak{A}_i)_{i\in\mathbb{N}}$ and $(\mathfrak{B}_i)_{i\in\mathbb{N}}$ such that naive RCR distinguishes $\mathfrak{A}_i$ and $\mathfrak{B}_i$ for all $i > 4$, but there does not contain a total, functional and acyclic structure that distinguishes $\mathfrak{A}_i$ and $\mathfrak{B}_i$ by number of homomorphism.

adapt to actual notation

Theorem 24: For two symmetric, relational structures $\mathfrak{A}$ and $\mathfrak{B}$, we have that RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ if, and only if, there is a symmetric, acyclic structure that distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ by number of homomorphism.

The methods used to achieve the above results rely heavily on the proofs from [16]. For the logical characterisation of structures with functions we translate both structures to relational structures and translate any formula to a valid $\mathsf{GF(C)}$-formula over a relational signature. The same is done in reverse as well. For the characterisation through homomorphism counting we present two families of structures. We prove that two structures from the families of the same size get distinguished by naive RCR but there cannot be an acyclic structure of the same signature such that it distinguishes them by homomorphism count.

The structure of this thesis is as follows. We begin by defining notation and fundamental definitions in section 2. Then in section 3 we present and explain the results from [16]. Afterwards, we continue in section 4 by considering structures with functions, where we will first show the logical characterisation for the two approaches, before then presenting the family of counterexamples for the characterisation through homomorphism counting. Lastly, we discuss the restriction to symmetric structures in section 5.

2 Preliminaries

3 Relational Colour Refinement

4 Relational Colour Refinement for structures with functions

4.1 Naive Encoding of functions

A simple way to apply relational colour refinement to non-relational structures is, to encode the functions as relations. Formally we transform a signature σ that includes function symbols to a new signature σ': For every relation symbol $R \in \sigma$, we introduce a relation symbol $R \in \sigma'$ with the same arity and for every function symbol $f \in \sigma$ with arity k, we introduce a relational symbol $R_f \in \sigma'$ of arity $k+1$.

Semantically, a structure $\mathfrak{A}$ of signature σ can then be encoded as a structure $\mathfrak{A}'$ of signature

σ' and with the same universe as $\mathfrak{A}$. For every relational symbol $R \in \sigma$ we set $R^{\mathfrak{A}'} := R^{\mathfrak{A}}$ and for every function symbol $f \in \sigma$ of arity k there exists a relation symbol $R_f \in \sigma'$ and we set $R_f^{\mathfrak{A}} := \{\mathbf{x}y : f^{\mathfrak{A}}(\mathbf{x}) = y\}$ where $\mathbf{x}$ is a tuple of arity k.

This procedure encodes a non-relational structure as a relational one, on which Relational Colour Refinement can now be performed. As such we say, that the Naive Relational Colour Refinement (nRCR) distinguishes two structures $\mathfrak{A}$ and $\mathfrak{B}$ if, and only if, RCR distinguishes their naive encodings $\mathfrak{A}'$ and $\mathfrak{B}'$. However, this results in a very weak logical characterisation, that does not allow nesting of terms, namely the nesting-free-fragment of $\mathsf{GF}(\mathsf{C})$.

Definition 1 ($\mathsf{nfGF}(\mathsf{C})$)**.** Consider the definition of $\mathsf{GF}(\mathsf{C})$ given in **??**. We obtain the nesting-free fragment, by allowing $f(\mathbf{x}) = y$ as a further atomic formula. Concretely, the only allowed atomic formulae are of the form $R(x_1, \ldots, x_\ell)$, $x = y$ and $f(x_1, \ldots, x_\ell) = y$, where f has arity ℓ, $\mathrm{free}(f(x_1, \ldots, x_\ell) = y) = \{x_1, \ldots, x_\ell\}$ and $\mathrm{gd}(f(\mathbf{x}) = y) = 0$.

The remaining definitions stay the same.

Theorem 2. *The two following statements are equivalent:*

1. *nRCR distinguishes* $\mathfrak{A}$ *and* $\mathfrak{B}$*.*
2. *There exists a sentence* $\varphi \in \mathsf{nfGF}(\mathsf{C})$ *such that* $\mathfrak{A} \models \varphi$ *and* $\mathfrak{B} \not\models \varphi$*.*

Proof. 1. $\Rightarrow$ 2.: By definition, $\mathfrak{A}$ and $\mathfrak{B}$ are distinguished by nRCR if, and only if, $\mathfrak{A}'$ and $\mathfrak{B}'$ are distinguished by RCR. Using the result of [16], we obtain a sentence $\varphi' \in \mathsf{GF}(\mathsf{C})$ that distinguishes the encoded structures. Via a structural induction on the formula, we can now translate φ' into a formula $\varphi \in \mathsf{nfGF}(\mathsf{C})$ This can be achieved by replacing formulae $R_f(x_1, \ldots, x_\ell, y)$ by $f(x_1, \ldots, x_\ell) = y$ for function symbols $f \in \sigma$ and letting everything else stay the same.

2. $\Rightarrow$ 1.: When considering $\mathsf{nfGF}(\mathsf{C})$, one can find that the transformation done at the end of the first direction can be applied in reverse. This then leads to a distinguishing sentence in $\mathsf{GF}(\mathsf{C})$ and with [16] to a distinguishing colouring of the encoded structures, which by definition is a distinguishing colouring for the structures themselves. $\square$

While the above theorem results in a nice characterisation of the naive encoding, the nesting of terms is often very desired when using functions. However, it can be shown that nesting is too powerful for the naive encoding.

Consider the two structures $\mathfrak{A}$ and $\mathfrak{B}$ of signature $\sigma = \{f/1\}$ which can be seen in Figure 1. Formally they are defined as $\mathfrak{A} = (A, f^{\mathfrak{A}})$ and $\mathfrak{B} = (B, f^{\mathfrak{B}})$ where

$$
\begin{aligned}
A =& \{a_1, a_2, a_3, a_4, a_5, a_6\}, \\
f^{\mathfrak{A}} =& \{a_1 \mapsto a_2, a_2 \mapsto a_3, a_3 \mapsto a_1, \\
& a_4 \mapsto a_5, a_5 \mapsto a_6, a_6 \mapsto a_4\}
\end{aligned}
\quad \text{and} \quad
\begin{aligned}
B =& \{b_1, b_2, b_3, b_4, b_5, b_6\}, \\
f^{\mathfrak{B}} =& \{b_1 \mapsto b_2, b_2 \mapsto b_3, b_3 \mapsto b_4, \\
& b_4 \mapsto b_5, b_5 \mapsto b_6, b_6 \mapsto b_1\}
\end{aligned}
$$

Consider the formula $\varphi = \exists^{\geq 1} x.(f(f(f(x))) = x)$ which utilizes term nesting to find a cycle of length three. It is obvious that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$. However, when encoding the two structures with the naive method described above, one finds that nRCR cannot distinguish them. Therefore, term nesting is too powerful for the naive encoding.

A method that allows for the nesting of terms will be described in the following section.

4.2 Using the transitive expansion

As a first remark we note that we only consider unary functions in this section. The key idea will be, to encode a function f as a family of relations, which then can capture the notion of nesting

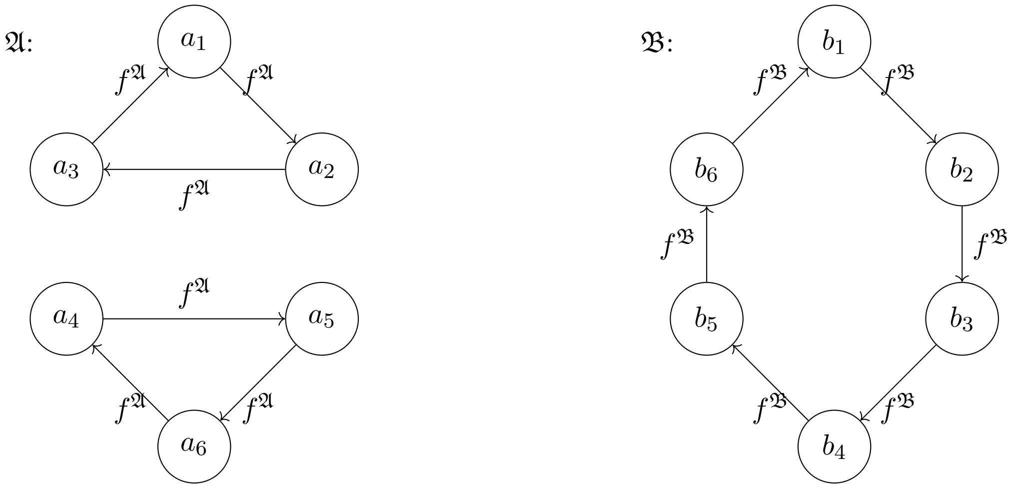

Figure 1: Two σ-structures $\mathfrak{A}$ and $\mathfrak{B}$ which can be distinguished by $\mathsf{GF}(\mathsf{C})$, but not by nRCR.

function applications. However, a bound on the alternation of different function symbols is necessary to ensure that the expanded signature is still finite, thus we will fixate a maximal alternation depth when discussing our new variant of RCR. Let us now concretely define, how we expand the signature.

Definition 3 (Transitive Expansion)**.** Let $\sigma := \sigma_{\text{Rel}} \,\dot\cup\, \sigma_{\text{Func}}$ be a signature with relation symbols σ_{Rel} and unary function symbols σ_{Func} and let $\mathfrak{A}$ be a structure of signature σ with $\|\mathfrak{A}\| = n$. For readability, we define the family of sets of alternations of function applications $\text{Alters}_n^0(\sigma) := \{\text{id}\}$ and

$$\begin{aligned}\text{Alters}_n^k(\sigma) := \text{Alters}_n^{k-1}(\sigma) \cup \{f_1^{m_1} f_2^{m_2} \dots f_k^{m_k} :\, & f_1 f_2 \dots f_k \in (\sigma_{\text{Func}})^k \\ & \wedge 0 < m_i \leq n \text{ for } i \in [k] \\ & \wedge \forall i \in \{1, \dots, k-1\}. f_{i-1} \neq f_i \neq f_{i+1}\}.\end{aligned}$$

We will now fixate an arbitrary $k \in \mathbb{N}$ which will be our bound on the alternation depth and will define a new signature $\widetilde{\sigma}$ as well as a structure $\widetilde{\mathfrak{A}}$ of said signature, which will be the transitive expansion with alternation depth k of $\mathfrak{A}$. For $k \in \mathbb{N}$, $\alpha, \beta, \alpha_1, \dots, \alpha_\ell \in \text{Alters}_n^k(\sigma)$ and a $R \in \sigma_{\text{Rel}}$ with arity ℓ, we define the binary relation

$$\text{Eq}_{\alpha,\beta}^{\widetilde{\mathfrak{A}}} := \{(a, b) : \alpha^{\mathfrak{A}}(a) = \beta^{\mathfrak{A}}(b)\},$$

and the relation of arity ℓ

$$R_{\alpha_1,\dots,\alpha_\ell}^{\widetilde{\mathfrak{A}}} := \{(a_1, \dots, a_\ell) : (\alpha_1^{\mathfrak{A}}(a_1), \dots, \alpha_\ell^{\mathfrak{A}}(a_\ell)) \in R^{\mathfrak{A}}\}.$$

We now define the transitive expansion with alternation depth k signature $\widetilde{\sigma}$, where

$$\begin{aligned}\widetilde{\sigma} :=& \{\text{Eq}_{\alpha,\beta} : \alpha, \beta \in \text{Alters}_n^k(\sigma)\}, \\ & \dot\cup \{R_{\alpha_1,\dots,\alpha_\ell} : R \in \sigma_{\text{Rel}}, \text{ar}(R) = \ell \text{ and } \alpha \in \text{Alters}_n^k(\sigma)\}.\end{aligned}$$

Since the following definitions will depend on this construction, let us consider an example. We define the signature $\sigma = \{R, f, g\}$ where R is a unary relation symbol and f and g are unary function symbols. Now consider a σ structure $\mathfrak{A} = (A, \sigma)$ with $A = \{a, b\}$, $R^{\mathfrak{A}} = \{a\}$, $f^{\mathfrak{A}} = \{a \mapsto b, b \mapsto a\}$ and $g^{\mathfrak{A}} = \{a \mapsto a, b \mapsto b\}$. A graphical representation of $\mathfrak{A}$ can be found in Figure 2. For the sake of simplicity we will define the transitive expansion with alternation depth 1 and because $\|\mathfrak{A}\| = 2$ we will use $\text{Alters}_2^1(\sigma)$ to do so. We see that $\text{Alters}_2^1(\sigma) = \{\text{id}, f, f^2, g, g^2\}$ and as such

$$\widetilde{\sigma} = \{R_{\text{id}}, R_f, R_{f^2}, R_g, R_{g^2}, \text{Eq}_{\text{id},\text{id}}, \text{Eq}_{\text{id},f}, \text{Eq}_{\text{id},f^2}, \dots, \text{Eq}_{g^2,g^2}\}.$$

Because of the relatively large size of $\widetilde{\sigma}$, we will only give the formal definitions for a few relations, while the rest of the relations in $\widetilde{\mathfrak{A}}$ can be seen in Figure 2. We find that $R_{\mathrm{id}}^{\widetilde{\mathfrak{A}}} = R_{f^2}^{\widetilde{\mathfrak{A}}} = R_{g}^{\widetilde{\mathfrak{A}}} = R_{g^2}^{\widetilde{\mathfrak{A}}} = \{a\}$ and that $R_{f}^{\widetilde{\mathfrak{A}}} = \{b\}$. Additionally, $\mathrm{Eq}_{g,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{g^2,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \{(a,a),(b,b)\} = \mathrm{Eq}_{\alpha,\alpha}^{\widetilde{\mathfrak{A}}}$ for all $\alpha \in \mathrm{Alters}_2^1(\sigma)$. To give another example, we have $\mathrm{Eq}_{g,f}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{g^2,f}^{\widetilde{\mathfrak{A}}} = \{(a,b),(b,a)\}$. The definitions of all $\mathrm{Eq}_{\alpha,\beta}^{\widetilde{\mathfrak{A}}}$ can be found in Figure 2.

