Relational Colour Refinement for Non-Relational Signatures

Bachelor's Thesis

Theodor Teslia

August 28, 2025

RWTH Aachen University

1 Introduction

2 Relational Colour Refinement

3 Relational Colour Refinement for structures with functions

3.1 Naive Encoding of functions

A simple way to apply relational colour refinement to non-relational structures is, to encode the functions in the signature as a relation. 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 [1], 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 expanding formulae $R_f(x_1, \ldots, x_\ell, y)$ to $f(x_1, \ldots, x_\ell) = y$ 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 [1] to a distinguishing colouring of the encoded structures, which by definition is a distinguishing colouring for the structures themselves. □

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 such a 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

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

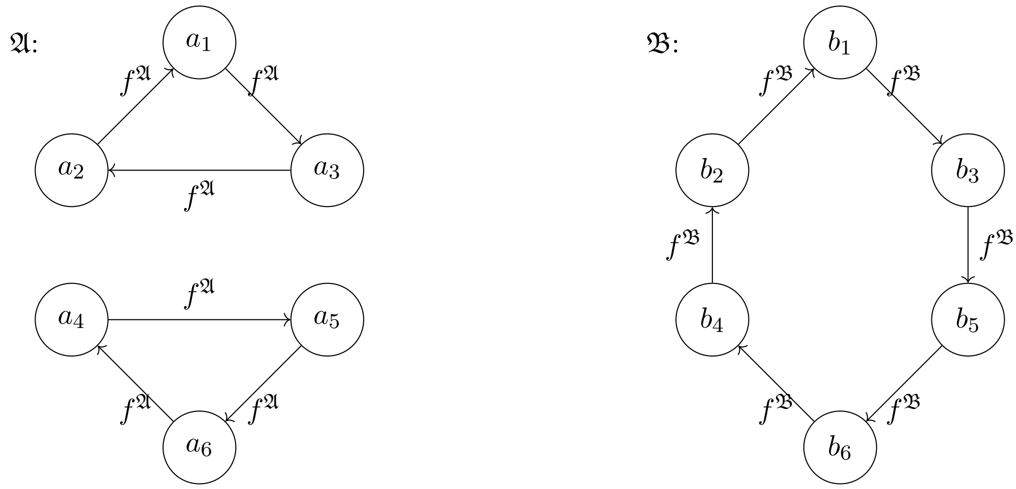

Figure 1: Two σ-structures $\mathfrak{A}$ and $\mathfrak{B}$

Consider the formula $\varphi = \exists x.(f(f(f(x))) = x)$ which utilizes term nesting to find a cycle with 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.

3.2 Using the transitive expansion

Let

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

The set will represents 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 $\mathcal{I}(n, m)$.

Lemma 3. *Let $\psi(x_1, x_2) \coloneqq f^m(x_1) = x_2$. 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, $m' \leq n$.

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

Otherwise, we define

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

where

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

and

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

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 use $f_{\mathfrak{A}}$ 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, 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 $k := i$, $\ell := j - i$ and $p := (m - i) \mod (j - i) = (m - i) \mod \ell$. 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.

$f^{k+p}(x_1) = x_2$: 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 $\mathfrak{A}, a_1, a_2 \models f^{k+p}(x_1) = x_2$.

$f^k(x_1) = f^{k+\ell}(x_1)$: 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 $\mathfrak{A}, a_1, a_2 \models f^k(x_1) = f^{k+\ell}(x_1)$.

$\mathrm{E}_f^{k,\ell}(x_1)$: This has to be 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. □

The above proof allows for the translation of formulae $f^m(x) = y$ 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 $g^m(f^{m'}(x)) = y$. Before we prove this, we want to give a small example of how the construction will work.

