Relational Colour Refinement for Non-Relational Signatures

Theodor Jurij Teslia

RWTH Aachen University

September 5, 2025

Introduction

- Colour Refinement is an important and interesting algorithm
- Applied in modern isomorphism solvers
- Can be characterised logically and combinatorially
- Extension to more than graphs seems desirable
- Scheidt and Schweikardt, 2025 introduced Relational Colour Refinement
- Conceptually similar to classical Colour Refinement
- Also has a logical and a combinatorial characterisation

Contents of this presentation

Classical Colour Refinement

Colour Refinement

- Also called *CR* or 1-*dimensional Weisfeiler-Leman* algorithm
- Iterative graph algorithm
- Constructs colour for every vertex, based on colours of neighbours
- Initial colour for every $v \in V$: $C_0(v) = 0$
- Next rounds: $C_{i+1}(v) \coloneqq (C_i(v), \{\!\{ C_i(u) : \{v, u\} \in E \}\!\})$
- CR distinguishes two graphs G and H, if
  - there exists $C_i(v)$ in colouring of G or H, such that the number of vertices with colour $C_i(v)$ is different in G than in H

Characterisations of CR

- There are equivalent characterisations for CR
- Logical characterisation:
  CR distinguishes G and H if, and only if, there exists $\varphi \in \mathsf{C}_2$, such that $G \models \varphi$ and $H \not\models \varphi$
- Combinatorial characterisation:
  CR distinguishes G and H if, and only if, there exists tree T, such that $\hom(T, G) \neq \hom(T, H)$

Relational Colour Refinement

Relational Colour Refinement (RCR)

- Applies variant of classical Colour Refinement on tuples of structure
- Uses set of relations that contain tuple as part of initial colouring
- Uses pairs of indices as edges to mark shared elements of tuples
- Formally:

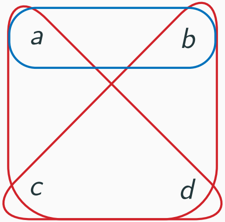
$$\mathrm{atp}(\mathbf{a}) = \{R \in \sigma : \mathbf{a} \in R\}$$

  and

$$\mathrm{stp}(\mathbf{a}, \mathbf{b}) = \{(i, j) \in [n] \times [m] : a_i = b_j\}$$

- For relational structure $\mathfrak{A}$ and all tuples $\mathbf{a} \in \mathbf{A}$:
- Initial colour: $\varrho_0(\mathbf{a}) = (\mathrm{atp}(\mathbf{a}), \mathrm{stp}(\mathbf{a}, \mathbf{a}))$
- For the next rounds: $\varrho_{i+1}(\mathbf{a}) = (\varrho_i(\mathbf{a}), \{\!\{(\mathrm{stp}(\mathbf{a}, \mathbf{b}), \varrho_i(\mathbf{b})) : \mathrm{stp}(\mathbf{a}, \mathbf{b}) \neq \emptyset\}\!\})$

An Example for RCR



- Structure $\mathfrak{A} = (A, R^{\mathfrak{A}}, T^{\mathfrak{A}})$

- $\varrho_0((a,b)) = (\{R\}, \{(1,1),(2,2)\})$ and
  $\varrho_0((a,c,d)) = \varrho_0((b,c,d)) = (\{T\}, \{(1,1),(2,2),(3,3)\})$
- $\varrho_1((a,c,d)) = (\varrho_0((a,c,d)), \{\!\{(\{(1,1)\}, \varrho_0((a,b))), \dots\}\!\})$ and
  $\varrho_1((b,c,d)) = (\varrho_0((b,c,d)), \{\!\{(\{(1,2)\}, \varrho_0((a,b))), \dots\}\!\})$

Distinguishing Relational Structures with RCR

- RCR distinguishes, if some colour appears differently often in the structures

- $\varrho_1((a, c, d))$ appears in colouring of left structure but not in right

Guarded Fragment of Counting Logic

- C_2 characterises CR on graphs
- Guarded fragment of counting logic GF(C) characterises RCR

