Tree-Representations of Binary Relations

Marc Hellmuth

Dptm. Mathematics and Computer Science
University of Greifswald, Germany

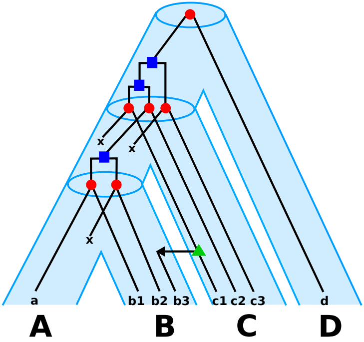
Joint work with Nic Wieseke and Peter F. Stadler

TBI WINTERSEMINAR 14-21. FEB. 2016

Outline

1. Motivation
2. Tree-Representation of
    - one symmetric relation
    - one non-symmetric relation
    - sets of symmetric relations
    - sets of non-symmetric relations
      (2-structures, Di-cographs and Symbolic Ultrametrics)

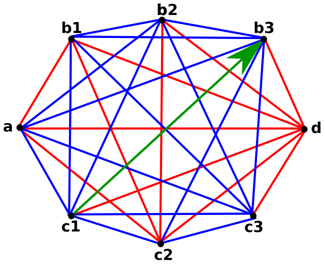
Motivation

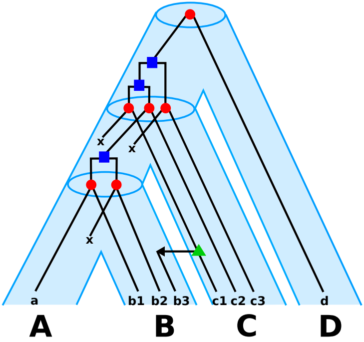


An ordered pair (x, y) of two genes is

- "lca"-orthologs if $\text{lca}(x, y) =$ ● $=$*speciation*
- "lca"-paralogs if $\text{lca}(x, y) =$ ■ $=$*duplication*
- "lca"-xenologs if $\text{lca}(x, y) =$ ▲ $=$*HGT* and ▲ "points from" x to y in T

Motivation

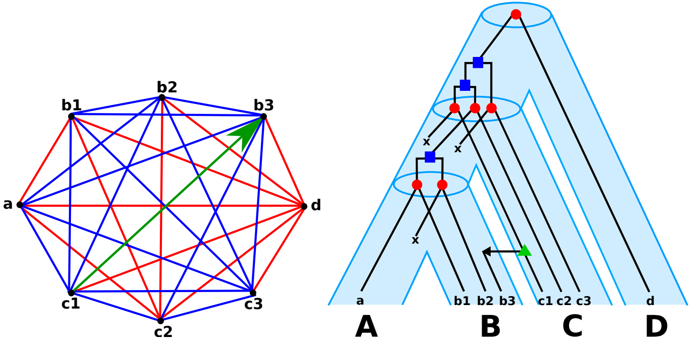




The gene-tree determines three distinct relations

- $R_{\bullet}$, the “lca”-orthologs ($\mathrm{lca}(x, y) = \bullet$)
- $R_{\blacksquare}$, the “lca”-paralogs ($\mathrm{lca}(x, y) = \blacksquare$)
- $R_{\blacktriangle}$, the “lca”-xenologs ($\mathrm{lca}(x, y) = \blacktriangle$, $\blacktriangle$ “points from” x to y in T)

Motivation



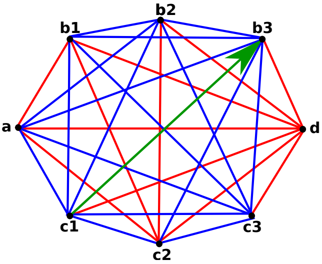
Orthologs can be estimated without inferring a gene- or species trees.