Consider the formula $\varphi(x_1, x_2) \coloneqq g^3(f^4(x_1)) = x_2$ and a structure with 2 elements. Let us begin by considering the formula $\varphi_1(y_1, y_2) \coloneqq f^4(y_1) = y_2$. By the above proof, we can translate that into the formula

$$\begin{aligned}\varphi_1' \coloneqq &(y_1 = y_2 \wedge y_1 = f(y_1) \wedge \mathrm{E}_f^{0,1}(y_1)) \\ &\vee (y_1 = y_2 \wedge y = 1 = f(f(y_1)) \wedge \mathrm{E}_f^{0,2}(y_1) \wedge y_1 \neq f(y_1)) \\ &\vee (f(y_1) = y_2 \wedge f(y_1) = f(f(y_1)) \wedge \mathrm{E}_f^{1,1}(y_1))\end{aligned}$$

which can also be written as

$$\varphi_1' \coloneqq \bigvee_{(k,\ell,p)\in\mathcal{I}(2,4)} \left(f^{k+p}(y_1) = y_2 \wedge \bigwedge \Psi_{1,(k,\ell,p)}\right).$$

Analogously we can write $\varphi_2 \coloneqq g^3(y_1) = y_2$ as

$$\varphi_2' \coloneqq \bigvee_{(k',\ell',p')\in\mathcal{I}(2,3)} \left(f^{k'+p'}(y_1) = y_2 \wedge \bigwedge \Psi_{2,(k',\ell',p')}\right).$$

As such, the translated formula will look like this:

$$\begin{aligned}\exists^{\geq 1} x_{f^4(x_1)}. \bigvee_{(k',\ell',p')\in\mathcal{I}(2,3)} \bigvee_{(k,\ell,p)\in\mathcal{I}(2,4)} &(f^{k+p}(x_1) = x_{f^4(x_1)} \wedge \Psi_{1,(k,\ell,p)} \\ &\wedge g^{k'+p'}(x_{f^4(x_1)}) = x_2 \wedge \Psi_{2,(k',\ell',p')})\end{aligned}$$

Genauere Erklärungen folgen noch.

Lemma 4. *Let $\psi(x_1, x_2) \coloneqq t(x_1) = x_2$ be an atomic formula. Then there exists a formula $\vartheta_t(x_1, x_2) \in$ $\mathsf{GF}(\mathsf{C})$, such that for any structure (of a fitting signature) $\mathfrak{A}$ with $\|\mathfrak{A}\| = n$ it holds*

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

Furthermore, $\vartheta_t(x_1, x_2)$ is of the form $\exists^{\geq 1}\mathbf{x}. \bigvee_i \zeta_i(x_1, x_2, \mathbf{x})$ where all $\zeta_i(x_1, x_2, \mathbf{x})$ are of the form

$$\begin{aligned}&t_{i,1}'(x_1) = x_{s_1(x_1)} \wedge \bigwedge \Psi_{i,1}(x_1) \\ &\wedge t_{i,2}'(x_{s_1(x_1)}) = x_{s_2(s_1(x_1))} \wedge \bigwedge \Psi_{i,2}(x_{s_1(x_1)}) \\ &\wedge \dots \\ &\wedge t_{i,k}'(x_{s_{k-1}(\dots(s_1(x_1)))}) = x_2 \wedge \bigwedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))})\end{aligned}$$

for terms $t_{i,j}'(x)$, $j \in [k]$, and $\mathrm{ar}(\mathbf{x}) = k - 1$. And for every $f^m(x)$ that appears in ϑ_t where f is a unary function and $m \in \mathbb{N}_{>0}$, it holds that $m \leq n$.

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

Base case: If $t(x_1) \coloneqq f^m(x_1)$ for a unary function symbol f and $m \in \mathbb{N}$, we use the formula constructed in the proof of Theorem 3. When setting $\mathbf{x}$ to the empty tuple, it can easily be verified that it is in the correct form.

Inductive step: Assume that $t(x_1) \coloneqq g^m(s_k(\dots(s_2(s_1(x_1)))))$ for a unary function symbol g, $m \in \mathbb{N}$ and terms s_1', s_2', …, s. For readability we define the term $s_{\text{total}}(x_1) \coloneqq s_k(\dots(s_2(s_1(x_1))))$.