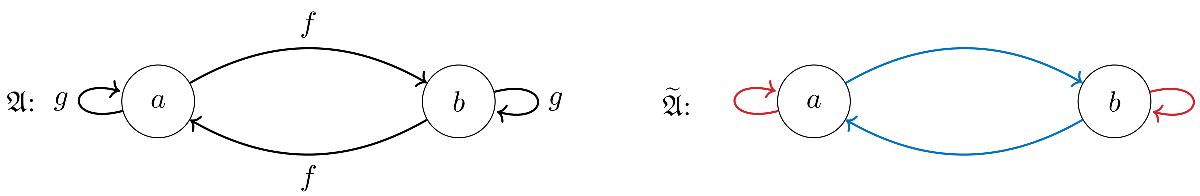

Figure 2: Graphical description of $\mathfrak{A}$ and $\widetilde{\mathfrak{A}}$. The blue transitions represent the relations $\mathrm{Eq}_{\alpha,\beta}$ with $(\alpha,\beta) \in \{(\mathrm{id},f),(f,\mathrm{id}),(f,f^2),(f,g),(f,g^2),(f^2,f),(g,f),(g^2,f)\}$, while the red transitions represent all other binary relations.

We can now define RCR for signatures that include unary function symbols.

Definition 4 (RCR for structures with unary functions)**.** Let σ be a signature with relation and unary function symbols and let $\mathfrak{A}$ and $\mathfrak{B}$ be structures of signature σ.

We say that $\mathfrak{A}$ and $\mathfrak{B}$ are being distinguished by RCR with alternation depth k (RCR_k), if $\|\mathfrak{A}\| \neq \|\mathfrak{B}\|$ or the transitive expansions with alternation depth k, $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$, are being distinguished by RCR.

To show that this definition may be sensible, we want to see, whether RCR_1 distinguishes the structures $\mathfrak{A}$ and $\mathfrak{B}$ from Figure 1. First we compute $\widetilde{\sigma}$ as $\{\mathrm{Eq}_{f^i,f^j}, \mathrm{Eq}_{f^i,\mathrm{id}}, \mathrm{Eq}_{\mathrm{id},f^j} : 0 \leq i,j \leq 6\} \cup \{\mathrm{Eq}_{\mathrm{id},\mathrm{id}}\}$. For easier readability, we will only give the definitions for the symbols in $\{\mathrm{Eq}_{f^i,\mathrm{id}} : 0 \leq i \leq n\}$. In fact, we find that

$$\mathrm{Eq}_{f^i,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \{(a_j, a_{j+i \mod 3}) : j \in [6]\}$$

and

$$\mathrm{Eq}_{f^i,\mathrm{id}}^{\widetilde{\mathfrak{B}}} = \{(a_j, a_{j+i \mod 6}) : j \in [6]\}.$$

By using [16], we know that RCR distinguishes $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$ if, and only if, there is a formula $\widetilde{\varphi} \in \mathsf{GF(C)}$ of signature $\widetilde{\sigma}$ that distinguishes them. Notice that $\mathrm{Eq}_{f^0,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{f^3,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{f^6,\mathrm{id}}^{\widetilde{\mathfrak{A}}}$, $\mathrm{Eq}_{f^1,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{f^4,\mathrm{id}}^{\widetilde{\mathfrak{A}}}$ and $\mathrm{Eq}_{f^2,\mathrm{id}}^{\widetilde{\mathfrak{A}}} = \mathrm{Eq}_{f^5,\mathrm{id}}^{\widetilde{\mathfrak{A}}}$, while only $\mathrm{Eq}_{f^0,\mathrm{id}}^{\widetilde{\mathfrak{B}}} = \mathrm{Eq}_{f^6,\mathrm{id}}^{\widetilde{\mathfrak{B}}}$. Therefore the sentence

$$\exists^{\geq 6}(x,y).\Big(\mathrm{Eq}_{f^1,\mathrm{id}}(x,y) \wedge \mathrm{Eq}_{f^4,\mathrm{id}}(x,y)\Big) \in \mathsf{GF(C)}$$

is satisfied by $\widetilde{\mathfrak{A}}$, but not $\widetilde{\mathfrak{B}}$. Furthermore, consider the formula $\varphi = \exists^{\geq 1}x.(f(f(f(x))) = x)$ that has been to used to distinguish $\mathfrak{A}$ and $\mathfrak{B}$. We can easily derive another formula $\varphi' \in \mathsf{GF(C)}$ to distinguish the transitive expansions, namely $\varphi' = \exists^{\geq 1}x.\,\mathrm{Eq}_{f^3,\mathrm{id}}(x,x)$.

We see that this procedure distinguishes structures that were not distinguished by nRCR. In the following, we want to investigate how much stronger this new algorithm is, by finding a logic that characterises it.

4.2.1 Logical characterisation of RCR_k

A first idea that may come to mind when looking at the definition of the transitive expansion, is to use the classical notion of atomic formula for guards, fixate a maximal alternation depth for terms and only allow $\|\mathfrak{A}\|$ applications of the same function symbol on series, that is, only allow $f^m(s(x))$ where $m < \|\mathfrak{A}\|$. However, we prove that we can allow any $f^m(s(x))$, while the bounded alternation depth is still needed. The reason why this is possible, hinges on the pigeonhole principle. When considering $f(x)$, $f^2(x)$, $f^3(x)$ and so forth, until $f^m(x)$, where $m > \|\mathfrak{A}\|$, there have to be to numbers i and j, such that $f^i(x) = f^j(x)$. Therefore, we can decompose the path into a path to a cycle, the cycle itself, and a last part of that cycle. To allow the following proofs to be more readable, we first want to define the set of all such valid decompositions.

Let

$$\begin{aligned}\mathcal{I}(n,m) = \{(k,l,p) \in [n]^3 \quad : \quad & k+p < k+l \leq n \land \\ & k + r \cdot l + p = m \text{ for some } r \in \mathbb{N}\}.\end{aligned}$$

This set will represents all the possible ways, to decompose a path into a cycle and the path to and from it. This means, that the triple (k,ℓ,p) will represent a path, that has a beginning part of length k, then a cycle of length ℓ and a last part that consists of the first p elements of the cycle. One can see that in a structure $\mathfrak{A}$ with a unary function f and n elements, any path along of f with length $m > n$ can be decomposed into a triple in the set $\mathcal{I}(n,m)$. A graphical description of such a triple (k,ℓ,p) can be found in Figure 3.

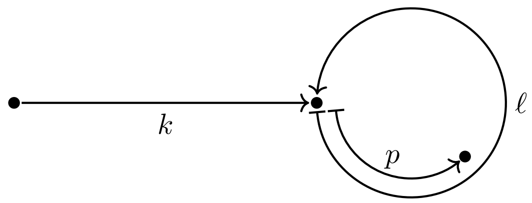

Figure 3: A description of how a path can be decomposed into a cycle, the path to it and a last part of it.

In the beginning we remarked that we have to fixate an alternation depth. This bound can be seen in the definition of the transitive expansion and will be used in the logic that will characterise the Colour Refinement algorithm. Therefore we can only reason about a fragment of $\mathsf{GF(C)}$, where the terms do not alternate too often. This is formally stated in the following definition.

Definition 5 (Alternation bounded $\mathsf{GF(C)}$)**.** The fragment of $\mathsf{GF(C)}$ with an bounded alternation depth of k ($\mathsf{GF(C)}_k$) is $\mathsf{GF(C)}$ with the constraint that for all formulae $\varphi \in \mathsf{GF(C)}_k$ of signature σ and every term t that appears in φ, there is an $n \in \mathbb{N}$ and an $\alpha \in \mathrm{Alters}_n^k(\sigma)$ such that $\alpha = t$. Atomic formulae are defined as usual, that is, the formulae $R(t_1(x_1), t_2(x_2), \dots, t_n(x_n))$ and $t_1(x_1) = t_2(x_2)$ for terms $t_1, t_2, \dots, t_n$ and variables $x_1, x_2, \dots, x_n$ are atomic formulae.

With this, we can prove the first result, which allows us to use every $f^m(x) = y$ in a formula.

Lemma 6. *Let $\psi(x_1,x_2) \in \mathsf{GF(C)}_1$ be of the form $f^m(x_1) = x_2$. Then there exists a formula $\vartheta(x_1,x_2) \in \mathsf{GF(C)}$ such that for any $\mathfrak{A}$ with $\|\mathfrak{A}\| = n$ it holds*

$$\mathfrak{A}, a_1, a_2 \models \psi(x_1,x_2) \textit{ if, and only if, } \mathfrak{A}, a_1, a_2 \models \vartheta(x_1,x_2)$$

and for any $f^{m'}(x)$ that appears in ϑ we have $m' \leq n$. Furthermore, $\vartheta(x_1,x_2)$ is of the form $\bigvee \Phi(x_1,x_2)$, and if $\mathfrak{A}, a_1, a_2 \models \vartheta(x_1,x_2)$, then there is exactly one $\varphi(x_1,x_2) \in \Phi$, such that $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi(x_1,x_2)$. Additionally, $\vartheta(x_1,x_2) \in \mathsf{GF(C)}_1$.

Proof. If $m \leq n$, we let $\vartheta \coloneqq \psi$ and the claim follows.

Otherwise, we define

$$\vartheta(x_1, x_2) \coloneqq \bigvee_{(k,\ell,p)\in\mathcal{I}(n,m)} \zeta_{(k,\ell,p)}(x_1, x_2)$$

where

$$\begin{aligned}\zeta_{(k,\ell,p)}(x_1, x_2) \coloneqq & f^{k+p}(x_1) = x_2 \wedge f^k(x_1) = f^{k+\ell}(x_1)\\ & \wedge \mathrm{E}_f^{k,\ell}(x_1)\\ & \wedge \bigwedge_{\ell'<\ell} f^k(x_1) \neq f^{k+\ell'}(x_1)\end{aligned}$$

and for some term $t(x_1)$ we have

$$\mathrm{E}_f^{k,\ell}(t(x_1)) = \begin{cases}\top & \text{if } k = 0\\ f^{k-1}(t(x_1)) \neq f^{k-1+\ell}(t(x_1)) & \text{otherwise.}\end{cases}$$

Due to the definition of $\mathcal{I}(n,m)$ it is obvious that only $f^{m'}$ with $m' \leq n$ appears. We now proceed to the proof of the equivalence. For the purpose of readability, we will write $f_{\mathfrak{A}}$ instead of $f^{\mathfrak{A}}$.

We will show that if $\mathfrak{A}, a_1, a_2 \models \vartheta(x_1, x_2)$, then $\mathfrak{A}, a_1, a_2 \models \psi(x_1, x_2)$. Let $\mathfrak{A}, a_1, a_2 \models \vartheta(x_1, x_2)$. By definition of ϑ, there are $(k, \ell, p) \in \mathcal{I}(n,m)$ with $\mathfrak{A}, a_1, a_2 \models \zeta_{(k,\ell,p)}(x_1, x_2)$. In particular $f_{\mathfrak{A}}^k(a_1) = f_{\mathfrak{A}}^{k+\ell}(a_1)$. It follows that

$$f_{\mathfrak{A}}^k(a_1) = f_{\mathfrak{A}}^{k+\ell}(a_1) = f_{\mathfrak{A}}^{k+2\ell}(a_1) = f_{\mathfrak{A}}^{k+3\ell}(a_1) = \cdots = f_{\mathfrak{A}}^{k+r\cdot\ell}(a_1)$$

for all $r \in \mathbb{N}$. By using the definition of $\mathcal{I}(n,m)$, we get

$$a_2 = f_{\mathfrak{A}}^{k+p}(a_1) = f_{\mathfrak{A}}^{k+r\cdot\ell+p}(a_1) = f_{\mathfrak{A}}^m(a_1).$$

From this we can deduce $\mathfrak{A}, a_1, a_2 \models \psi(x_1, x_2)$, where $\psi(x_1, x_2)$ has the form $f^m(x_1) = x_2$.

Now we prove that if $\mathfrak{A}, a_1, a_2 \models \psi(x_1, x_2)$, then $\mathfrak{A}, a_1, a_2 \models \vartheta(x_1, x_2)$. Let $\mathfrak{A}, a_1, a_2 \models \psi(x_1, x_2)$. By assumption $m > n$ and by the pigeonhole principle there have to be distinct i and j such that $f_{\mathfrak{A}}^i(a_1) = f_{\mathfrak{A}}^j(a_1)$. Choose such i, j such that they are lexicographically minimal. Now choose $k \coloneqq i$, $\ell \coloneqq j - i$ and $p \coloneqq (m-i) \mod (j-i) = (m-i) \mod \ell$. Obviously $(k, \ell, p) \in \mathcal{I}(n,m)$ and what remains to be shown is that $\mathfrak{A}, a_1, a_2 \models \zeta_{(k,\ell,p)}(x_1, x_2)$. For that, we consider the parts of the conjunction and show for each one that it is satisfied.

- $f^{k+p}(x_1) = x_2$ is satisfied. We use the fact that $a = b \mod c \Leftrightarrow b = r \cdot c + a$ for some $r \in \mathbb{N}$. Then
$$f_{\mathfrak{A}}^{k+p}(a_1) = f_{\mathfrak{A}}^{i+(m-i)-r\cdot\ell}(a_1) = f_{\mathfrak{A}}^{i+r\cdot\ell+m-i-r\cdot\ell}(a_1) = f_{\mathfrak{A}}^m(a_1) = a_2.$$
Therefore $\mathfrak{A}, a_1, a_2 \models f^{k+p}(x_1) = x_2$.