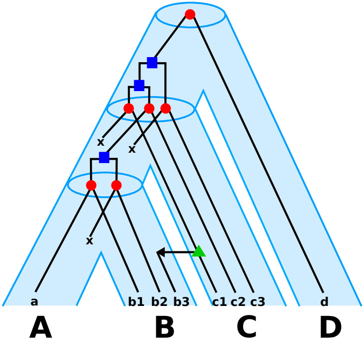
Assume we have *estimated* binary relations $R_1, \dots, R_k$ s.t.

$$(xy) \in R_i \text{ iff } \mathrm{lca}(xy) = i \text{ in ordered tree } T$$

Thus, it is important to understand, when those relations $R_1, \dots, R_k$ can be "represented" in a single tree.

Motivation





We consider irreflexive relations $(x, x) \notin R$ for all $x \in X$.

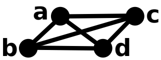
If both pairs $(x, y), (y, x) \in R$ we simply write $x - y \in R$

One binary relation

One symmetric relation R over X

Tree

Graph



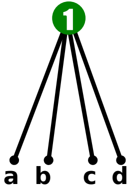
Set $R = X \times X_{|\mathrm{irr}}$

A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

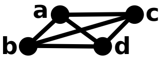
$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

One symmetric relation R over X

Tree



Graph



Set

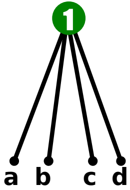
$R = X \times X_{|\mathrm{irr}}$

A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

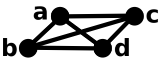
$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

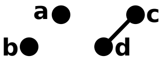
One symmetric relation R over X

Tree



Graph





Set

$R = X \times X_{|\mathrm{irr}}$ $R = \{c - d\}$

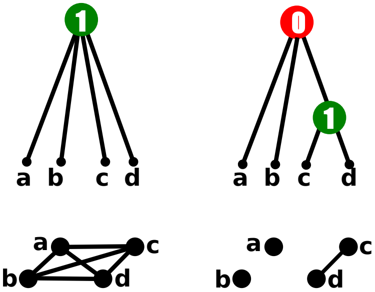
A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

One symmetric relation R over X

Tree

Graph



Set

$R = X \times X_{|\mathrm{irr}}$ $\qquad R = \{c - d\}$

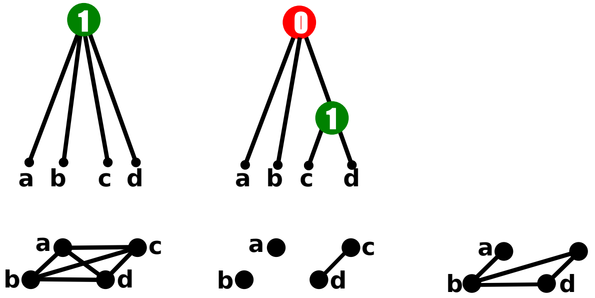
A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

One symmetric relation R over X

Tree

Graph

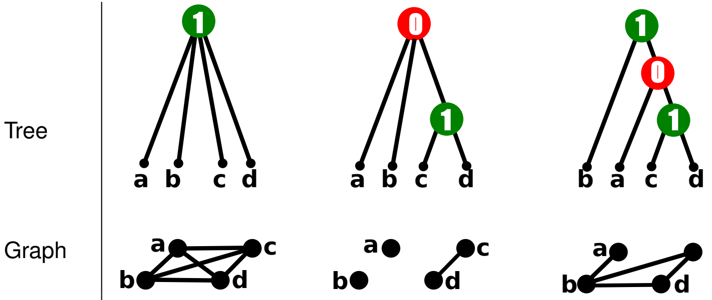


Set: $R = X \times X_{|\mathrm{irr}}$ $\quad$ $R = \{c - d\}$ $\quad$ $R = X \times X_{|\mathrm{irr}} \setminus \{a - c, a - d\}$

A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

One symmetric relation R over X

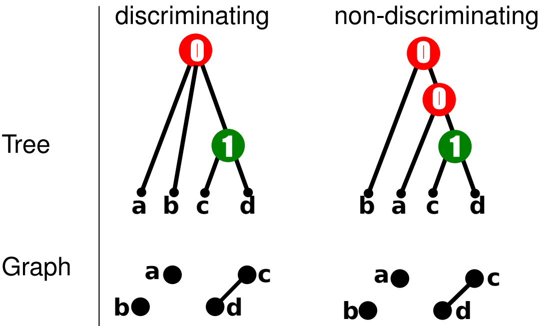


Set: $R = X \times X_{|\mathrm{irr}}$; $R = \{c - d\}$; $R = X \times X_{|\mathrm{irr}} \setminus \{a - c, a - d\}$