Guarded Fragment of Counting Logic

- Everything except for quantifiers defined as in classical counting logic
- For atomic formula $\Delta \in \mathsf{GF}(\mathsf{C})$ and formula $\varphi \in \mathsf{GF}(\mathsf{C})$, we call Δ a guard for φ, if $\mathsf{free}(\Delta) \supseteq \mathsf{free}(\varphi)$
- Quantifiers appear only in form $\exists^{\geq i}\mathbf{v}\,.(\Delta \wedge \varphi)$, where Δ is guard for φ and $\mathsf{set}(\mathbf{v}) \subseteq \mathsf{free}(\Delta)$

- Examples:
  - $\exists^{\geq 2}(x, y)\,.(E(x, y) \wedge T(y)) \in \mathsf{GF}(\mathsf{C})$
  - $\exists^{\geq 3}(x, y, z)\,.(E(x, y) \wedge E(y, z) \wedge E(z, x)) \notin \mathsf{GF}(\mathsf{C})$

Characterising RCR Using Logic

Theorem B (Scheidt and Schweikardt, 2025)

Let $\mathfrak{A}$ and $\mathfrak{B}$ be two relational structures. Then the two following statements are equivalent.

1. RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$
2. There exists a sentence in GF(C) that is satisfied by $\mathfrak{A}$, but not by $\mathfrak{B}$

Example for Logical Characterisation of RCR

- We have seen RCR distinguishes the structures
- Formula $\exists^{\geq 1}(x, y, z).\,\big(T(x, y, z) \wedge \exists^{\geq 1}(y).\,(R(x, y))\big)$ satisfied by left and not by right structure

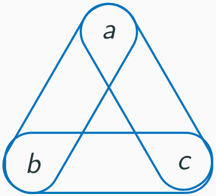
Acyclic Structures

- Counting homomorphisms from trees characterises CR on graphs
- Abstraction from trees to relational structures is needed: α-acyclic structures (in the following only acyclic structures)

Acyclic Structures

- Relational structure $\mathfrak{C}$ is acyclic if it has a join tree J
- Join tree J is tree with $V(J) = \bigcup_{R \in \sigma} R^{\mathfrak{C}}$ and fulfils join-tree-property:
  - For every $e \in C$, the set $\{\mathbf{x} \in V(J) : e \in \mathrm{set}(\mathbf{x})\}$ induces a connected subtree

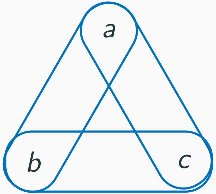
Examples for Acyclic Structures



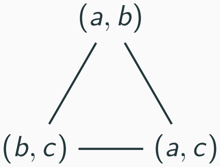
No:

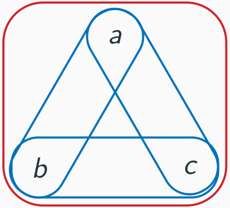
$$\begin{tikzcd}[column sep=small] & (a, b) \arrow[dl, dash] \arrow[dr, dash] & \\ (b, c) \arrow[rr, dash] & & (a, c) \end{tikzcd}$$

Examples for Acyclic Structures

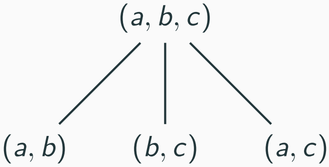


No:





Yes:



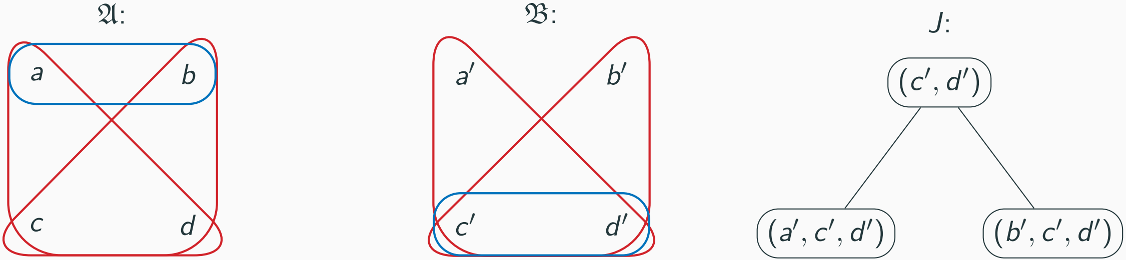
Characterising RCR Using Homomorphism Counting

Theorem A (Scheidt and Schweikardt, 2025)

Let $\mathfrak{A}$ and $\mathfrak{B}$ be relational structures. Then the two following statements are equivalent.

1. RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$
2. There exists an acyclic relational structure $\mathfrak{C}$, such that it distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ by homomorphism count

Example for Combinatorial Characterisation of RCR



- J is join tree for $\mathfrak{B}$, therefore $\mathfrak{B}$ is acyclic
- Identity is homomorphism, so $\mathfrak{B}$ has at least one homomorphism to itself
- $\mathfrak{B}$ has no homomorphisms to $\mathfrak{A}$

Relational Colour Refinement for Structures With Functions

Relational Colour Refinement for Structures With Functions

- Many interesting structures use functions
- Colour Refinement algorithm for such structures seems desirable
- Will use the results of Scheidt and Schweikardt and investigate how robust they are
- Following structure:
    1. Presentation of two approaches for Colour Refinement for non-relational signatures
    2. Logical characterisation of both approaches
    3. Discussion on combinatorial characterisation

Naive RCR

- Goal: Encode non-relational structures and signatures as relational ones
- Functions can directly be interpreted as relations:

$$f(\mathbf{x}) = y \iff (\mathbf{x}y) \in R_f$$

- For non-relational signature σ define relational signature σ':
  - Inherit relation symbols from σ
  - Function symbol $f \in \sigma$ of arity $n \to$ introduce $R_f \in \sigma'$ of arity $n+1$
- Encode σ-structure $\mathfrak{A}$ as σ'-structure $\mathfrak{A}'$:
  - Relations like in $\mathfrak{A}$
  - For function symbol $f \in \sigma$: $R_f^{\mathfrak{A}'} := \{(\mathbf{x}y) : f^{\mathfrak{A}}(\mathbf{x}) = y\}$
- We say *naive RCR* distinguishes $\mathfrak{A}$ and $\mathfrak{B}$, if RCR distinguishes the encodings

Idea of the Transitive Expansion

- Approach is only defined for unary function symbols
- Encoding emulates the nesting of function applications
- Encode function f as family of relations $R_{f^1}, R_{f^2}, \ldots,$ where $(x, y) \in R_{f^i}$ if $\underbrace{f(f(\ldots f(}_{i \text{ times}}x))) = y$
- In the following: $f^i(x)$ written for $\underbrace{f(f(\ldots f(}_{i \text{ times}}x)))$
- For multiple functions, also encode alternations, for example R_{fg} or $R_{g^2 f^3}$

Alternations of Function Applications

- Let σ be signature with unary function symbols
- Define set of all allowed function application alternations Alters_n^k as all sequences of up to k function symbols, where
    1. Every function symbol has exponent in $[n]$
    2. Two succeeding function symbols are different

- Example:
    - $\sigma = \{f/1, g/1\}$
    - $\mathrm{Alters}_2^2(\sigma) = \underbrace{\{\mathrm{id}\}}_{k=0} \cup \underbrace{\{f, f^2, g, g^2\}}_{k=1} \cup \underbrace{\{fg, fg^2, f^2g, f^2g^2, gf, \dots\}}_{k=2}$