- $f^k(x_1) = f^{k+\ell}(x_1)$ is satisfied. Consider that
$$f_{\mathfrak{A}}^k(a_1) = f_{\mathfrak{A}}^i(a_1) = f_{\mathfrak{A}}^j(a_1) = f_{\mathfrak{A}}^{j+i-i}(a_1) = f_{\mathfrak{A}}^{i+j-i}(a_1) = f_{\mathfrak{A}}^{k+\ell}(a_1).$$
This leads to $\mathfrak{A}, a_1, a_2 \models f^k(x_1) = f^{k+\ell}(x_1)$.

- $\mathrm{E}_f^{k,\ell}(x_1)$ is satisfied. Otherwise $f_{\mathfrak{A}}^{k-1}(a_1) = f_{\mathfrak{A}}^{k-1+\ell}(a)$, but then $(k-1, \ell)$ would be lexicographically smaller than (i, j).

- The same reasoning applies to $\bigwedge_{\ell'<\ell} f^k(x_1) \neq f^{k+\ell'}(x_1)$. If it weren't satisfied, there would be a (i, j') with $j' < j$ and $f_{\mathfrak{A}}^i(a_1) = f_{\mathfrak{A}}^{i+j'}(a_1)$ which would be lexicographically smaller than (i, j).

Thus we have shown that every subformula of the conjunction and therefore the formula is being fulfilled.

Lastly, it remains to prove that if ϑ is satisfied, then there is exactly one $(k,\ell,p) \in \mathcal{I}(n,m)$ such that $\exists^{\geq 1} x_2.\zeta_{(k,\ell,p)}(x_1,x_2)$ is fulfilled. We prove this by contradiction. Assume that $\mathfrak{A}, a_1, a_2 \models \vartheta(x_1,x_2)$ and that there are $\zeta_{(k,\ell,p)}(x_1,x_2)$ and $\zeta_{(k',\ell',p')}(x_1,x_2)$ with $(k,\ell,p) \neq (k',\ell',p')$, such that $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\zeta_{(k,\ell,p)}(x_1,x_2)$ and $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\zeta_{(k',\ell',p')}(x_1,x_2)$.

We proceed with a case distinction. Let $k = k'$ and $\ell = \ell'$. Then there are $r, r' \in \mathbb{N}$ such that

$$k + r \cdot \ell + p = k' + r' \cdot \ell' + p' = m.$$

Thus we can infer that $r \cdot \ell + p = r' \cdot \ell' + p'$. By definition of $\mathcal{I}(n,m)$ we know that $p, p' < \ell = \ell'$ and as such

$$r \cdot \ell + p, r' \cdot \ell' + p' \in \{r \cdot \ell, r \cdot \ell + 1, \ldots, r \cdot \ell + (\ell - 1)\}$$

and because p is a non-negative integer, $r = r'$ has to follow and further we get $p = p'$. However this would contradict that $(k,\ell,p) \neq (k',\ell',p')$. Now assume that $\ell \neq \ell'$ and without loss of generality assume that $\ell < \ell'$. But then $\mathfrak{A}, a_1 \not\models \bigwedge_{\hat{\ell}<\ell'} f^{k'}(x_1) \neq f^{k'+\hat{\ell}}(x_1)$, because

$$f_{\mathfrak{A}}^{k'+\ell}(a_1) = f_{\mathfrak{A}}^{k+\ell} = f_{\mathfrak{A}}^{k}(a_1) = f_{\mathfrak{A}}^{k'}(a_1)$$

and $k' + \ell < k' + \ell'$. Thus this cannot be the case as well.

Consider that $k \neq k'$ and without loss of generality assume that $k < k'$. If $\ell = \ell'$, then by the principle of induction, we get that $f_{\mathfrak{A}}^{k}(a_1) = f_{\mathfrak{A}}^{k+\ell}(a_1)$, $f_{\mathfrak{A}}^{k+1}(a_1) = f_{\mathfrak{A}}^{k+1+\ell}(a_1)$ and then $f_{\mathfrak{A}}^{k'}(a_1) = f_{\mathfrak{A}}^{k'+\ell'}(a_1)$. But this contradicts $\mathfrak{A}, a_1 \models \mathrm{E}_f^{k',\ell'}(x_1)$. If $\ell < \ell'$, then

$$f_{\mathfrak{A}}^{k'}(a_1) = f_{\mathfrak{A}}^{k+(k'-k)}(a_1) = f_{\mathfrak{A}}^{k+(k'-k)+\ell}(a_1) = f_{\mathfrak{A}}^{k'+\ell}(a_1),$$

but this again contradicts $\mathfrak{A}, a_1 \models \bigwedge_{\hat{\ell}<\ell'} f^{k'}(x_1) \neq f^{k'+\hat{\ell}}(x_1)$. If $\ell' < \ell$, then there exists a $t \in \mathbb{N}$ such that

$$k + t \cdot \ell < k' \leq k + (t+1) \cdot \ell.$$

We now define $r := k + (t+1) \cdot \ell - k'$ and get $f_{\mathfrak{A}}^{k'+r}(a_1) = f_{\mathfrak{A}}^{k'+r+\ell'}$ and by using $f_{\mathfrak{A}}^{k'+r}(a_1) = f_{\mathfrak{A}}^{k+(t+1)\cdot\ell}(a_1) = f_{\mathfrak{A}}^{k}(a_1)$ it follows that $f_{\mathfrak{A}}^{k}(a_1) = f_{\mathfrak{A}}^{k+\ell'}(a_1)$. This contradicts $\mathfrak{A}, a_1 \models \bigwedge_{\hat{\ell}<\ell} f^{k}(x_1) \neq f^{k'+\hat{\ell}}(x_1)$.

One can see that we did not use x_2 or a_2. Therefore its interpretation is irrelevant, which is why we can existentially quantify it in the claim. As all possible cases lead to a contradiction, the first assumption cannot be true and we proved the claim.

As we did not use any function symbols other than f, $\vartheta(x_1,x_2) \in \mathsf{GF}(\mathsf{C})_1$ follows obviously. $\square$

The above proof allows for the translation of a formula $f^m(x) = y$ to a formula $\vartheta(x,y)$ that is equivalent for structures with n elements. A natural extension would be, to allow alternation of functions, for example formulae like $g^m(f^{m'}(x)) = y$. This is also possible and will be proved in the following.

Lemma 7. *Let $d \in$ and $\psi(x_1,x_2) \in \mathsf{GF}(\mathsf{C})_d$ be of the form $t(x_1) = x_2$ for a term t. Then there exists a formula $\vartheta_t(x_1,x_2) \in \mathsf{GF}(\mathsf{C})_d$, such that for any structure $\mathfrak{A}$ with $\|\mathfrak{A}\| = n$ it holds*

$$\mathfrak{A}, a_1, a_2 \models \psi(x_1,x_2) \textit{ if, and only if, } \mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1,x_2).$$

Furthermore, $\vartheta_t(x_1,x_2)$ is of the form $\bigvee \Phi(x_1,x_2)$ where all $\varphi(x_1,x_2) \in \Phi(x_1,x_2)$ are of the form

$$t'(x_1) = x_2 \wedge \bigwedge \Psi(x_1)$$

for some term $t'(x_1)$, and for every function symbol f in the signature, there does not appear a term of the form $f^m(s(x))$ where $m > n$. Additionally, if $\mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1,x_2)$, then there is exactly one $\varphi \in \Phi$, such that $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi(x_1,x_2)$.

Proof. We prove this via an induction on the term $t(x_1)$.

Base case: If $t(x_1)$ is of the form $f^m(x_1)$ for a unary function symbol f and $m \in \mathbb{N}$, we use the formula constructed in the proof of Theorem 6. It can easily be verified that it is in the correct form and from the same proof we get that if the translated formula is fulfilled, exactly one subformula of the disjunction is satisfied.

Inductive step: Assume that $t(x_1)$ is of the form $g^m(s(x_1))$ for a unary function symbol g, $m \in \mathbb{N}$ and term s. By the induction hypothesis, there is a formula $\vartheta_s(x_1, x_2) \in \mathsf{GF}(\mathsf{C})_{d-1}$ of the form $\bigvee \Phi_s(x_1, x_2)$ defined above with $\mathfrak{A}, a_1, a_2 \models s(x_1) = x_2$ if, and only if, $\mathfrak{A}, a_1, a_2 \models \vartheta_s(x_1, x_2)$.

If $m \leq n$, we set $\vartheta_t(x_1, x_2)$ to

$$\bigvee \Phi'(x_1, x_2),$$

where $\Phi'(x_1, x_2) \coloneqq \{g^m(t'(x_1)) = x_2 \land \bigwedge \Psi(x_1) : t'(x_1) = x_2 \land \bigwedge \Psi(x_1) \in \Phi_s(x_1, x_2)\}$.

If $m > n$, then we set $\vartheta_t(x_1, x_2)$ to

$$\bigvee_{(k,\ell,p) \in \mathcal{I}(n,m)} \bigvee \Phi'_{(k,\ell,p)}(x_1, x_2),$$

where

$$\begin{aligned}\Phi'_{(k,\ell,p)} \coloneqq \{g^{k+p}(t'(x_1)) = x_2 \land g^k(t'(x_1)) = g^{k+l}(t'(x_1)) \\ \land \mathrm{E}_g^{k,l}(t'(x_1)) \land \bigwedge_{\ell' < \ell} g^k(t'(x_1)) \neq g^{k+\ell'}(t'(x_1)) \\ \land \Psi(x_1) : t'(x_1) = x_2 \land \bigwedge \Psi(x_1) \in \Phi_s(x_1, x_2)\}\end{aligned}$$

By using the above definitions, we get $\mathfrak{A}, a_1, a_2 \models s(x_1) = x_2$ if, and only if, $\mathfrak{A}, a_1, a_2 \models \varphi_s(x_1, x_2)$ for some $\varphi_s \in \Phi_s$ where $\varphi_s(x_1, x_2)$ is of the form $t'(x_1) = x_2 \land \bigwedge \Psi(x_1)$. Therefore

$$\mathfrak{A}, a_1, a_2 \models s(x_1) = x_2 \text{ if, and only if, } \mathfrak{A}, a_1, a_2 \models t'(x_1) = x_2 \land \bigwedge \Psi(x_1). \tag{1}$$

We now prove that

$$\mathfrak{A}, a_1, a_2 \models t(x_1) = x_2 \text{ if, and only if, } \mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1, x_2).$$

Assume $m \leq n$. Let $\mathfrak{A}, a_1, a_2 \models \vartheta_t$. Then there is some $\varphi(x_1, x_2)$ of the form $g^m(t'(x_1)) = x_2 \land \bigwedge \Psi(x_1)$ such that $\mathfrak{A}, a_1, a_2 \models \varphi(x_1, x_2)$. We then get

$$\begin{aligned} &\mathfrak{A}, a_1, a_2 \models g^m(t'(x_1)) = x_2 \land \bigwedge \Psi(x_1) \\ \Leftrightarrow\ &\mathfrak{A}, a_1, a_2, a_3 \models g^m(x_3) = x_2 \land \bigwedge \Psi(x_1) \land t'(x_1) = x_3 \text{ for some } a_3 \in A \\ \overset{(1)}{\Leftrightarrow}\ &\mathfrak{A}, a_1, a_2, a_3 \models g^m(x_3) = x_2 \land s(x_1) = x_3 \text{ for some } a_3 \in A \\ \Leftrightarrow\ &\mathfrak{A}, a_1, a_2 \models g^m(s(x_1)) = x_2. \end{aligned}$$

Now let $m > n$. Then there is a

$$\begin{aligned}\varphi(x_1, x_2) \coloneqq g^{k+p}(t'(x_1)) = x_2 \land g^k(t'(x_1)) = g^{k+l}(t'(x_1)) \\ \land \mathrm{E}_g^{k,l}(t'(x_1)) \land \bigwedge_{\ell' < \ell} g^k(t'(x_1)) \neq g^{k+\ell'}(t'(x_1)) \\ \land \bigwedge \Psi(x_1)\end{aligned}$$

for some $(k, \ell, p) \in \mathcal{I}(n, m)$ with $\mathfrak{A}, a_1, a_2 \models \varphi(x_1, x_2)$. And now

$$\begin{aligned}
\mathfrak{A}, a_1, a_2 \models& \varphi(x_1, x_2) \\
\Leftrightarrow A, a_1, a_2, a_3 \models& g^{k+p}(x_3) = x_2 \wedge g^k(x_3) = g^{k+l}(x_3) \\
& \wedge \mathrm{E}_g^{k,l}(x_3) \wedge \bigwedge_{\ell' < \ell} g^k(x_3) \neq g^{k+\ell'}(x_3) \\
& \wedge \bigwedge \Psi(x_1) \wedge t'(x_1) = x_3 \text{ for some } a_3 \in A \\
\overset{\textit{Theorem 6}}{\Leftrightarrow} \mathfrak{A}, a_1, a_2, a_3 \models& g^m(x_3) = x_2 \wedge t'(x_1) = x_3 \wedge \bigwedge \Psi(x_1) \text{ for some } a_3 \in A \\
\overset{\textit{Equation (1)}}{\Leftrightarrow} \mathfrak{A}, a_1, a_2, a_3 \models& g^m(x_3) = x_2 \wedge s(x_1) = x_3 \text{ for some } a_3 \in A \\
\Leftrightarrow \mathfrak{A}, a_1, a_2 \models& g^m(s(x_1)) = x_2.
\end{aligned}$$

The other direction follows in both cases, as only equivalent steps have been used and it is obvious that the disjunction of a set is being fulfilled, if a formula of the set is satisfied.