By the induction hypothesis, we have a formula $\vartheta_{s_{\text{total}}}(x_1, x_2) = \exists^{\geq 1}\mathbf{x}. \bigvee_i \zeta_i(x_1, x_2, \mathbf{x})$, where ζ_i is of the form

$$\begin{aligned} & s'_{i,1}(x_1) = x_{s_1(x_1)} \wedge \bigwedge \Psi_{i,1}(x_1) \\ \wedge & s'_{i,2}(x_{s_1(x_1)}) = x_{s_2(s_1(x_1))} \wedge \bigwedge \Psi_{i,2}(x_{s_1(x_1)}) \\ \wedge & \dots \\ \wedge & s'_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) = x_2 \wedge \bigwedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) \end{aligned}$$

with $\mathfrak{A}, a_1, a_2 \models s_{\text{total}}(x_1) = x_2$ if, and only if, $\mathfrak{A}, a_1, a_2 \models \vartheta_{s_{\text{total}}}(x_1, x_2)$.

Furthermore, we get $\mathfrak{A}, a_1, a_2 \models s_{\text{total}}(x_1) = x_2$ if, and only if, $\mathfrak{A}, a_1, a_2, \mathbf{a} \models \zeta_i(x_1, x_2, \mathbf{x})$ for an $i \in [k]$ and $\mathbf{a} \in A^k$. Thus it follows that

$$\mathfrak{A}, a_1, a_2 \models s_{\text{total}}(y_1) = y_2 \text{ if, and only if, } \mathfrak{A}, a_1, a_2, \mathbf{a} \models \zeta_i(x_1, x_2, \mathbf{x}). \tag{1}$$

If $m \leq n$, we set

$$\vartheta_t(x_1, x_2) = \exists^{\geq 1}\mathbf{x}, x_{s_{\text{total}}(x_1)}. \bigvee_i \zeta_i(x_1, x_2, \mathbf{x}, x_{s_{\text{total}}})$$

and ζ_i is of the form

$$\begin{aligned} & s'_{i,1}(x_1) = x_{s_1(x_1)} && \wedge \bigwedge \Psi_{i,1}(x_1) \\ \wedge & s'_{i,2}(x_{s_1(x_1)}) = x_{s_2(s_1(x_1))} && \wedge \bigwedge \Psi_{i,2}(x_{s_1(x_1)}) \\ \wedge & \dots \\ \wedge & s'_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) = x_{s_{\text{total}}(x_1)} && \wedge \bigwedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) \\ \wedge & g^m(x_{s_{\text{total}}(x_1)}) = x_2 && \wedge \bigwedge \Psi_{i,k+1}(x_{s_{\text{total}}(x_1)}), \end{aligned}$$

where $\Psi_{i,k+1}(x_{s_{\text{total}}(x_1)}) = \{\top\}$.

Now assume $m > n$. Then we set

$$\vartheta_t(x_1, x_2) = \exists^{\geq 1}\mathbf{x}, x_{s_{\text{total}}(x_1)}. \bigvee_{(k,\ell,p) \in \mathcal{I}(n,m)} \bigvee_i \zeta_{(k,\ell,p),i}(x_1, x_2, \mathbf{x}, x_{s_{\text{total}}})$$

and $\zeta_{(k,\ell,p),i}$ is of the form

$$\begin{aligned} & s'_{i,1}(x_1) = x_{s_1(x_1)} && \wedge \bigwedge \Psi_{i,1}(x_1) \\ \wedge & s'_{i,2}(x_{s_1(x_1)}) = x_{s_2(s_1(x_1))} && \wedge \bigwedge \Psi_{i,2}(x_{s_1(x_1)}) \\ \wedge & \dots \\ \wedge & s'_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) = x_{s_{\text{total}}(x_1)} && \wedge \bigwedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1)))}) \\ \wedge & g^{k+p}(x_{s_{\text{total}}(x_1)}) = x_2 && \wedge \bigwedge \Psi_{i,k+1}(x_{s_{\text{total}}(x_1)}), \end{aligned}$$

where

$$\begin{aligned} \Psi_{i,k+1}(x_{s_{\text{total}}(x_1)}) = \{ & g^k(x_{s_{\text{total}}(x_1)}) = g^{k+\ell}(x_{s_{\text{total}}(x_1)}), \\ & \mathrm{E}_g^{k,\ell}(x_{s_{\text{total}}(x_1)}), \\ & \bigwedge_{\ell' < \ell} g^k(x_{s_{\text{total}}(x_1)}) \neq g^{k+\ell'}(x_{s_{\text{total}}(x_1)})\}. \end{aligned}$$