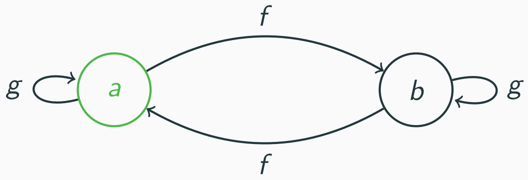
Transitive Expansion ii

Transitive Expansion

- For alternation depth k and σ-structure $\mathfrak{A}$ with $|\mathfrak{A}| = n$ define signature $\widetilde{\sigma}$ and transitive expansion $\widetilde{\mathfrak{A}}$ as $\widetilde{\sigma}$-structure
- For all $\alpha, \beta, \alpha_1, \dots, \alpha_\ell \in \mathsf{Alters}_n^k(\sigma)$ and relation symbol $R \in \sigma$ of arity ℓ, insert relation symbol $\mathsf{Eq}_{\alpha,\beta}$ of arity 2 and relation symbol $R_{\alpha_1,\dots,\alpha_\ell}$ of arity ℓ into $\widetilde{\sigma}$
- Define $\mathsf{Eq}_{\alpha,\beta}^{\widetilde{\mathfrak{A}}} := \{(x, y) : \alpha^{\mathfrak{A}}(x) = \beta^{\mathfrak{A}}(y)\}$ and $R_{\alpha_1,\dots,\alpha_\ell}^{\widetilde{\mathfrak{A}}} := \{(x_1, \dots, x_\ell) : (\alpha_1^{\mathfrak{A}}(x_1), \dots, \alpha_\ell^{\mathfrak{A}}(x_\ell)) \in R^{\mathfrak{A}}\}$

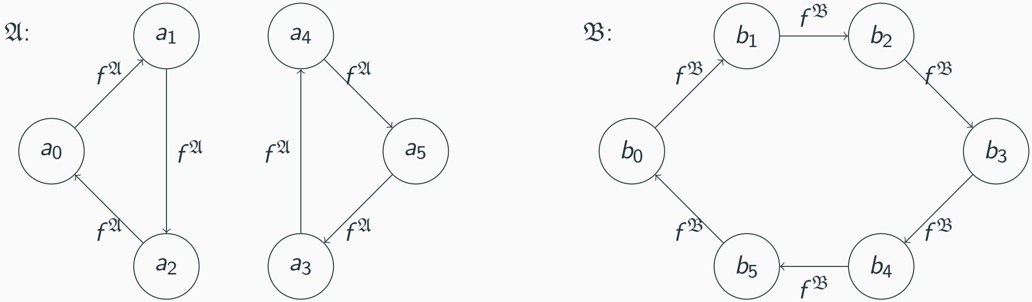
- For $k \in \mathbb{N}$ we say that RCR_k distinguishes structures $\mathfrak{A}$ and $\mathfrak{B}$, if RCR distinguishes the transitive expansions with alternation depth k

Example for the Transitive Expansion



- Structure $\mathfrak{A} = (A, R^{\mathfrak{A}}, f^{\mathfrak{A}}, g^{\mathfrak{A}})$
- $k = 1$ and $n = 2$: $\text{Alters}_2^1(\sigma) = \{\text{id}, f, f^2, g, g^2\}$
- $\widetilde{\sigma} = \{R_{\text{id}}, R_f, R_{f^2}, R_g, R_{g^2}, \text{Eq}_{\text{id},\text{id}}, \text{Eq}_{\text{id},f}, \text{Eq}_{\text{id},f^2}, \ldots, \text{Eq}_{g^2,g^2}\}$
- Examples:
  - $R_f^{\widetilde{\mathfrak{A}}} = \{b\}$
  - $\text{Eq}_{f^2,\text{id}}^{\widetilde{\mathfrak{A}}} = \{(a, a), (b, b)\}$
  - $\text{Eq}_{g,f}^{\widetilde{\mathfrak{A}}} = \{(a, b), (b, a)\}$

Naive Encoding versus Transitive Expansion



- Cannot be distinguishes by naive RCR: Encodings result in regular graphs
- But: Distinguished by Transitive Expansion Encoding
  - We find that $\mathrm{Eq}^{\widetilde{\mathfrak{A}}}_{f^1,\mathrm{id}} = \mathrm{Eq}^{\widetilde{\mathfrak{A}}}_{f^4,\mathrm{id}}$, not for $\widetilde{\mathfrak{B}}$
  - Sentence $\exists^{\geq 6}(x, y) \,.\, \big(\mathrm{Eq}_{f^1,\mathrm{id}}(x, y) \wedge \mathrm{Eq}_{f^4,\mathrm{id}}(x, y)\big) \in \mathrm{GF}(\mathrm{C})$ distinguishes encodings

Relational Colour Refinement for Structures With Functions

Logical Characterisations for Both Approaches

Nesting-Free Guarded Fragment of Counting Logic

nfGF(C)