Lastly, we show that if $\mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1, x_2)$, where ϑ_t is of the form $\bigvee \Phi$, there is exactly one $\varphi \in \Phi$, such that $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi(x_1, x_2)$. As in the proof of Theorem 6, we are going to use a proof by contradiction and we will look at the cases where $m \leq n$ and $m > n$ separately. If $m \leq n$, assume that $\mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1, x_2)$ and that there are $\varphi_1, \varphi_2 \in \Phi'(x_1, x_2)$ with $\varphi_1 \neq \varphi_2$, $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi_1(x_1, x_2)$ and $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi_2(x_1, x_2)$. It is easy to see that

$$\mathfrak{A}, a_1, a_2 \models g^m(t_1'(x_1)) = x_2 \wedge \bigwedge \Psi_1(x_1) \wedge g^m(t_2'(x_1)) = x_2 \wedge \bigwedge \Psi_2(x_1)$$

for some a_2, which is equivalent to

$$\mathfrak{A}, a_1, a_2, a_3, a_4 \models g^m(x_3) = x_2 \wedge t_1'(x_1) = x_3 \wedge \Psi_1(x_1) \wedge g^m(x_4) = x_2 \wedge t_2'(x_1) = x_4 \wedge \Psi_2(x_1)$$

when using the correct a_3 and a_4. However, $t_1'(x_1) = x_2 \wedge \Psi_1(x_1), t_2'(x_1) = x_s \wedge \Psi_2(x_1) \in \Phi_s$ and thus there would be $\psi_1(x_1, x_3/x_2), \psi_2(x_1, x_4/x_2) \in \Phi_s$, such that $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_3.\psi(x_1, x_3)$ and $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_4.\psi(x_1, x_4)$. This is a contradiction to the induction hypothesis.

If $m > n$, we again assume that $\mathfrak{A}, a_1, a_2 \models \vartheta_t(x_1, x_2)$ and that there are $\varphi_1(x_1, x_2) \in \Phi'_{(k,\ell,p)}(x_1, x_2)$ and $\varphi_2(x_1, x_2) \in \Phi'_{(k',\ell',p')}(x_1, x_2)$ with $\varphi_1 \neq \varphi_2$, $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi_1(x_1, x_2)$ and $\mathfrak{A}, a_1 \models \exists^{\geq 1} x_2.\varphi_2(x_1, x_2)$. By looking at the structure of the formulae as they are defined in this proof and by substituting terms and variables like in the first case, we again find that

$$\mathfrak{A}, a_1, a_3, a_4 \models t_1'(x_1) = x_3 \wedge \Psi_1(x_1) \wedge t_2'(x_1) = x_4 \wedge \Psi_2(x_1),$$

where $t_1'(x_1) = x_2 \wedge \bigwedge \Psi_1(x_1), t_2'(x_1) = x_2 \wedge \bigwedge \Psi_2(x_1) \in \Phi_2$. By using the same arguments as before, we as well arrive at a contradiction. As such, the assumption must be false and we have finished the proof. $\square$

A corollary of the above lemma is that the same statement also holds for an arbitrary relation, in addition to equality.

Lemma 8. *Let $d \in \mathbb{N}$ and $\psi(x_1, \ldots, x_m) \coloneqq R(t_1(x_1), \ldots, t_m(x_m)) \in \mathsf{GF}(\mathsf{C})_d$ be an atomic formula. Then there exists a formula $\vartheta_\psi \in \mathsf{GF}(\mathsf{C})_d$, such that for any given structure (of fitting signature) $\mathfrak{A}$ with $\|\mathfrak{A}\| = n$ it holds*

$$\mathfrak{A}, a_1, \ldots, a_m \models \psi(x_1, \ldots, x_m) \textit{ if, and only if, } \mathfrak{A}, a_1, \ldots, a_m \models \vartheta_\psi(x_1, \ldots, x_m).$$

Furthermore, $\vartheta_\psi(x_1, \ldots, x_m)$ is of the form $\bigvee \Phi(x_1, \ldots, x_m)$ where all $\varphi \in \Phi$ are of the form

$$R(t_1'(x_1), \ldots, t_m'(x_m)) \wedge \bigwedge \Psi_1(x_1) \wedge \cdots \wedge \bigwedge \Psi_m(x_m),$$

and for every $f^m(s(x))$ that appear in ϑ_ψ, where f is a unary function symbol and s is a term, $m \leq n$. Additionally, if $\mathfrak{A}, a_1, \ldots, a_, \models \vartheta_\psi(x_1, \ldots, x_m)$, then there exists exactly one $\varphi(x_1, \ldots, x_m) \in \Phi(x_1, \ldots, x_m)$, such that $\mathfrak{A}, a_1, \ldots, a_m \models \varphi(x_1, \ldots, x_m)$.

Proof. Let $\mathfrak{A}, a_1, \ldots, a_m \models \psi(x_1, \ldots, x_m)$. This is equivalent to

$$\mathfrak{A}, a_1, \ldots, a_m, b_1, \ldots, b_m \models R(b_1, \ldots, b_m) \wedge t_1(x_1) = b_1 \wedge \cdots \wedge t_m(x_m) = b_m$$

for some $b_1, \ldots, b_m \in A$. By applying the previous lemma, we get the equivalent statement

$$\begin{aligned}\mathfrak{A}, a_1, \ldots, a_m, b_1, \ldots, b_m \models R(y_1, \ldots, y_m) \wedge & \bigvee_{i_1} \left(t'_{1,i_1}(x_1) = y_1 \wedge \bigwedge \Psi_{1,i_1}(x_1)\right) \\ & \wedge \ldots \\ & \wedge \bigvee_{i_m} \left(t'_{m,i_m}(x_m) = y_m \wedge \bigwedge \Psi_{m,i_m}(x_m)\right).\end{aligned}$$

Through distribution of boolean formulae we get

$$\begin{aligned}\mathfrak{A}, a_1, \ldots, a_m, b_1, \ldots, b_m \models \bigvee_{i_1} \cdots \bigvee_{i_m} (R(y_1, \ldots, y_m) \wedge & t'_{1,i_1}(x_1) = y_1 \wedge \bigwedge \Psi_{1,i_1}(x_1) \\ & \wedge \ldots \\ & \wedge t'_{m,i_m}(x_m) = y_m \wedge \bigwedge \Psi_{m,i_m}(x_m)).\end{aligned} \tag{2}$$

Finally, we can resubstitute variables and get

$$\begin{aligned}\mathfrak{A}, a_1, \ldots, a_m \models \bigvee_{i_1} \cdots \bigvee_{i_m} (& R(t'_{1,i_1}(x_1), \ldots, t'_{m,i_m}(x_m)) \\ & \wedge \bigwedge \Psi_{1,i_1}(x_1) \\ & \wedge \ldots \\ & \wedge \bigwedge \Psi_{m,i_m}(x_m)) =: \vartheta_\psi(x_1, \ldots, x_m).\end{aligned}$$

One can see that ϑ_ψ is of the correct form. The equality follows from the fact that only equivalences have been used to derive ϑ_ψ from ψ.

Lastly, we prove that if ϑ_ψ is satisfied, there is exactly one formula of the disjunction that is satisfied. For this, consider the equivalent formula from Equation (2). Assume that $\mathfrak{A}, a_1, \ldots, a_m \models \vartheta_\psi$ and that there are two subformulae φ_1 and φ_2 of the formula in Equation (2), where φ_1 is of the form

$$\begin{aligned}R(y_1, \ldots, y_m) \wedge & t'_{1,i_1}(x_1) = y_1 \wedge \bigwedge \Psi_{1,i_1}(x_1) \\ & \wedge \ldots \\ & \wedge t'_{m,i_m}(x_m) = y_m \wedge \bigwedge \Psi_{m,i_m}(x_m)\end{aligned}$$

and φ_2 is of the form

$$\begin{aligned}R(y_1, \ldots, y_m) \wedge & s'_{1,i_1}(x_1) = y_1 \wedge \bigwedge \Psi'_{1,i_1}(x_1) \\ & \wedge \ldots \\ & \wedge s'_{m,i_m}(x_m) = y_m \wedge \bigwedge \Psi'_{m,i_m}(x_m),\end{aligned}$$

such that $\varphi_1 \neq \varphi_2$, $\mathfrak{A}, a_1, \ldots, a_m, b_1, \ldots, b_m \models \varphi_1$ and $\mathfrak{A}, a_1, \ldots, a_m, b_1, \ldots, b_m \models \varphi_2$. As $\varphi_1 \neq \varphi_2$, there must be a j such that ψ_1 is of the form $t'_{j,i_j}(x_j) = y_j \wedge \bigwedge \Psi_{j,i_j}(x_j)$, ψ_2 is of the form $s'_{j,i_j}(x_j) = y_j \wedge \bigwedge \Psi'_{j,i_j}(x_j)$ and $\psi_1 \neq \psi_2$. From the construction of the formula we know, that there is a term t_j, a formula ϑ_{t_j} of the form $\bigvee \Phi_{t_j}$ and $\psi_1, \psi_2 \in \Phi_{t_j}$. However, $\mathfrak{A}, a_j \models \exists^{\geq 1} y_j.\psi_1(x_j, y_j)$ and $\mathfrak{A}, a_j \models \exists^{\geq 1} y_j.\psi_2(x_j, y_j)$ would contradict the claim that has been proved in Theorem 7. $\square$

To illustrate how this translation works, let us consider the formula ψ of the form $g^4(f^3(x)) = y$ for a structure with 2 elements. As in the proof, we inductively translate the inner terms and as such get for the formula $f^3(x) = y$, the formula φ of the form

$$\bigvee_{(k,\ell,p)\in\mathcal{I}(2,3)} \left(f^k(x) = f^{k+\ell}(x) \wedge f^{k+p}(x) = y \wedge \mathrm{E}_f^{k,l}(x) \wedge \bigwedge_{\hat{\ell}<\ell} f^k(x) \neq f^{k+\hat{\ell}}(x) \right)$$

and with $\mathcal{I}(2,3) = \{(0,2,1),(1,1,0),(0,1,0)\}$ we get that φ equals

$$\begin{aligned} &\big(x = f^2(x) \wedge f(x) = y \wedge x \neq f(x)\big) \\ \vee &\big(f(x) = f^2(x) \wedge f(x) = y \wedge x \neq f(x)\big) \\ \vee &\,(x = f(x) \wedge x = y \wedge \top)\,. \end{aligned}$$

Now we can construct ϑ_ψ from ψ. From the proof, we know that ϑ_ψ is of the form

$$\begin{aligned} \bigvee_{(k',\ell',p')\in\mathcal{I}(2,4)} \ \bigvee_{(k,\ell,p)\in\mathcal{I}(2,3)} &(f^k(x) = f^{k+\ell}(x) \wedge \mathrm{E}_f^{k,l}(x) \wedge \bigwedge_{\hat{\ell}<\ell} f^k(x) \neq f^{k+\hat{\ell}}(x) \\ &\wedge g^{k'}(f^{k+p}(x)) = g^{k'+\ell'}(f^{k+p}(x)) \wedge g^{k'+p'}(f^{k+p}(x)) = y \\ &\wedge E_g^{k',\ell'}(f^{k+p}(x)) \wedge \bigwedge_{\hat{\ell}<\ell'} g^{k'}(f^{k+p}(x)) \neq g^{k'+\hat{\ell}}(f^{k+p}(x)) \end{aligned}$$

and with $\mathcal{I}(2,4) = \{(1,1,0),(0,1,0),(0,2,0)\}$ we can analogous find that ϑ_ψ is equal to

Das ist ja eine Disjunktion mit 3*3=9 Formeln die jeweils etwa eine Zeile lang sind. Sollte ich die trotzdem komplett aufschreiben

This now allows us to proof the logical characterisation of our Colour Refinement Algorithm.

Theorem 9. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two structures of the same signature σ with relation and unary function symbols and let $k \in \mathbb{N}$ The two following statements are equivalent:*

1. RCR_k *distinguishes $\mathfrak{A}$ and $\mathfrak{B}$.*
2. *There exists a sentence $\varphi \in \mathsf{GF}(\mathsf{C})_k$ such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$.*

Proof. We prove that *1.* implies *2.*. Let $\mathfrak{A}$ and $\mathfrak{B}$ be distinguished by RCR_k. If they are of different sizes, assume without loss of generality that

$$\|\mathfrak{A}\| = n > n' = \|\mathfrak{B}\|.$$

Then define $\varphi := \exists^{\geq n} x.\top \in \mathsf{GF}(\mathsf{C})_k$, which obviously distinguishes the structures.

Now assume $\|\mathfrak{A}\| = \|\mathfrak{B}\| = n$. By definition, RCR distinguishes $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$. When using the proof from [16], we obtain a formula $\widetilde{\varphi} \in \mathsf{GF}(\mathsf{C})$ of signature $\widetilde{\sigma}$ that distinguishes the expansions. This formula $\widetilde{\varphi}$ can then be translated to a formula $\varphi \in \mathsf{GF}(\mathsf{C})_k$ of signature σ. For every atomic subformula $\mathrm{Eq}_{\alpha,\beta}(x,y)$, where $\alpha, \beta \in \mathrm{Alters}_n^k(\sigma)$, replace it by the formula $\alpha(x) = \beta(y)$, and every atomic subformula $R_{\alpha_1,\ldots,\alpha_\ell}(x_1,\ldots,x_\ell)$, replace it by the formula $R(\alpha_1(x_1),\ldots,\alpha_\ell(x_\ell))$. Obviously, if a structure's expansion satisfied $\widetilde{\varphi}$, it also satisfies φ and vice versa. Therefore, we get a formula $\varphi \in \mathsf{GF}(\mathsf{C})_k$ that distinguishes $\mathfrak{A}$ and $\mathfrak{B}$.

Now we prove that *2.* implies *1.*. Let $\varphi \in \mathsf{GF}(\mathsf{C})_k$ such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$. Using Theorem 8 we can obtain a formula ϑ_ψ for every atomic subformula ψ of φ with $\mathfrak{A} \models \psi$ if, and only if, $\mathfrak{A} \models \vartheta_\psi$. With this we can construct an equivalent formula $\varphi' \in \mathsf{GF}(\mathsf{C})_k$, which then allows us, to easily translate it to $\widetilde{\sigma}$. We will construct this formula φ' inductively and directly prove the equivalence.