We now proof 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 a ζ_i, $\mathbf{a} \in A^k$ and $a_{s_{\text{total}}} \in A$, such that $\mathfrak{A}, a_1, a_2, \mathbf{a}, a_{s_{\text{total}}} \models \zeta_i(x_1, x_2, \mathbf{a}, x_{s_{\text{total}}})$. We then get

$$\begin{aligned}
\mathfrak{A}, a_1, a_2, \mathbf{a}, a_{s_{\text{total}}} \models & s'_{i,1}(x_1) = x_{s_1(x_1)} \wedge \Psi_{i,1}(x_1) \\
& \wedge \dots \\
& \wedge s'_{i,k}(x_{s_{k-1}(\dots(s_1(x_1))\dots)}) = x_{s_{\text{total}}} \wedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1))\dots)}) \\
& \wedge g^m(x_{s_{\text{total}}}) = x_2 \\
\overset{\text{Equation (1)}}{\Leftrightarrow} \mathfrak{A}, a_1, a_2, a_{s_{\text{total}}} \models & s_{\text{total}}(x_1) = x_{s_{\text{total}}} \wedge g^m(x_{s_{\text{total}}}) = x_2 \\
\Leftrightarrow \mathfrak{A}, a_1, a_2 \models & g^m(s_{s_{\text{total}}}(x_1)) = x_2.
\end{aligned}$$

Now let $m > n$. Then there is a $\zeta_{(k,\ell,p),i}$, $\mathbf{a} \in A^k$ and $a_{s_{\text{total}}} \in A$, such that

$$\mathfrak{A}, a_1, a_2, \mathbf{a}, a_{s_{\text{total}}} \models \zeta_{(k,\ell,p),i}(x_1, x_2, \mathbf{a}, x_{s_{\text{total}}}).$$

And now

$$\begin{aligned}
\mathfrak{A}, a_1, a_2, \mathbf{a}, a_{s_{\text{total}}} \models & s'_{i,1}(x_1) = x_{s_1(x_1)} \wedge \Psi_{i,1}(x_1) \\
& \wedge \dots \\
& \wedge s'_{i,k}(x_{s_{k-1}(\dots(s_1(x_1))\dots)}) = x_{s_{\text{total}}} \wedge \Psi_{i,k}(x_{s_{k-1}(\dots(s_1(x_1))\dots)}) \\
& \wedge g^{k+p}(x_{s_{\text{total}}}) = x_2 \wedge \Psi_{i,k+1}(x_{s_{\text{total}}}) \\
\overset{\text{Equation (1)}}{\Leftrightarrow} \mathfrak{A}, a_1, a_2, a_{s_{\text{total}}} \models & s_{\text{total}}(x_1) = x_{s_{\text{total}}} \wedge g^{k+p}(x_{s_{\text{total}}}) = x_2 \wedge \Psi_{i,k+1}(x_{s_{\text{total}}}) \\
\overset{\text{Theorem 3}}{\Leftrightarrow} \mathfrak{A}, a_1, a_2, a_{s_{\text{total}}} \models & s_{\text{total}}(x_1) = x_{s_{\text{total}}} \wedge g^m(x_{s_{\text{total}}}) = x_2 \\
\Leftrightarrow \mathfrak{A}, a_1, a_2 \models & g^m(s_{s_{\text{total}}}(x_1)) = x_2.
\end{aligned}$$

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.

Therefore we have finished the proof. □

It can easily be seen that a variable $x_{t(x_1)}$ for a term t that gets introduced in the above proof, always has to correspond with its suffix. Formally, this means that $\mathfrak{A}, \mathfrak{I} \models \bigvee_i \zeta_i(x_1, x_2, \mathbf{x})$ implies that $\mathfrak{A}, \mathfrak{I} \models x_{t(x_1)} = t(x_1)$. Therefore, there is always exactly one value for a variable $x_{t(x_1)}$.