- Extends given definition of GF(C) for non-relational signatures
- Allow atomics of the following forms
  - Relation symbols and variable equations like in GF(C)
  - For function symbol f of arity ℓ and variables $x_1, \ldots, x_\ell, y$:
    $f(x_1, \ldots, x_\ell) = y \in \text{nfGF(C)}$

- Forbid nesting of terms, for example $f(g(x), y) = z$
- Informally: Usage of function symbols like relation symbols

Characterising Naive RCR Logically

Logical Characterisation of Naive RCR

Let $\mathfrak{A}$ and $\mathfrak{B}$ be structures.

Naive RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$
iff.
There exists a sentence $\varphi \in \text{nfGF(C)}$ which is fulfilled by $\mathfrak{A}$, but not by $\mathfrak{B}$

Proof idea:

- Naive RCR distinguishes structures iff. RCR distinguishes encodings iff. there exists a sentence in GF(C) that distinguishes the encodings
- Define translation of GF(C) to and from nfGF(C)
  - Replace $R_f(\mathbf{x}y)$ by $f(\mathbf{x}) = y$

GF(C) **with alternation depth** k **(**$\mathrm{GF(C)}_k$**)**

GF(C) **with alternation depth** k

- Fixate $k \in \mathbb{N}$
- Atomics are defined like in natural extension to non-relational signatures, with one restriction
- For every formula in $\mathrm{GF(C)}_k$ and every term t that appears in it, there must exist a $n \in \mathbb{N}$, such that $t = \alpha$ for a $\alpha \in \mathrm{Alters}_n^k(\sigma)$

- Restrict number of alternations of function applications to k
- No restriction of number of application of same function in series
- Examples:
  - $f^2(g(h^3(x))) = y \notin \mathrm{GF(C)}_2$, but in $\mathrm{GF(C)}_3$
  - $f^i(x) = y \in \mathrm{GF(C)}_1$ for all $i \in \mathbb{N}$

Characterising RCR_k Logically

Logical Characterisation of RCR_k

Let $k \in \mathbb{N}$ and let $\mathfrak{A}$ and $\mathfrak{B}$ be two structures.

RCR_k distinguishes $\mathfrak{A}$ and $\mathfrak{B}$
iff.
There exists a sentence in $\mathrm{GF(C)}_k$ that is fulfilled by $\mathfrak{A}$, but not by $\mathfrak{B}$

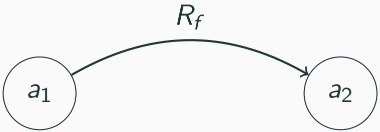
- 1. to 2.: Similar to before, translation from GF(C) to $\mathrm{GF(C)}_k$ very simple
- 2. to 1.:
  - Assume $n = |\mathfrak{A}| = |\mathfrak{B}|$
  - Translate and replace all atomic subformulae by formula that:
    - is equivalent for all structures with n elements
    - only contains terms $f^i(s(x))$ with $i \leq n$
  - Rearrange resulting formula to get valid $\mathrm{GF(C)}_k$-sentence
  - Results in equivalent formula for structures with n elements and for every term t there exists an $\alpha \in \mathrm{Alters}_n^k(\sigma)$, such that $t = \alpha$
  - Can easily be translated into sentence in GF(C) of signature $\widetilde{\sigma}$

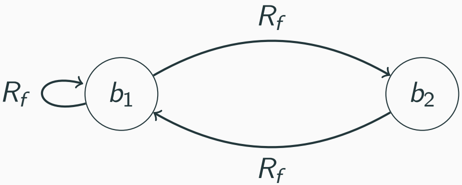
Relational Colour Refinement for Structures With Functions

Discussion on the Combinatorial Characterisation

Total and Functional Structures

- Let σ be a signature, σ' its naive encoding and $\mathfrak{A}'$ a σ'-structure
- We call $\mathfrak{A}'$ total if for every function symbol $f \in \sigma$ and every tuple $\mathbf{x}$ there is a y, such that $(\mathbf{x}y) \in R_f^{\mathfrak{A}'}$
- We call $\mathfrak{A}'$ functional if for every function symbol $f \in \sigma$ there are no two tuples $(\mathbf{x}y), (\mathbf{x}z) \in R_f^{\mathfrak{A}'}$