Claim 10. *The two formulae φ and φ' are equivalent.*

Proof. **Base cases:** If φ is an atomic formula, that is, either a term equivalence or a relation, then set φ' to ϑ_φ. The equivalence follows directly from the above lemmas 7 and 8.

Inductive cases: In the cases where φ is of the form $\neg\vartheta$ or $\vartheta_1 \wedge \vartheta_2$, we set φ' to $\neg\vartheta'$ or $\vartheta'_1 \wedge \vartheta'_2$ and the claim follows directly using the induction hypothesis.

Let φ be of the form $\exists^{\geq \ell}\mathbf{v}.\Delta \wedge \vartheta$. In addition to translating Δ and ϑ to ϑ_Δ and ϑ' respectively, we also will need to transform the formula, so that it still is a valid formula in $\mathsf{GF}(\mathsf{C})_k$. When looking at the possible translations from the atomic formula $\Delta(x_1, \dots, x_m)$, we see that it must be of the form $\bigvee_{i\in[o]}(\Delta'_i(x_1,\dots,x_m) \wedge \bigwedge \Psi_i(x_1,\dots,x_m))$. When considering the transformed formula

$$\exists^{\geq \ell}\mathbf{v}.\left(\bigvee_{i\in[o]}(\Delta'_i \wedge \bigwedge \Psi_i) \wedge \vartheta'\right),$$

we then will distribute ϑ over the disjunction and thus define

$$\psi \coloneqq \exists^{\geq \ell}\mathbf{v}.\left(\bigvee_{i\in[o]}\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta'\right)$$

In the following we prove the equivalence of φ and ψ. Let $\mathfrak{A} \models \varphi$. This means there are at least ℓ tuples $\mathbf{a} \in A$, such that $(\mathfrak{A}, \mathbf{a}) \models \Delta(\mathbf{v}) \wedge \vartheta(\mathbf{v})$. Using the induction hypothesis we get that this is equivalent to $(\mathfrak{A}, \mathbf{a}) \models \bigvee(\Delta' \wedge \bigwedge \Psi) \wedge \vartheta'$, which, using the distributive law of propositional logic, is equivalent to $(\mathfrak{A}, \mathbf{a}) \models \bigvee(\Delta' \wedge \bigwedge \Psi \wedge \vartheta')$. Therefore the number of tuples that satisfy $\Delta \wedge \vartheta$ must be the same as for $\bigvee(\Delta' \wedge \bigwedge \Psi \wedge \vartheta')$ and $\mathfrak{A} \models \exists^{\geq \ell}\mathbf{v}.\bigvee(\Delta' \wedge \bigwedge \Psi \wedge \vartheta')$ follows.

However, we are not finished, because $\psi \notin \mathsf{GF}(\mathsf{C})_k$. We will solve this, by considering all possible segmentations of the disjunction. Formally, for $o, n \in \mathbb{N}$ we define $\mathrm{Parts}(o, n)$ as the set of all multisets with exactly n elements of $[o]$, respecting their multiplicity. We then define φ' as

$$\bigvee_{(M,\mathrm{mult}_M)\in \mathrm{Parts}(o,\ell)} \bigwedge_{i\in M} \exists^{\geq \mathrm{mult}_M(i)}\mathbf{v}.(\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta')$$

and will prove the equivalence between ψ and φ' in the following.

Let $\mathfrak{A} \models \psi$. Then there are ℓ different tuples $\mathbf{a}$, such that $\mathfrak{A}, \mathbf{a} \models \bigvee_{i\in[o]}(\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta')$. From the above lemmas we know that for every such tuple, there is exactly one i such that $\mathfrak{A}, \mathbf{a} \models \Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta'$. Now construct a multiset (M, mult_M) with exactly these i that are being satisfied and with the multiplicity of the amount of tuples satisfying them. One can see that $(M, \mathrm{mult}_M) \in \mathrm{Parts}(o, n)$ and that

$$\mathfrak{A} \models \bigwedge_{i\in M} \exists^{\geq \mathrm{mult}_M(i)}\mathbf{v}.(\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta').$$

It directly follows that $\mathfrak{A} \models \varphi'$.

Let $\mathfrak{A} \models \varphi'$. From the construction we know, that every $\mathbf{a}$ that is being quantified satisfies only the $\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta'$ they are being quantified for. By the definition of $\mathrm{Parts}(o, \ell)$, we thus get exactly ℓ tuples that satisfy some $\Delta'_i \wedge \bigwedge \Psi_i \wedge \vartheta'$ and $\mathfrak{A} \models \psi$ follows. $\square$

Note that for every term α that appears in φ', it holds that $\alpha \in \mathrm{Alters}^k_n(\sigma)$. This follows from the properties of the translation in Theorem 8. Furthermore, for every atomic subformula, we have a corresponding relation symbol in $\widetilde{\sigma}$. With this, we can transform φ' to a formula $\widetilde{\sigma} \in \mathsf{GF}(\mathsf{C})$ of signature $\widetilde{\sigma}$, such that $\mathfrak{A} \models \varphi'$ if, and only if, $\widetilde{\mathfrak{A}} \models \widetilde{\varphi}$.

It can be seen that the only subformulae that need to be changed are atomic. Let ψ be an atomic formula that appears in φ'. If ψ is a term equation, that is, it is of the form $t(x) = s(y)$, we know through the construction of φ' and the definition of the transitive expansion, that there are $\alpha, \beta \in \mathrm{Alters}^k_n(\sigma)$ with $\alpha = t$ and $\beta = s$. As such, we can replace ψ with $\mathrm{Eq}_{\alpha,\beta}(x, y)$.

If ψ is a relation, that is, it is of the form $R(t_1(x_1), \ldots, t_m(x_m))$, we again have $\alpha_1, \ldots, \alpha_m \in \text{Alters}_n^k(\sigma)$, such that $\alpha_i = t_i$ for $i \in [m]$. We then can replace ψ with $R_{\alpha_1,\ldots,\alpha_m}(x_1, \ldots, x_m)$. From the semantic definition of the transitive expansion, it can be easily seen that $\mathfrak{A} \models \varphi'$ if, and only if, $\widetilde{\mathfrak{A}} \models \widetilde{\varphi}$.

With this, we have obtained a formula $\widetilde{\varphi} \in \mathsf{GF(C)}$ of signature $\widetilde{\sigma}$, where $\widetilde{\mathfrak{A}} \models \widetilde{\varphi}$ and $\widetilde{\mathfrak{B}} \not\models \widetilde{\varphi}$. Using [16], we thus know that RCR distinguishes $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$ and by definition we can deduce that RCR_k distinguishes $\mathfrak{A}$ and $\mathfrak{B}$. □

4.3 Characterisation through homomorphism counting

One very interesting property of classical, as well as Relational Colour Refinement is that aside from its logical characterisation, it can be characterised by counting homomorphisms from certain structures. As we showed, the logical characterisation has two possible extensions to structures with functions. Thus we now want to consider, whether those extensions can also be characterised by counting homomorphisms.

In the following we will see that, using the two established approaches, it is in general not possible, to find acyclic structures with functions that divide two structures by homomorphism count. We will concentrate on the encoding defined in section 4.1, where we will show one positive and one negative result. Afterwards, we will widen the negative results to show that a corresponding characterisation for the encoding from section 4.2 cannot exist as well.

4.3.1 Acyclic structures cannot characterise naive RCR

To begin, let us define two concepts for relational structures that encode a non-relational structure.

Definition 11 (Total structures)**.** Let $\sigma := \sigma_{\text{Rel}} \,\dot\cup\, \sigma_{\text{Func}}$ be a signature where σ_{Func} contains exactly all functional symbols and is not empty. Now let σ' be the relational encoding of σ and let $\mathfrak{A}'$ ba a σ'-structure. We call $\mathfrak{A}'$ total, if for every $R_f \in \widehat{\sigma}$ with arity $n+1$ where $f \in \sigma_{\text{Func}}$ has arity n, we have that for every n-ary tuple $\mathbf{x}$ there is a y such that $(\mathbf{x}y) \in R_f^{\mathfrak{A}'}$.

Total structures therefore capture the notion that every relation that encodes a function is defined for the complete value-domain. However, functions also only valuate to exactly one value. This idea is captured by the following definition.

Definition 12 (Functional structures)**.** We define σ, σ' and $\mathfrak{A}'$ exactly as in definition 11. We call $\mathfrak{A}'$ functional, if for every $R_f \in \widehat{\sigma}$ with arity $n+1$ where $f \in \sigma_{\text{Func}}$ has arity n, we have that if $(\mathbf{x}y) \in R_f^{\mathfrak{A}'}$ there is no $y' \neq y$ such that $(\mathbf{x}y') \in R_f^{\mathfrak{A}'}$.

A graphical representation of total and functional structures can be found in figure 4.

To continue, we have to define what it means to be acyclic for a structure with functions.

Definition 13 (Acyclic, non-relational structures)**.** Let σ be a signature with function-symbols and σ' its encoding as it is defined in section 4.1. Let $\mathfrak{C}$ be a σ-structure and $\mathfrak{C}'$ be its encoding of signature σ'. We then call $\mathfrak{C}$ acyclic, if $\mathfrak{C}'$ is acyclic, with respect to acyclicity as it is defined in definition **??**.

We now want to find the equivalence between the existence of an acyclic structure with functions and an encoding of an acyclic structure with the above properties (with respect to homomorphism counting).

Lemma 14. *Let σ be a signature with function-symbols and σ' be the relational encoding of it. Let $\mathfrak{A}$ and $\mathfrak{B}$ be σ-structures and $\mathfrak{A}'$ and $\mathfrak{B}'$ be their respective encodings of signature σ'. Then the two following statements are equivalent.*

1. *There exists an acyclic structure $\mathfrak{C}$ of signature σ such that* $\hom(\mathfrak{C}, \mathfrak{A}) \neq \hom(\mathfrak{C}, \mathfrak{B})$.

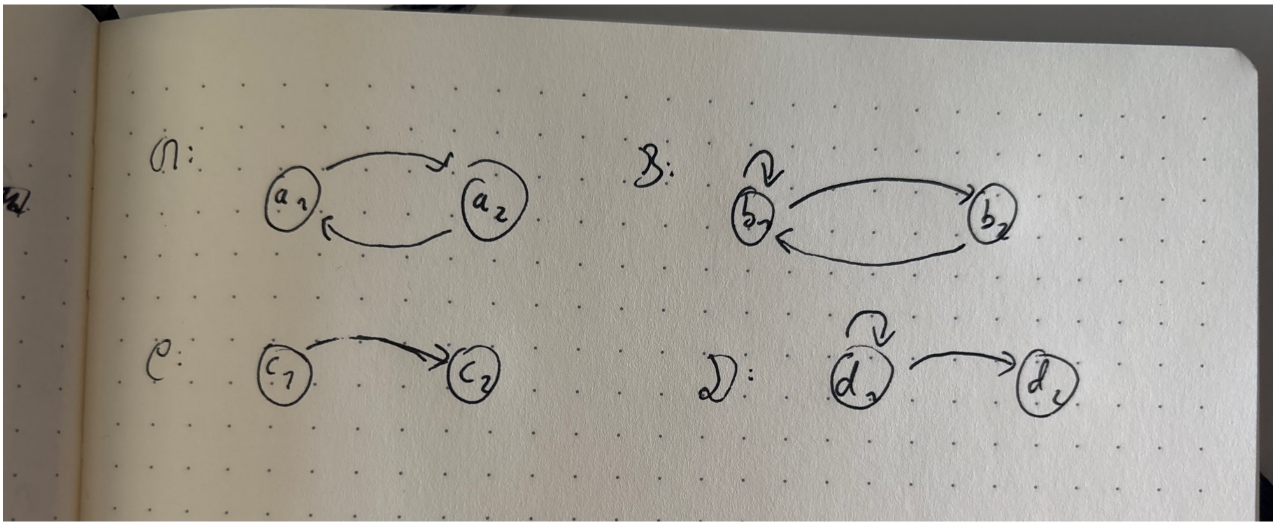

Figure 4: Examples for a total and functional ($\mathfrak{A}$), a total and not-functional ($\mathfrak{B}$), a not-total but functional ($\mathfrak{C}$) and a not-total and not-functional structure ($\mathfrak{D}$), where $\sigma = \{f/1\}$ and thus $\sigma' = \{R_f/2\}$.

2. *There exists an acyclic, total and functional structure* $\mathfrak{C}'$ *of signature* σ' *such that* $\hom(\mathfrak{C}', \mathfrak{A}') \neq \hom(\mathfrak{C}', \mathfrak{B}')$.

Proof. We begin by proving that *1.* implies *2.*. Let $\mathfrak{C}$ be such a σ-structure. Now let $\mathfrak{C}'$ be its encoding as a σ'-structure. By definition is $\mathfrak{C}'$ acyclic and when considering the definition of the encoding, we find that it also has to be total and functional. Furthermore we have $\mathrm{Hom}(\mathfrak{C}, \mathfrak{A}) = \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$ and respectively with $\mathfrak{B}$ and $\mathfrak{B}'$.