As a convention, we will abbreviate the $\zeta_{(k,\ell,p),i}$ formulae as $\zeta_{(k,\ell,p),i} \coloneqq \bigwedge \bar{\Phi}_{i,(k,\ell,p)}(x_1, x_2, \mathbf{x}) \wedge \bigwedge \bar{\Psi}_{i,(k,\ell,p)}(x_1, \mathbf{x})$, where $\bar{\Phi}_{i,(k,\ell,p)}$ contains the term equations $s'_{i,1}(x_1) = x_{s_1(x_1)}$, $s'_{i,2}(x_{s_1(x_1)}) = x_{s_2(s_1(x_1))}$ and so on, and $\bar{\Psi}_{i,(k,\ell,p)}$ contains the sets $\Psi_{i,j}$.

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

Corollary 5. *Let $\psi(y_1, \dots, y_m) \coloneqq R(t_1(y_1), \dots, t_m(y_m))$ be an atomic formula. Then there exists a formula $\vartheta_\psi \in \mathsf{GF}(\mathsf{C})$, such that for any given structure (of fitting signature) $\mathfrak{A}$ with $\|\mathfrak{A}\| = n$ it holds*

$$\mathfrak{A}, a_1, \dots, a_m \models \psi(y_1, \dots, y_m) \text{ if, and only if, } \mathfrak{A}, a_1, \dots, a_m \models \vartheta_\psi(y_1, \dots, y_m).$$

Furthermore, $\vartheta_\psi(y_1, \dots, y_m)$ is of the form $\exists^{\geq 1}\mathbf{x}, \mathbf{z}. \bigvee \Phi(y_1, \dots, y_m, \mathbf{x})$ where all $\varphi \in \Phi$ are of the form

$$R(z_1, \dots, z_m) \wedge \bigwedge \bar{\Phi}_1(y_1, z_1, \mathbf{x}) \wedge \bigwedge \bar{\Psi}_1(y_1, \mathbf{x}) \wedge \dots \wedge \bigwedge \bar{\Phi}_m(y_m, z_m, \mathbf{x}) \wedge \bigwedge \Psi_m(y_m, \mathbf{x}),$$

and for every $f^m(x)$ that appear in ϑ_ψ, where f is a unary function symbol, it holds that $m \leq n$.

Proof. Let $\mathfrak{A}, a_1, \ldots, a_m \models \psi(y_1, \ldots, y_m)$. This is equivalent to

$$\mathfrak{A}, a_1, \ldots, a_m \models \exists^{\geq 1}\mathbf{z}.\, (R(z_1, \ldots, z_m) \wedge t_1(y_1) = z_1 \wedge \cdots \wedge t_m(y_m) = z_m)$$

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

$$\begin{aligned}\mathfrak{A}, a_1, \ldots, a_m \models \exists^{\geq 1}\mathbf{z}.(R(z_1, \ldots, z_m) &\wedge \exists^{\geq 1}\mathbf{x_1}. \bigvee_{i_1} \Big(\bigwedge \bar{\Phi}_{1,i_1}(y_1, z_1, \mathbf{x_1}) \wedge \bigwedge \bar{\Psi}_{1,i_1}(y_1, \mathbf{x_1}) \Big) \\ &\wedge \ldots \\ &\wedge \exists^{\geq 1}\mathbf{x_m}. \bigvee_{i_m} \Big(\bigwedge \bar{\Phi}_{m,i_m}(y_m, z_m, \mathbf{x_m}) \wedge \bigwedge \bar{\Psi}_{m,i_m}(y_1, \mathbf{x_m}) \Big)).\end{aligned}$$

Through distribution of boolean formulae and because there is always exactly one value for the quantified variables, we can derive the equivalent statement

$$\begin{aligned}\mathfrak{A}, a_1, \ldots, a_m \models \exists^{\geq 1}\mathbf{x}.\exists^{\geq 1}\mathbf{z}.(\bigvee_{i_1} \cdots \bigvee_{i_m} (R(z_1, \ldots, z_m) &\wedge \bigwedge \bar{\Phi}_{1,i_1}(y_1, z_1, \mathbf{x_1}) \wedge \bigwedge \bar{\Psi}_{1,i_1}(y_1, \mathbf{x_1}) \\ &\wedge \ldots \\ &\wedge \bigwedge \bar{\Phi}_{m,i_m}(y_m, z_m, \mathbf{x_m}) \wedge \bigwedge \bar{\Psi}_{m,i_m}(y_1, \mathbf{x_m}))).\end{aligned}$$

We will define this last formula as ϑ_ψ and one can see that this formula is of the correct form. The equality follows from the fact that only equivalences have been used to derive ϑ_ψ from ψ. $\square$

To obtain our characterising result for structures with (unary) functions, we have to define how the functions should be encoded.