Non-Relational Acyclic Structures

- Will define acyclicity w.r.t. the naive encoding

Non-Relational Acyclic Structures

- Let $\mathfrak{A}$ be a non-relational structure
- We call $\mathfrak{A}$ acyclic, if its naive encoding $\mathfrak{A}'$ is acyclic

Total and Functional Structures as Encodings

- Desired equivalence:

Non-relational, acyclic structure distinguishes $\mathfrak{A}$ and $\mathfrak{B}$ by homomorphism count
?
Naive RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$

- Result: Forward direction holds, backwards does not
- First step: Reformulate first statement:

Some non-relational, acyclic structure dist. $\mathfrak{A}$ and $\mathfrak{B}$ by hom. count
iff.
Some total, functional and acyclic structure dist. encodings $\mathfrak{A}'$ and $\mathfrak{B}'$ by hom. count

Enforcing Functionality

- We can show:

$$\begin{gathered}\text{Acyclic } \sigma'\text{-structure dist. } \mathfrak{A}' \text{ and } \mathfrak{B}' \text{ by hom. count}\\ \text{iff.}\\ \text{Functional and acyclic } \sigma'\text{-structure dist. } \mathfrak{A}' \text{ and } \mathfrak{B}' \text{ by hom. count}\end{gathered}$$

Proof idea:

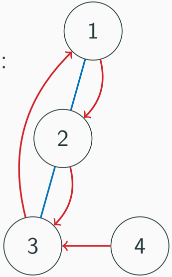
- Backwards direction is obvious
- Forwards direction eliminates collisions of the form $(\mathbf{x}y), (\mathbf{x}z) \in R_f$ by contracting y and z
- This can be done while maintaining the homomorphisms and acyclicity and can be repeated until no collisions remain

Non-Enforceability of Totality

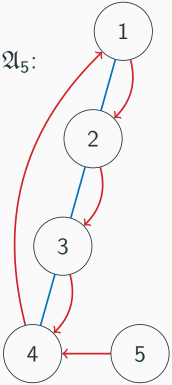
- There are structures that are distinguished by naive RCR, but there is no acyclic and total structure that distinguishes the encodings by homomorphism count
- Define signature $\sigma = \{E/2, f/1\}$
- Two families of σ-structures $(\mathfrak{A}_i)_{i \in \mathbb{N}_{\geq 4}}$ and $(\mathfrak{B}_i)_{i \in \mathbb{N}_{\geq 4}}$
- For all $i \in \mathbb{N}_{\geq 4}$: Naive RCR distinguishes $\mathfrak{A}_i$ and $\mathfrak{B}_i$, but no total and acyclic structure can distinguish the encodings by hom. count

Non-Enforceability of Totality ii

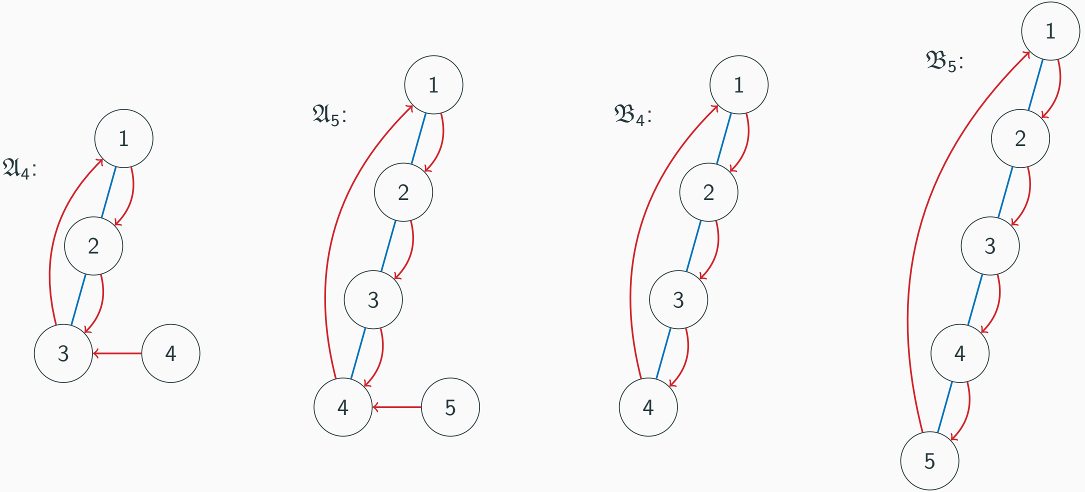
$\mathfrak{A}_4$:



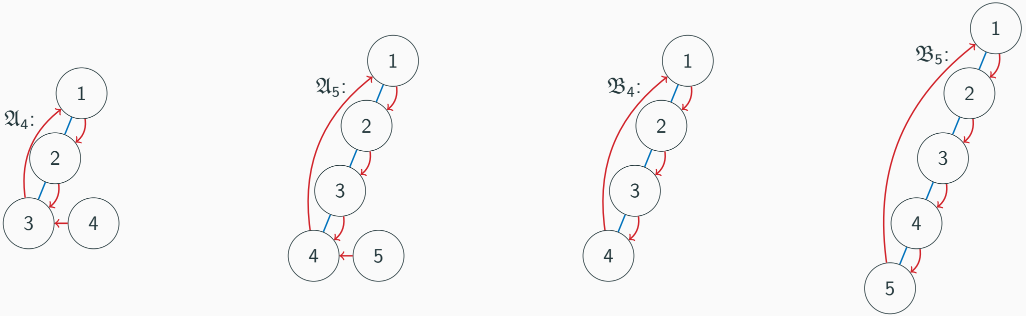
$\mathfrak{A}_5$:



Non-Enforceability of Totality ii



Non-Enforceability of Totality iii



- Obviously distinguished by naive RCR
- If structure has R_f-loops or R_f-2-cycles, then no homomorphisms to either structure
- Because total, it has to contain larger R_f-cycles, but then cannot be acyclic

Results of combinatorial characterisation of naive RCR

We have the following results:

Naive RCR distinguishes $\mathfrak{A}$ and $\mathfrak{B}$

$\Updownarrow$

There exists acyclic structure that dist. encodings $\mathfrak{A}'$ and $\mathfrak{B}'$ by hom. count

$\Updownarrow$

There exists *functional* and acyclic structure that dist. encodings by hom. count

$\Uparrow$, but $\not\Downarrow$

There exists *total*, functional and acyclic structure that dist. encodings by hom. count

$\Updownarrow$

There exists *non-relational* and acyclic structure that dist. $\mathfrak{A}$ and $\mathfrak{B}$ by hom. count

Restricting RCR to Subclasses of Relational Structures

Restricting the Class of Structures

- For what subclass $\mathcal{S}$ of relational structures do we have the following equivalence:

  Two structures from $\mathcal{S}$ get distinguished by RCR
  iff.
  There exists an acyclic structure from $\mathcal{S}$ that dist. the structures by hom. count
- Does not hold for class of total structures
  - Encodings of families of structures from before are total, but no total and acyclic structure dist. them by hom. count
- Another class to investigate: Class of symmetric structures

Restriction to Symmetric Structures

- Relational Structure is symmetric, if for every relation R and for every tuple $\mathbf{x} \in R$, every permutation of the elements in $\mathbf{x}$ is also in R

Restriction to Symmetric Structures

- Relational Structure is symmetric, if for every relation R and for every tuple $\mathbf{x} \in R$, every permutation of the elements in $\mathbf{x}$ is also in R
- For two symmetric structures we can show

  There exists acyclic structure that dist. the structures by hom. count

  iff.

  There exists symmetric and acyclic structure that dist. the structure by hom. count
- From this, restriction to symmetric structures is possible

Conclusion

Conclusion

- We presented classical CR and Scheidt's and Schweikardt's RCR algorithm
- We defined two possible ways to apply their algorithm to non-relational signatures
- We showed a logical characterisation for each of the approaches
- We disproved the characterisation by homomorphism counting
- We showed results for the restriction to two subclasses of the relational structures