Let $\varphi \in \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$. We show that $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$. Let $(\mathbf{x}y) \in R_f^{\mathfrak{C}'}$ for a function-symbol $f \in \sigma$. Then by definition of the encoding $f^{\mathfrak{C}}(\mathbf{x}) = y$ and then $f^{\mathfrak{A}}(\varphi(\mathbf{x})) = \varphi(f^{\mathfrak{C}}(\mathbf{x})) = \varphi(y)$. Thus we have that $(\varphi(\mathbf{x})\varphi(y)) \in R_f^{\mathfrak{A}'}$, but this is equal to $\varphi(\mathbf{x}y) \in R_f^{\mathfrak{A}'}$. Let $\mathbf{x} \in R^{\mathfrak{C}'}$ where $R \neq R_f$ for all function symbols $f \in \sigma$. Then we have that $R^{\mathfrak{C}'} = R^{\mathfrak{C}}$ and $R^{\mathfrak{A}} = R^{\mathfrak{A}'}$. Therefore because $\mathbf{x} \in R^{\mathfrak{C}}$, we have $\varphi(\mathbf{x}) \in R^{\mathfrak{A}} = R^{\mathfrak{A}'}$. This was to be shown

Let $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$. We show that $\varphi \in \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$. Let $\mathbf{x} \in R^{\mathfrak{C}}$ for a relational-symbol R. By the construction of the encoding we know that $R^{\mathfrak{C}} = R^{\mathfrak{C}'}$ and $R^{\mathfrak{A}'} = R^{\mathfrak{A}}$ and with the same argument as before we can conclude that $\varphi(\mathbf{x}) \in R^{\mathfrak{A}}$. Let $\mathbf{x}$ and y be such that $f^{\mathfrak{C}}(\mathbf{x}) = y$. Then by construction we know that $(\mathbf{x}y) \in R_f^{\mathfrak{C}'}$. Because φ is a homomorphism we also know that $\varphi(\mathbf{x}y) \in R_f^{\mathfrak{A}'}$ and because $\mathfrak{A}'$ is also an encoding we have that $f^{\mathfrak{A}}(\varphi(\mathbf{x})) = \varphi(y) = \varphi(f^{\mathfrak{C}}(\mathbf{x}))$. This was to be shown. $\mathfrak{A}$ and $\mathfrak{A}'$ can be replaced by $\mathfrak{B}$ and $\mathfrak{B}'$, respectively, to get the analogous result for the other structure.

We now show that *2.* implies *1.*. Let $\mathfrak{C}'$ be an acyclic, total and functional σ'-structure. We can now construct a σ-structure $\mathfrak{C}$ by decoding $\mathfrak{C}'$ and will also get that $\mathrm{Hom}(\mathfrak{C}', \mathfrak{A}') = \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$. For a relation-symbol $R \in \sigma$ we can define $R^{\mathfrak{C}} := R^{\mathfrak{C}'}$. For a function-symbol $f \in \sigma$ we can define $f^{\mathfrak{C}}$ as follows: Let f be of arity n. Then for a n-ary tuple $\mathbf{x}$ there must be a y such that $(\mathbf{x}y) \in R_f^{\mathfrak{C}'}$ because $\mathfrak{C}'$ is total and there must be exactly one such y because $\mathfrak{C}'$ is functional. Therefore we define $f^{\mathfrak{C}}(\mathbf{x}) = y$. The claim that the sets of homomorphisms is equal can be verified using exactly the same arguments as in the analogous proof of the other direction. □

We can now continue to put this in relation to the statement regarding RCR. We have that naive RCR distinguishes two σ-structures $\mathfrak{A}$ and $\mathfrak{B}$ if, and only if, RCR distinguishes the encodings $\mathfrak{A}'$ and $\mathfrak{B}'$ of signature σ'. Due to the results of [16] this is the case if, and only if, there is an acyclic σ'-structure $\mathfrak{C}'$ such that $\hom(\mathfrak{C}', \mathfrak{A}') \neq \hom(\mathfrak{C}', \mathfrak{B}')$. We would like to achieve the result that this is equivalent to there being an acyclic σ structure that distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ by homomorphism count (however this will not be the case). By the above lemma, this is equivalent to there being an acyclic, total and functional σ'-structure $\mathfrak{C}''$ that distinguishes $\mathfrak{A}'$ and $\mathfrak{B}'$ by homomorphism count. Our goal will therefore be, to study the relationship between the two following statements.

1. There is an acyclic σ'-structure $\mathfrak{C}'$ such that $\hom(\mathfrak{C}', \mathfrak{A}') \neq \hom(\mathfrak{C}', \mathfrak{B}')$.
2. There is an acyclic, total and functional σ'-structure $\mathfrak{C}''$ such that $\hom(\mathfrak{C}'', \mathfrak{A}') \neq \hom(\mathfrak{C}'', \mathfrak{B}')$.

It is obvious that *2.* implies *1.*, however the other direction will not hold in general. In fact, we will be able to construct a functional structure from $\mathfrak{C}'$, but totality will not be able to be constructed. We will in the following prove the former claim and then show the latter claim, using a family of counterexamples.

Lemma 15. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be structures of signature σ and $\mathfrak{A}'$, $\mathfrak{B}'$ and σ' the respective encodings. If there is an acyclic structure $\mathfrak{C}'$ of signature σ' with $\hom(\mathfrak{C}', \mathfrak{A}') \neq \hom(\mathfrak{C}', \mathfrak{B}')$, then we can construct a functional, acyclic structure $\mathfrak{C}''$ of signature σ' such that $\hom(\mathfrak{C}'', \mathfrak{A}') = \hom(\mathfrak{C}', \mathfrak{A}')$ and $\hom(\mathfrak{C}'', \mathfrak{B}') = \hom(\mathfrak{C}', \mathfrak{B}')$.*

Proof. The proof for the above lemma will work as follows. If $\mathfrak{C}'$ is not functional, then there is a function-symbol $f \in \sigma$ and two different tuples $(\mathbf{x}y), (\mathbf{x}z) \in R_f^{\mathfrak{C}'}$, we call this a collision. We will give a procedure, to iteratively remove such collisions. The procedure will reduce the number of elements by one, will keep the acyclicity property and will result in a structure with the same amount of homomorphisms to $\mathfrak{A}'$ and $\mathfrak{B}'$. Thus, by continuously applying that procedure to all collisions, we will get $\mathfrak{C}''$. The algorithm must terminate, as the number of elements strictly decreases and we only consider finite structures.

The remainder of this proof will be dedicated to describing the procedure and proving the above claims. Assume that there is a function symbol $f \in \sigma$ and two different tuples $(\mathbf{x}y), (\mathbf{x}z) \in R_f^{\mathfrak{C}'}$. Our goal will be to construct a structure $\mathfrak{C}''$ of signature σ' without this particular collision and with the following properties:

a. $\|\mathfrak{C}''\| = \|\mathfrak{C}'\| - 1$
b. $\mathfrak{C}''$ is acyclic
c. $\hom(\mathfrak{C}'', \mathfrak{A}') = \hom(\mathfrak{C}', \mathfrak{A}')$ and $\hom(\mathfrak{C}'', \mathfrak{B}') = \hom(\mathfrak{C}', \mathfrak{B}')$.

We define the function $\chi : \mathbf{C}' \to \mathbf{C}''$ which will map tuples of $\mathfrak{C}'$ to tuples of $\mathfrak{C}''$ with the same arity. Concretely, for an arbitrary tuple $\mathbf{c} \in \mathbf{C}'$ of arity k and for every $i \in [k]$, we have

$$\chi(\mathbf{c})_i := \begin{cases} c_i & \text{if } c_i \notin \{y, z\} \\ v_{y,z} & \text{if } c_i \in \{y, z\} \end{cases}$$

for a new element $v_{y,z}$. In words, we replace every occurrence of y and z by $v_{y,z}$, while letting everything else stay the same. Now we can define

$$\mathfrak{C}'' := ((C' \setminus \{y, z\}) \cup \{v_{y,z}\}, \sigma),$$

where for all $R \in \sigma'$ we have $R^{\mathfrak{C}''} := \{\chi(\mathbf{c}) : \mathbf{c} \in R^{\mathfrak{C}'}\}$. We will now proceed by proving the above properties.

Property a.: This property follows directly from the definition of C''. We have $y \neq z$, $y, z \in C'$ and thus $|C' \setminus \{y, z\}| = |C'| - 2$. Furthermore, we have $v_{y,z} \notin C'$ and therefore $|(C' \setminus \{y, z\}) \cup \{v_{y,z}\}| = |C'| - 1$. This was to be shown.

Property b.: To show that $\mathfrak{C}''$ is acyclic, we first will define an undirected graph J'', will prove that it is connected and cycle free, thus a tree, and that it fulfils the join tree property for $\mathfrak{C}''$. By the assumption we know that $\mathfrak{C}'$ is acyclic and thus has a join tree J'. We further notice that $\mathbf{C}'' = \{\chi(\mathbf{c}) : \mathbf{c} \in \mathbf{C}'\}$. Now we define $V(J'') \coloneqq \mathbf{C}''$ and $E(J'') \coloneqq \{\{\chi(\mathbf{u}), \chi(\mathbf{v})\} : \{\mathbf{u}, \mathbf{v}\} \in E(J')\}$.

Claim 16. *J'' is connected.*

Proof. Consider $\mathbf{u}, \mathbf{v} \in V(J'')$. Then there are $\mathbf{a}, \mathbf{b} \in V(J')$, such that $\chi(\mathbf{a}) = \mathbf{u}$ and $\chi(\mathbf{b}) = \mathbf{v}$. By assumption, J' is a tree, so there are $\mathbf{a}_0, \mathbf{a}_1, \ldots, \mathbf{a}_k$ with $\mathbf{a}_0 = \mathbf{a}$, $\mathbf{a}_k = \mathbf{b}$ and $\{\mathbf{a}_{i-1}, \mathbf{a}_i\} \in E(J')$ for all $i \in [k]$. By definition, we have $\chi(\mathbf{a}_0), \chi(\mathbf{a}_1), \ldots, \chi(\mathbf{a}_k) \in V(J'')$ with $\chi(\mathbf{a}_0) = \chi(\mathbf{a}) = \mathbf{u}$, $\chi(\mathbf{a}_k) = \chi(\mathbf{b}) = \mathbf{v}$ and $\{\chi(\mathbf{a}_{i-1}), \chi(\mathbf{a}_i)\} \in E(J'')$ for all $i \in [k]$. Thus $\mathbf{u}$ and $\mathbf{v}$ are connected. ☐

In the following, for an arbitrary $e \in C'$, we define the set $V_e \coloneqq \{\mathbf{c} \in V(J') : e \in \mathbf{c}\}$. One can see that $V(J'_e) = V_e$, where J'_e is the subgraph of J', induced by all elements containing e.

Claim 17. *J'' is cycle-free.*

Proof. Assume that there were $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k \in V(J'')$ with $\{\mathbf{u}_{i-1}, \mathbf{u}_i\} \in E(J'')$ for all $i \in [k]$ and $\{\mathbf{u}_k, \mathbf{u}_0\} \in E(J'')$. We now define directed edges such that $e_0 = (\mathbf{u}_0, \mathbf{u}_1)$, $e_1 = (\mathbf{u}_1, \mathbf{u}_2)$, $\ldots$, $e_k = (\mathbf{u}_k, \mathbf{u}_0)$, therefore we have $e_i = (\mathbf{u}_i, \mathbf{u}_{i+1 \bmod k})$. For every edge e_i choose two elements $\mathbf{a}_i, \mathbf{b}_i \in V(J')$ such that $\{\mathbf{a}_i, \mathbf{b}_i\} \in E(J')$, $\chi(\mathbf{a}_i) = \mathbf{u}_i$ and $\chi(\mathbf{b}_i) = \mathbf{u}_{i+1 \bmod k}$. These elements must exist by the definition of $E(J'')$.

We now prove that there exists a cycle in J', which would contradict our assumption of J' being a join tree for $\mathfrak{C}'$. To show this, we prove that for all $i \in \{0\} \cup [k]$, the elements $\mathbf{b}_i$ and $\mathbf{a}_{i+1 \bmod k}$ are connected in J'. We see that $\chi(\mathbf{b}_i) = \mathbf{u}_{i+1 \bmod k} = \chi(\mathbf{a}_{i+1 \bmod k})$.

If there is a $c \in \mathrm{set}(\mathbf{u}_{i+1 \bmod k})$ with $c \neq v_{y,z}$, then by definition of χ, we have $c \in \mathrm{set}(\mathbf{b}_i) \cap \mathrm{set}(\mathbf{a}_{i+1 \bmod k})$. Therefore $\mathbf{b}_i, \mathbf{a}_{i+1 \bmod k} \in V_c$ and because J' is a join tree, $\mathbf{b}_i$ and $\mathbf{a}_{i+1 \bmod k}$ have to be connected.

If $\mathrm{set}(\mathbf{u}_{i+1 \bmod k}) = \{v_{y,z}\}$, we have four possible cases. If $y \in \mathrm{set}(\mathbf{b}_i) \cap \mathrm{set}(\mathbf{a}_{i+1 \bmod k})$ or $z \in \mathrm{set}(\mathbf{b}_i) \cap \mathrm{set}(\mathbf{a}_{i+1 \bmod k})$, then we can do the same as before by setting $c = y$ or $c = z$, respectively. Otherwise we have $y \in \mathrm{set}(\mathbf{b}_i)$ and $z \in \mathrm{set}(\mathbf{a}_{i+1 \bmod k})$, or $z \in \mathrm{set}(\mathbf{b}_i)$ and $y \in \mathrm{set}(\mathbf{a}_{i+1 \bmod k})$. We will only consider the former option, as the latter can be proven analogously. From our beginning assumption we know that $(\mathbf{x}y), (\mathbf{x}z) \in R_f$. Choose some $x \in \mathbf{x}$, and $(\mathbf{x}y), (\mathbf{x}z) \in V_x$ follows. Furthermore, we have $\mathbf{b}_i, (\mathbf{x}y) \in V_y$ and $\mathbf{a}_{i+1 \bmod k}, (\mathbf{x}z) \in V_z$. Since J' is a join tree, we thus know that $\mathbf{b}_i$ is connected with $(\mathbf{x}y)$, which in turn is connected with $(\mathbf{x}z)$, which again is connected with $\mathbf{a}_{i+1 \bmod k}$. Therefore $\mathbf{b}_i$ and $\mathbf{a}_{i+1 \bmod k}$ are connected.