Definition 6 (Transitive Expansion)**.** Let $\sigma := \sigma_{\mathrm{Rel}} \,\dot\cup\, \sigma_{\mathrm{Func}}$ 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 $\|\mathfrak{A}\| = n$.

We define the transitive expansion of $\mathfrak{A}$ as a structure $\widetilde{\mathfrak{A}}$ of signature $\widetilde{\sigma}$, where

$$\widetilde{\sigma} := \sigma_{\mathrm{Rel}} \,\dot\cup \{F_\alpha : \alpha = f^i \text{ for } f \in \sigma_{\mathrm{Func}} \text{ and } i \in \{0\} \cup [n]\}$$

and the F_α are binary relations. Semantically, we have

$$F_\alpha^{\widetilde{\mathfrak{A}}} := \{(a, b) : \alpha^{\mathfrak{A}}(a) = b\}.$$

We now can define the algorithm for relational colour refinement for (unary) functions.

Definition 7 (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 functional-RCR (f-RCR), if $\|\mathfrak{A}\| \neq \|\mathfrak{B}\|$ or the transitive expansions, $\widetilde{\mathfrak{A}}$ and $\widetilde{\mathfrak{B}}$, are being distinguished by RCR.

To show that this definition may be sensible, we want to execute f-RCR on the structures $\mathfrak{A}$ and $\mathfrak{B}$ from Figure 1. First we compute $\widetilde{\sigma}$ as $\{F_{f^i} : 0 \leq i \leq 6\}$ and by performing the translation we obtain:

$$\begin{aligned}\widetilde{\mathfrak{A}} = (A, F_{f^0}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_1), (a_2, a_2), (a_3, a_3), (a_4, a_4), (a_5, a_5), (a_6, a_6)\}, \\ F_{f^1}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_3), (a_2, a_1), (a_3, a_2), (a_4, a_5), (a_5, a_6), (a_6, a_4)\}, \\ F_{f^2}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_2), (a_2, a_3), (a_3, a_1), (a_4, a_6), (a_5, a_4), (a_6, a_5)\}, \\ F_{f^3}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_1), (a_2, a_2), (a_3, a_3), (a_4, a_4), (a_5, a_5), (a_6, a_6)\}, \\ F_{f^4}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_3), (a_2, a_1), (a_3, a_2), (a_4, a_5), (a_5, a_6), (a_6, a_4)\}, \\ F_{f^5}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_2), (a_2, a_3), (a_3, a_1), (a_4, a_6), (a_5, a_4), (a_6, a_5)\}, \\ F_{f^6}^{\widetilde{\mathfrak{A}}} &= \{(a_1, a_1), (a_2, a_2), (a_3, a_3), (a_4, a_4), (a_5, a_5), (a_6, a_6)\})\end{aligned}$$

and

$$\begin{aligned}
\widetilde{\mathfrak{B}} = (B, F_{f^0}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_1),(a_2,a_2),(a_3,a_3),(a_4,a_4),(a_5,a_5),(a_6,a_6)\},\\
F_{f^1}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_3),(a_2,a_1),(a_3,a_5),(a_4,a_2),(a_5,a_6),(a_6,a_4)\},\\
F_{f^2}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_5),(a_2,a_3),(a_3,a_6),(a_4,a_1),(a_5,a_4),(a_6,a_2)\},\\
F_{f^3}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_6),(a_2,a_5),(a_3,a_4),(a_4,a_3),(a_5,a_2),(a_6,a_1)\},\\
F_{f^4}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_4),(a_2,a_6),(a_3,a_2),(a_4,a_5),(a_5,a_1),(a_6,a_3)\},\\
F_{f^5}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_2),(a_2,a_4),(a_3,a_1),(a_4,a_6),(a_5,a_3),(a_6,a_5)\},\\
F_{f^6}^{\widetilde{\mathfrak{B}}} &= \{(a_1,a_1),(a_2,a_2),(a_3,a_3),(a_4,a_4),(a_5,a_5),(a_6,a_6)\})
\end{aligned}$$

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

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

is satisfied by $\widetilde{\mathfrak{A}}$, but not $\widetilde{\mathfrak{B}}$.

We see, that this procedure distinguishes structures, that were not distinguished by nRCR. To formalise this, we want to characterise this algorithm logically, as well as combinatorially.

3.2.1 Logical characterisation of f-RCR

Theorem 8. *Let $\mathfrak{A}$ and $\mathfrak{B}$ be two structures of the same signature σ with relation and unary function symbols. The two following statements are equivalent:*

1. f-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 $\mathfrak{B} \not\models \varphi$.*

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 [1], 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 $\varphi \in \mathsf{GF}(\mathsf{C})$ of signature σ.

For every atomic subformula $R(\mathbf{x})$, where $R \in \sigma_{\mathrm{Rel}}$, let the formula stay the same. For every atomic subformla $F_\alpha(x,y)$, where $\alpha \in \mathrm{Alters}_n^k(\sigma)$, replace it by the formula $\alpha(x) = y$. Obviously, if a structure's expansion satisfied $\widetilde{\varphi}$, it also satisfies φ and vice versa.

Therefore, we get a formula $\varphi \in \mathsf{GF}(\mathsf{C})$ of signature "σ that distinguishes $\mathfrak{A}$ and $\mathfrak{B}$. Now we prove that *2.* implies *1.*. Let $\varphi \in \mathsf{GF}(\mathsf{C})$ such that $\mathfrak{A} \models \varphi$ and $\mathfrak{B} \not\models \varphi$. Our approach will be, to transform φ to a formula $\widetilde{\varphi}$ that only uses symbols from $\widetilde{\sigma}$. This transformation will result in a formula that is not in $\mathsf{GF}(\mathsf{C})$. However, *das sollte immer noch gehen, ich bin mir nur noch nicht sicher wie. Muss ich mit Moritz besprechen*

Using Theorem 5 we can obtain a formula ϑ_ψ for every atomic subformula ψ of φ with $\mathfrak{A} \models \psi$ if, and only if, $\mathfrak{A} \models \vartheta_\psi$. Now replace every subformula ψ in φ with this newly constructed formula. This yields us $\varphi' \in \mathsf{C}$.

Claim 9. *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 replace φ with ϑ_φ. The equivalence follows directly from the above lemmas.

Inductive cases: In the cases where φ is of the form $\neg\vartheta$ and $\vartheta_1 \wedge \vartheta_2$, 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 respectively, we also want to distribute them to allow for easier definitions in the following proofs. As such, we want to translate φ to $\varphi' := \exists^{\geq \ell}\mathbf{v}. \bigvee (\Delta' \wedge \bigwedge \Psi \wedge \vartheta')$.

In the following we prove the equivalence of these two formulae. 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 indsuction 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. □

Note that for every term α that appears in φ' it is either a variable or there is a $F_\alpha \in \widetilde{\sigma}$. This follows from the properties of the translation in Theorem 5. Furthermore, for guards we have to distribute the formulae a bit more.

Now we can transform φ' to a formula of signature $\widetilde{\sigma}$. We will define this transformation inductively over the structure of the formula ψ. Furthermore, we use this induction to prove that this does not change, whether a structure satisfies the formula.

Claim 10. *It holds that $\mathfrak{A} \models \varphi'$ if, and only if, $\widetilde{\mathfrak{A}} \models \widetilde{\varphi}$.*

Proof. **Base cases:** If ψ is of the form $t(x) = y$ for a term t, then there exists a relation F_t in $\widetilde{\sigma}$. Therefore we set the transformed formula to $F_t(x, y)$.