Thus we have found a cycle in J', which is a contradiction to it being a join tree. Therefore our assumption of the existence of the elements $\mathbf{u}_0, \mathbf{u}_1, \ldots, \mathbf{u}_k$ has to be false. ☐

The last missing piece to prove the acyclicity of $\mathfrak{C}''$ is to show that J'' fulfils the join tree property. That is, for any $c \in C''$, the set $\{\mathbf{c} \in V(J'') : c \in \mathrm{set}(\mathbf{c})\}$ induces a connected subgraph of J''.

Claim 18. *J'' is a valid join tree.*

Proof. Consider any $c \in C''$ and two elements $\mathbf{u}$ and $\mathbf{v}$ from the set $S \coloneqq \{\mathbf{c} \in V(J'') : c \in \mathrm{set}(\mathbf{c})\}$. We show that there is a path from $\mathbf{u}$ to $\mathbf{v}$ in S. If $v \neq v_{y,z}$, then $v \in C'$, thus consider V_c. By definition there are $\mathbf{a}, \mathbf{b} \in V_c$ such that $\chi(\mathbf{a}) = \mathbf{u}$ and $\chi(\mathbf{b}) = \mathbf{v}$. Because J' is a join tree, V_c must induce a connected subtree. Thus there must be $\mathbf{a}_0, \ldots, \mathbf{a}_k \in V_c$ such that $\mathbf{a}_0 = \mathbf{a}$, $\mathbf{a}_k = \mathbf{b}$ and $\{\mathbf{a}_{i-1}, \mathbf{a}_i\} \in E(J')$ for all $i \in [k]$. By definition we then get, that there must be $\chi(\mathbf{a}_0), \chi(\mathbf{a}_k) \in S$ and we get $\chi(\mathbf{a}_0) = \chi(\mathbf{a}) = \mathbf{u}$, $\chi(\mathbf{a}_k) = \chi(\mathbf{b}) = \mathbf{v}$ and $\{\chi(\mathbf{a}_{i-1}, \mathbf{a}_i)\} \in E(J''))$ for all $i \in [k]$. Therefore, $\mathbf{u}$ and $\mathbf{v}$ are connected in S.

If $v = v_{y,z}$, we again define $\mathbf{a}, \mathbf{b} \in V(J')$ such that $\chi(\mathbf{a}) = \mathbf{u}$ and $\chi(\mathbf{b}) = \mathbf{v}$. If $\mathbf{a}, \mathbf{b} \in V_y$ or $\mathbf{a}, \mathbf{b} \in V_z$, then we can proceed exactly as in the former case with $v = y$ or $v = z$, respectively. If that is not the case, then either $\mathbf{a} \in V_y and \mathbf{b} \in V_z$, or the other way round. We will only prove the former case, as the latter case can proven analogously. Since the following will depend on it, we will now prove the following equality: $S = \{\chi(\mathbf{c}) : \mathbf{c} \in V_y\} \cup \{\chi(\mathbf{c}) : \mathbf{c} \in V_z\}$.

$\supseteq$: Let $\mathbf{u} \in \{\chi(\mathbf{c}) : \mathbf{c} \in V_y\} \cup \{\chi(\mathbf{c}) : \mathbf{c} \in V_z\}$. We then get that $\mathbf{u} = \chi(\mathbf{a})$ for an $\mathbf{a} \in V(J')$. From the definition it follows that $y \in \mathrm{set}(\mathbf{a})$ or $z \in \mathrm{set}(\mathbf{a})$ has to hold. Thus we get that $v_{y,z} \in \mathrm{set}(\chi(\mathbf{a})) = \mathrm{set}(\mathbf{u})$ and therefore $\mathbf{u} \in S$ has to hold.

$\subseteq$: Let $\mathbf{u} \in S$, then $v_{y,z} \in \mathrm{set}(\mathbf{u})$ follows. By definition there must exist an $\mathbf{a} \in V(J')$ such that $\chi(\mathbf{a}) = \mathbf{u}$ and $y \in \mathrm{set}(\mathbf{a})$ or $z \in \mathrm{set}(\mathbf{a})$ has to hold. Thus we get $\chi(\mathbf{a}) = \mathbf{u} \in \{\chi(\mathbf{c}) : \mathbf{c} \in V_y\} \cup \{\chi(\mathbf{c}) : \mathbf{c} \in V_z\}$

It is obvious that $(\mathbf{x}y) \in V_y$ and $(\mathbf{x}z) \in V_z$. We now get a path in V_y with the elements $\mathbf{a}_0, \dots, \mathbf{a}_k \in V_b$ such that $\mathbf{a}_0 = \mathbf{a}$, $\mathbf{a}_k = (\mathbf{x}y)$ and $\{\mathbf{a}_{i-1}, \mathbf{a}_i\} \in E(J')$ for all $i \in [k]$. We also get a path in V_z with the elements $\mathbf{b}_0, \dots, \mathbf{b}_\ell \in V_z$ such that $\mathbf{b}_0 = (\mathbf{x}z)$, $\mathbf{b}_\ell = \mathbf{b}$ and $\{\mathbf{b}_{i-1}, \mathbf{b}_i\} \in E(J')$. Therefore with the above equation, we have two paths in S: $\chi(\mathbf{a}_0), \dots, \chi(\mathbf{a}_k)$ and $\chi(\mathbf{b}_0), \dots, \chi(\mathbf{b}_\ell)$, where $\chi(\mathbf{a}_0) = \mathbf{u}$ and $\chi(\mathbf{b}_\ell) = \mathbf{v}$. Now we have that $\chi(\mathbf{a}_k) = \chi((\mathbf{x}y)) = (\mathbf{x}' v_{y,z}) = \chi((\mathbf{x}z)) = \chi(\mathbf{b}_0)$. Therefore, we get one path from $\mathbf{u}$ to $\mathbf{v}$ in S. $\square$

Property c.: To show that $\hom(\mathfrak{C}'', \mathfrak{A}') = \hom(\mathfrak{C}', \mathfrak{A}')$ and $\hom(\mathfrak{C}'', \mathfrak{B}') = \hom(\mathfrak{C}', \mathfrak{B}')$, we will give a mapping $\pi : \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}') \to \mathrm{Hom}(\mathfrak{C}'', \mathfrak{A}')$ from homomorphisms from $\mathfrak{C}'$ to $\mathfrak{A}'$ to homomorphisms from $\mathfrak{C}''$ to $\mathfrak{A}'$. We will then show that π is a bijection, from which follows that both sets have the same cardinality, which then proves the claim for $\mathfrak{A}'$. For $\mathfrak{B}'$, the prove is completely analogous, which is why it will be omitted. Let $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$. We now define $\varphi' \coloneqq \pi(\varphi)$ as

$$\varphi'(x) = \begin{cases} \varphi(x) & \text{if } x \neq v_{y,z} \\ \varphi(y) & \text{if } x = v_{y,z}. \end{cases}$$

In the following we will be using that for any $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$, we have that $\varphi(y) = \varphi(z)$. Otherwise, from $(\mathbf{x}y), (\mathbf{x}z) \in R_f^{\mathfrak{C}'}$, it follows that $(\varphi(\mathbf{x})\varphi(y)), (\varphi(\mathbf{x})\varphi(z)) \in R_f^{\mathfrak{A}'}$ for two different tuples $(\varphi(\mathbf{x})\varphi(y))$ and $(\varphi(\mathbf{x})\varphi(z))$. However, this would contradict that, by definition of the encoding, $\mathfrak{A}'$ is functional.

Claim 19. *For all $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$, we have $\pi(\varphi) \in \mathrm{Hom}(\mathfrak{C}'', \mathfrak{A}')$.*

Proof. Let $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$ and $\varphi' \coloneqq \pi(\varphi)$. Consider a relation-symbol R and a tuple $\mathbf{c} \in R^{\mathfrak{C}''}$. If $v_{y,z} \notin \mathrm{set}(\mathbf{c})$, then $\chi(\mathbf{c}) = \mathbf{c}$, $\varphi(\mathbf{c}) = \varphi'(\mathbf{c})$ and $\mathbf{c} \in R^{\mathfrak{C}'}$ follows. Then we get $\varphi(\mathbf{c}) \in R^{\mathfrak{A}'}$ and further $\varphi'(\mathbf{c}) \in R^{\mathfrak{A}'}$.

If $v_{y,z} \in \mathrm{set}(\mathbf{c})$, then there exists a $\mathbf{c}' \in R^{\mathfrak{C}'}$ such that $\chi(\mathbf{c}') = \mathbf{c}$. We thus get $\varphi(\mathbf{c}') \in R^{\mathfrak{A}'}$. We also have that $\varphi(\mathbf{c}') = \varphi'(\mathbf{c})$, because for all $x \in \mathrm{set}(\mathbf{c}) \setminus \{v_{y,z}\}$, we also have $x \in \mathrm{set}(\mathbf{c}')$ and $\varphi(x) = \varphi'(x)$. For $v_{y,z}$ we have that $\varphi'(v_{y,z}) = \varphi(y) = \varphi(z)$. Therefore, $\varphi'(\mathbf{c}) \in R^{\mathfrak{A}'}$ follows, which was to be shown. $\square$

This shows that π is correctly defined as a mapping between homomorphisms. The two following proofs will show that π is bijective.

Claim 20. *π is injective.*

Proof. Let $\varphi_1, \varphi_2 \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$, $\varphi_1 \neq \varphi_2$ and define $\varphi_1' \coloneqq \pi(\varphi_1)$ and $\varphi_2' \coloneqq \pi(\varphi_2)$. Our goal is to show that $\varphi_1' \neq \varphi_2'$. There has to be a $u \in C'$ such that $\varphi_1(u) \neq \varphi_2(u)$, otherwise they would be the same function. We now do a case distinction.

Case 1: $u \notin \{y, z\}$. Then we have

$$\varphi_1'(u) = \varphi_1(u) \neq \varphi_2(u) = \varphi_2'(u).$$

Case 2: $u \in \{y, z\}$. Then we have

$$\varphi_1'(v_{y,z}) = \varphi_1(y) = \varphi_1(z) = \varphi_1(u) \neq \varphi_2(u) = \varphi_2(z) = \varphi_2(y) = \varphi_2'(v_{y,z}).$$

In both cases we thus have found an element that gets mapped differently, thus $\varphi_1' \neq \varphi_2'$ must follow. ☐

Claim 21. *π is surjective.*

Proof. Let $\varphi' \in \mathrm{Hom}(\mathfrak{C}'', \mathfrak{A}')$. We now construct a $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$ such that $\pi(\varphi) = \varphi'$. We define

$$\varphi(x) = \begin{cases} \varphi'(x) & \text{if } x \notin \{y, z\} \\ \varphi'(v_{y,z}) & \text{if } x \in \{y, z\}. \end{cases}$$

Using that $\psi(y) = \psi(z)$ for all $\psi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$, it can easily be verified that $\pi(\varphi) = \varphi'$. We now only have to show that $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A}')$. Let $\mathbf{c} \in R^{\mathfrak{C}'}$ for a relation-symbol R. We now define $\mathbf{c}' \coloneqq \chi(\mathbf{c})$ and get by construction that $\mathbf{c}' \in R^{\mathfrak{C}''}$. Because we know that φ' is a homomorphism, it follows that $\varphi'(\mathbf{c}') \in R^{\mathfrak{A}'}$. But since $\varphi'(a) = \varphi(a)$ for all $a \notin \{y, z, v_{y,z}\}$ and $\varphi'(v_{y,z}) = \varphi(y) = \varphi(z)$, it follows that $\varphi'(\mathbf{c}') = \varphi(\mathbf{c})$. Therefore $\varphi(\mathbf{c}) \in R^{\mathfrak{A}'}$. ☐

We now have proved all properties. Therefore we have devised a procedure, to iteratively construct a functional, acyclic structure, that divides $\mathfrak{A}'$ and $\mathfrak{B}'$ by homomorphism count. ☐

5 Relational Colour Refinement for symmetric structures

One very interesting subclass of relational structures is the class of symmetric structures. A special case of these are for example undirected graphs, as their edge relations are symmetric. This notion of symmetry can be generalized to any relational signature.

Definition 22 (Symmetric Structures)**.** Let σ be a relational signature. A structure $\mathfrak{A}$ of signature σ is a symmetric structure, if for every relation and every tuple in those relations, the order of the elements is irrelevant. This means, that every relation R with arity k is a subset of all possible subsets of A with exactly k elements. Formally, that means

$$R \subseteq \binom{A}{k}.$$

An equivalent characterisation uses the symmetric groups $\mathcal{S}_k$. We call a σ structure $\mathfrak{A}$ symmetric, if for every $R \in \sigma$ of arity k, every k-tuple $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in R^{\mathfrak{A}}$ and every k-permutation $\pi \in \mathcal{S}_k$, we have that

$$(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)}) \in R^{\mathfrak{A}}.$$

In the following, we will use $\pi(\mathbf{x})$ as a shorthand notation for $(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)})$.

As symmetric structures are a subset of relational structures, the results from [16] obviously apply to them. Thus, we have that the following three statements are equivalent for two symmetric σ structures $\mathfrak{A}$ and $\mathfrak{B}$:

1. RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$.
2. There exists a sentence $\varphi \in \mathsf{GF}(\mathsf{C})$, such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$.
3. There exists an acyclic σ structure $\mathfrak{C}$, such that $\hom(\mathfrak{C}, \mathfrak{A}) \neq \hom(\mathfrak{C}, \mathfrak{B})$.

However, as we restricted the class of structures for $\mathfrak{A}$ and $\mathfrak{B}$, this poses the question, whether the same can be done to the acyclic structures. Concretely, we want to investigate, whether the first statement is also equivalent to there being an acyclic, symmetric σ structure, such that it has a different homomorphism count to $\mathfrak{A}$ than to $\mathfrak{B}$.