If ψ is of the form $R(t_1(x_1), \ldots, t_m(x_m))$, then there are relations F_{t_i} for $1 \leq i \leq m$ in $\widetilde{\sigma}$. Therefore we set the transformed formula to $R(y_1, \ldots, y_m) \wedge F_{t_1}(x_1, y_1) \wedge \cdots \wedge F_{t_m}(x_m, y_m)$.

The claim obviously follows from the definition of the transitive expansion.

Inductive cases: Given the formulae ϑ_1 and ϑ_2, as well as their transformed forms $\widetilde{\vartheta_1}$ and $\widetilde{\vartheta_2}$, which fulfil the above claim.

We then have the following translations, for which the claim can easily be shown using the induction hypothesis.

- $\neg\vartheta_1$ to $\neg\widetilde{\vartheta_1}$ and
- $\vartheta_1 \wedge \vartheta_2$ to $\widetilde{\vartheta_1} \wedge \widetilde{\vartheta_2}$.

The case for quantifiers is a bit more involved and will be described in the following.

Der Teil muss verändert werden, da sich meine Übersetzung in Theorem 5 auch verändert hat.

Let ψ be of the form $\exists^{\geq \ell}\mathbf{v}.\vartheta(\mathbf{v})$. Using the induction hypothesis, we already obtained a translation $\widetilde{\vartheta}(\mathbf{v})$ of $\vartheta(\mathbf{v})$. By considering the analogous case of the above claim, we see that ϑ must be of the form

$$\bigvee \left(\Delta(\mathbf{v}) \wedge \bigwedge \Psi(\mathbf{v}) \wedge \vartheta'(\mathbf{v})\right)$$

for an atomic formula Δ, a set of formulae Ψ and a formula ϑ'. However, to allow for an easier proof, our first goal will be, to quantify the variables for each subformula of the disjunction separately. More precisely, this means that instead of $\exists^{\geq \ell}\mathbf{v}. \bigvee(\ldots)$ we want our formula to be of the form $\bigvee(\exists^{\geq \ell}\mathbf{v}. \ldots)$.

Furthermore, by considering the translations from this proof, we see that $\widetilde{\vartheta}$ then must be of the form

$$\bigvee \left(R(\mathbf{u}) \wedge \bigwedge_{i \in [k]} R_i(v_i, u_i) \wedge \bigwedge \left\{ \widetilde{\zeta} : \zeta \in \Psi \right\} \wedge \widetilde{\vartheta'} \right),$$

where R and R_i for $i \in [k]$ are relation symbols and $k = 0$ in the case where Δ is an equality and $k = \mathrm{ar}(R)$ otherwise. $\square$

The resulting formula is a formula in the logic disjunctive-$\mathsf{GF}(\mathsf{C})$, which we will define in the following.

Die Definition muss sich analog zu dem Beweis oben ändern.

Definition 11 (disjunctive-$\mathsf{GF}(\mathsf{C})$)**.** The logic disjunctive-$\mathsf{GF}(\mathsf{C})$ is a syntactical extension of $\mathsf{GF}(\mathsf{C})$. As such it is defined by the rules given in **??** of $\mathsf{GF}(\mathsf{C})$ in addition to a sixth rule:

For two disjunct sets of variables $\mathbf{x} = \{x_1, \ldots, x_m\}$ and $\mathbf{y} = \{y_1, \ldots, y_m\}$, atomic formulae Δ_i with $\mathrm{free}(\Delta_i) \subseteq \mathbf{y}$, formulae ϑ_i with $\mathrm{free}(\vartheta_i) \subseteq \mathbf{x}$, binary relations R_j, $\ell \in \mathbb{N}$ and a set $\mathbf{x'} \subseteq \mathbf{x}$, the formula

$$\exists^{\geq \ell} \mathbf{x}'. \bigvee_i \left(\Delta(\mathbf{y}) \wedge \bigwedge_j R_j(x_j, y_j) \wedge \vartheta(\mathbf{x}) \right)$$

is a formula of disjunctive-$\mathsf{GF}(\mathsf{C})$.

Irgendwie sollte man daraus beweisen, dass RCR die Transitiven Erweiterungen trennt, bin mir aber nicht sicher wie genau.

$\square$

4 Relational Colour Refinement for symmetric structures

5 Conclusion

References

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