As we will prove in the following, it is indeed the case that we can restrict the class of acyclic structures to only include structures that are acyclic and symmetric. However, before we prove this, we have to show a lemma which will be used in the proof. As a reminder on notation, for a k-tuple $\mathbf{x} = (x_1, x_2, \ldots, x_k)$, a homomorphism φ and a permutation π, we write $\varphi(\mathbf{x})$ for $(\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_k))$ and $\pi(\mathbf{x})$ for $(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)})$.

Lemma 23. *Let $\pi \in \mathcal{S}_k$, φ be a homomorphism, R a relation of arity k and $\mathbf{x} = (x_1, x_2, \ldots, x_k) \in R$. Then $\varphi(\pi(\mathbf{x})) = \pi(\varphi(\mathbf{x}))$.*

Proof. We prove this by contradiction. Assume the contrary. Then there exists an $i \in [k]$, such that $\varphi(\pi(\mathbf{x}))_i \neq \pi(\varphi(\mathbf{x}))_i$. Note that the definitions of $\varphi(\pi(\mathbf{x}))$ and $\pi(\varphi(\mathbf{x}))$ are

$$\varphi(\pi(\mathbf{x})) = (\varphi(x_{\pi(1)}), \varphi(x_{\pi(2)}), \ldots, \varphi(x_{\pi(k)}))$$

and

$$\pi(\varphi(\mathbf{x})) = (\varphi(\mathbf{x})_{\pi(1)}, \varphi(\mathbf{x})_{\pi(2)}, \ldots, \varphi(\mathbf{x})_{\pi(k)}).$$

From these, we directly get

$$\varphi(\pi(\mathbf{x}))_i = \varphi(x_{\pi(i)}) = (\varphi(x_1), \varphi(x_2), \ldots, \varphi(x_2))_{\pi(i)} = \varphi(\mathbf{x})_{\pi(i)} = \pi(\varphi(\mathbf{x}))_i.$$

Contradiction! Therefore the lemma must hold. □

We now prove the above claim:

Theorem 24. *Let σ be a relational signature and $\mathfrak{A}$ and $\mathfrak{B}$ be two σ structures. Then the following two statements are equivalent:*

1. RCR *distinguishes* $\mathfrak{A}$ *and* $\mathfrak{B}$.
2. *There exists an acyclic, symmetric σ structure $\mathfrak{C}$ with* $\hom(\mathfrak{C}, \mathfrak{A}) \neq \hom(\mathfrak{C}, \mathfrak{B})$.

Proof. We first prove that *2.* implies *1.* Let $\mathfrak{C}$ be an acyclic, symmetric σ structure with $\hom(\mathfrak{C}, \mathfrak{A}) \neq \hom(\mathfrak{C}, \mathfrak{B})$. As $\mathfrak{C}$ is acyclic, we can apply the equivalence seen in **??** and get that RCR must distinguish $\mathfrak{A}$ and $\mathfrak{B}$.

We now prove that *1.* implies *2.* Assume that RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$. From **??** we know that there exists an acyclic structure $\mathfrak{C}'$ with $\hom(\mathfrak{C}', \mathfrak{A}) \neq \hom(\mathfrak{C}', \mathfrak{B})$. Our goal will be to construct a σ structure $\mathfrak{C}$ from $\mathfrak{C}'$ that is both acyclic and symmetric. Informally, $\mathfrak{C}'$ will have the same elements as $\mathfrak{C}$ and for every tuple that appears in some relation, we will add all possible permutations of that tuple to the relation as well. Formally, we define $\mathfrak{C} := (C', \sigma)$ and for all $R \in \sigma$ with arity k, we have

$$R^{\mathfrak{C}} := \{(x_{\pi(1)}, x_{\pi(2)}, \ldots, x_{\pi(k)}) : \text{for every } (x_1, x_2, \ldots, x_k) \in R^{\mathfrak{C}'} \text{ and every } \pi \in \mathcal{S}_k\}.$$

From the second characterisation of symmetric structures given above, it is obvious that $\mathfrak{C}$ is symmetric.

Claim 25. *$\mathfrak{C}$ is acyclic.*

Proof. We define a join-tree J for $\mathfrak{C}$. Since $\mathfrak{C}'$ is acyclic, we have a join-tree J' for $\mathfrak{C}'$. From the definition we know that $V(J) = \mathbf{C}$, thus we only have to define the set of edges. Let $\mathbf{x} \in V(J) \setminus V(J')$. From the construction there exists a permutation $\pi_{\mathbf{x}}$, such that $\pi_{\mathbf{x}}(\mathbf{x}) \in V(J')$. We now define $E(J) := E(J') \cup \{\{\pi_{\mathbf{x}}(\mathbf{x}), \mathbf{x}\} : \mathbf{x} \in \mathbf{C} \setminus \mathbf{C}'\}$. This construction can be seen in figure 5.

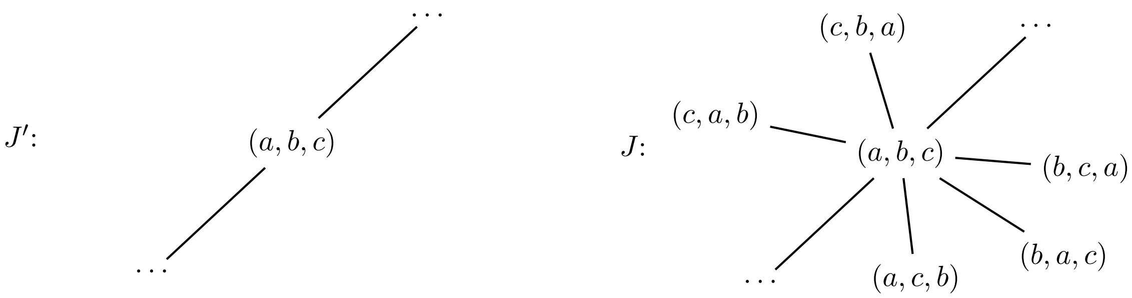

Figure 5: A section from the join tree J' and the join tree J generated from it. We consider a tuple $\mathbf{x} = (a,b,c)$, for which no other permutation appears in $\mathbf{C}'$.

The connectedness and cycle-freeness follows directly from the fact that J' is also a tree. As such, it only remains to show the join-tree property. Consider an arbitrary $v \in C$. Since $C = C'$, we have that $v \in C'$ and the set of all $\mathbf{x} \in \mathbf{C}'$ with $v \in \mathrm{set}(\mathbf{x})$ induces a connected subgraph. Let $\mathbf{x} \in \mathbf{C} \setminus \mathbf{C}'$ and $v \in \mathrm{set}(\mathbf{x})$. Then $\pi_{\mathbf{x}}(\mathbf{x}) \in \mathbf{C}'$ and $\{\pi_{\mathbf{x}(\mathbf{x})}, \mathbf{x}\} \in E(J)$, thus $\mathbf{x}$ is also connected and the set $\{\mathbf{x} \in V(J) : v \in \mathrm{set}(\mathbf{x})\}$ also induces a connected subgraph. This was to be shown. $\square$

It now remains to prove, that $\mathfrak{C}$ also has a different number of homomorphisms to $\mathfrak{A}$, than to $\mathfrak{B}$. In fact, we will show that $\mathfrak{C}$ and $\mathfrak{C}'$ have exactly the same homomorphisms to $\mathfrak{A}$ and $\mathfrak{B}$, respectively. Formally, we will prove that $\mathrm{Hom}(\mathfrak{C}', \mathfrak{A}) = \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$ and $\mathrm{Hom}(\mathfrak{C}', \mathfrak{B}) = \mathrm{Hom}(\mathfrak{C}, \mathfrak{B})$. However, we will only prove the claim for $\mathfrak{A}$, as the case for $\mathfrak{B}$ can be proven completely analogously.

Let $\varphi \in \mathrm{Hom}(\mathfrak{C}', \mathfrak{A})$. Then for every $R \in \sigma$, we have that if $\mathbf{x} \in R^{\mathfrak{C}'}$, then $\varphi(\mathbf{x}) \in R^{\mathfrak{A}}$. Now consider $\mathbf{x} \in R^{\mathfrak{C}}$ for a $R \in \sigma$ with arity k and we will proceed with a case distinction. If $\mathbf{x} \in \mathbf{C}'$, then we have $\mathbf{x} \in R^{\mathfrak{C}'}$ and by assumption $\varphi(\mathbf{x}) \in R^{\mathfrak{A}}$. If $\mathbf{x} \in \mathbf{C} \setminus \mathbf{C}'$, then there must be a $\pi \in \mathcal{S}_k$, such that $\pi(\mathbf{x}) \in \mathbf{C}'$ and further $\pi(\mathbf{x}) \in R^{\mathfrak{C}'}$. Then by assumption we have that $\varphi(\pi(\mathbf{x})) \in R^{\mathfrak{A}}$. Using Lemma 23, we know that $\varphi(\pi(\mathbf{x})) = \pi(\varphi(\mathbf{x})) \in R^{\mathfrak{A}}$. Now let $\pi' \in \mathcal{S}_k$, such that $\pi' \circ \pi = \mathrm{id} \in \mathcal{S}_k$. As $\mathfrak{A}$ is symmetric, we know that $\pi'(\pi(\varphi(\mathbf{x}))) \in R^{\mathfrak{A}}$ and further we get that $\pi'(\pi(\varphi(\mathbf{x}))) = \varphi(\mathbf{x})$. Therefore $\varphi(\mathbf{x}) \in R^{\mathfrak{A}}$ and $\varphi \in \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$ follows.

Now let $\varphi \notin \mathrm{Hom}(\mathfrak{C}', \mathfrak{A})$. Then there is a $R \in \sigma$ with arity k and a $\mathbf{x} \in R^{\mathfrak{C}'}$ with $\varphi(\mathbf{x}) \notin R^{\mathfrak{A}}$. From the definition we get that $\mathbf{x} \in \mathbf{C}$ and thus $\mathbf{x} \in R^{\mathfrak{C}}$ and from the assumption we get that $\varphi(\mathbf{x}) \notin R^{\mathfrak{A}}$. Therefore $\varphi \notin \mathrm{Hom}(\mathfrak{C}, \mathfrak{A})$. $\square$

With this we have proven that it is possible to only consider symmetric acyclic structures with a different homomorphism count, when trying to distinguish symmetric structures.

6 Conclusion

References

[1] László Babai. Graph Isomorphism in Quasipolynomial Time, January 2016. arXiv:1512.03547, doi:10.48550/arXiv.1512.03547.

[2] László Babai, Paul Erdős, and Stanley M. Selkow. Random Graph Isomorphism. *SIAM Journal on Computing*, 9(3):628–635, August 1980. doi:10.1137/0209047.

[3] Christoph Berkholz, Paul Bonsma, and Martin Grohe. Tight Lower and Upper Bounds for the Complexity of Canonical Colour Refinement. *Theory of Computing Systems*, 60(4):581–614, May 2017. doi:10.1007/s00224-016-9686-0.

[4] Jan Böker. Color Refinement, Homomorphisms, and Hypergraphs. In Ignasi Sau and Dimitrios M. Thilikos, editors, *Graph-Theoretic Concepts in Computer Science*, pages 338–350, Cham, 2019. Springer International Publishing. doi:10.1007/978-3-030-30786-8_26.

[5] Silvia Butti and Victor Dalmau. Fractional homomorphism, Weisfeiler-Leman invariance, and the Sherali-Adams hierarchy for the Constraint Satisfaction Problem, July 2021. arXiv:2107.02956, doi:10.48550/arXiv.2107.02956.

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

[7] Holger Dell, Martin Grohe, and Gaurav Rattan. Lovász Meets Weisfeiler and Leman, May 2018. arXiv:1802.08876, doi:10.48550/arXiv.1802.08876.

[8] Zdeněk Dvořák. On recognizing graphs by numbers of homomorphisms. *Journal of Graph Theory*, 64(4):330–342, 2010. doi:10.1002/jgt.20461.

[9] Martin Grohe, Kristian Kersting, Martin Mladenov, and Pascal Schweitzer. Color Refinement and Its Applications. August 2021. doi:10.7551/mitpress/10548.003.0023.

[10] Martin Grohe, Kristian Kersting, Martin Mladenov, and Erkal Selman. Dimension Reduction via Colour Refinement. In Andreas S. Schulz and Dorothea Wagner, editors, *Algorithms - ESA 2014*, pages 505–516, Berlin, Heidelberg, 2014. Springer. doi:10.1007/978-3-662-44777-2_42.

[11] Martin Grohe and Pascal Schweitzer. The graph isomorphism problem. *Commun. ACM*, 63(11):128–134, October 2020. doi:10.1145/3372123.

[12] Neil Immerman and Eric Lander. Describing Graphs: A First-Order Approach to Graph Canonization. In Alan L. Selman, editor, *Complexity Theory Retrospective: In Honor of Juris Hartmanis on the Occasion of His Sixtieth Birthday, July 5, 1988*, pages 59–81. Springer, New York, NY, 1990. doi:10.1007/978-1-4612-4478-3_5.

[13] H. L. Morgan. The Generation of a Unique Machine Description for Chemical Structures-A Technique Developed at Chemical Abstracts Service. *Journal of Chemical Documentation*, 5(2):107–113, May 1965. doi:10.1021/c160017a018.

[14] Louis C. Ray and Russell A. Kirsch. Finding Chemical Records by Digital Computers. *Science*, 126(3278):814–819, October 1957. doi:10.1126/science.126.3278.814.

[15] Benjamin Scheidt and Nicole Schweikardt. Counting Homomorphisms from Hypergraphs of Bounded Generalised Hypertree Width: A Logical Characterisation, August 2023. arXiv:2303.10980, doi:10.48550/arXiv.2303.10980.

[16] Benjamin Scheidt and Nicole Schweikardt. Color Refinement for Relational Structures, January 2025. arXiv:2407.16022, doi:10.48550/arXiv.2407.16022.

[17] S Vichy N Vishwanathan, Nicol N Schraudolph, Risi Kondor, and Karsten M Borgwardt. Graph kernels. *The Journal of Machine Learning Research*, 11:1201–1242, 2010.

[18] 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.