A tree-representation of a Relation R over X is
a tree with leaf set X and event-labels 0(•) and 1(•) s.t.:

$$\mathrm{lca}(xy) = 1 \Leftrightarrow (x, y) \in R$$

Discriminating Trees

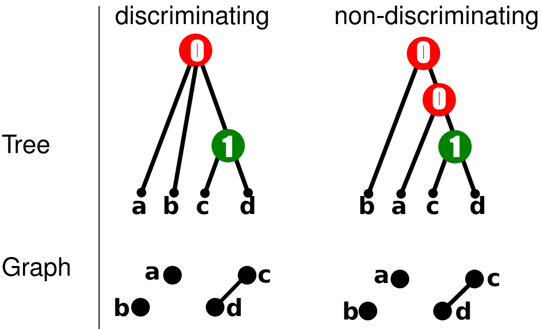


Set $R = \{c - d\}$

Here, discriminating trees, since those trees

- contain all information about the relation
- are unique (up to isom.)
- don't pretend higher resolution than actually supported by the data.

Discriminating Trees



$R = \{c - d\}$

Here, discriminating trees, since those trees

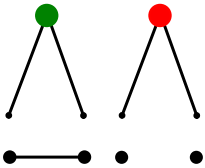
- contain all information about the relation
- are unique (up to isom.)
- don't pretend higher resolution than actually supported by the data.

Do all symmetric relations R have a tree-representation?

Relation R over X

Do all symmetric relations R have a tree-representation?

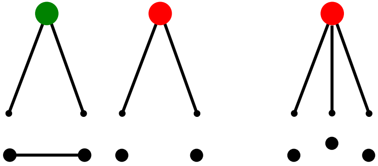
Relation R over X



$|X| = 2$

Do all symmetric relations R have a tree-representation?

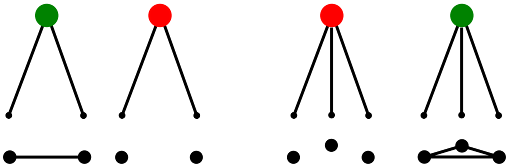
Relation R over X



$|X| = 2$ $|X| = 3$

Do all symmetric relations R have a tree-representation?

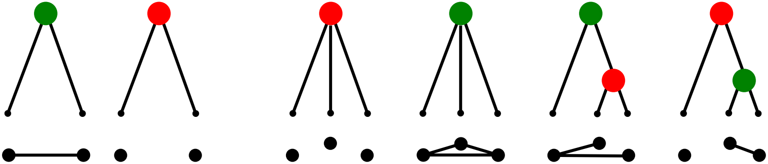
Relation R over X



$|X| = 2$ $|X| = 3$

Do all symmetric relations R have a tree-representation?

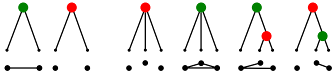
Relation R over X



$|X| = 2$ $|X| = 3$

Do all symmetric relations R have a tree-representation?

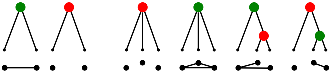
Relation R over X



$|X| = 2$ $|X| = 3$

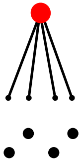
Do all symmetric relations R have a tree-representation?

Relation R over X



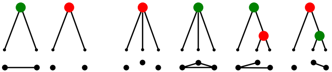
$|X| = 2$ $|X| = 3$

$|X| = 4$



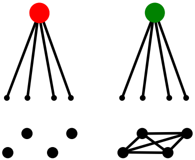
Do all symmetric relations R have a tree-representation?

Relation R over X



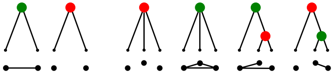
$|X| = 2$ $|X| = 3$

$|X| = 4$



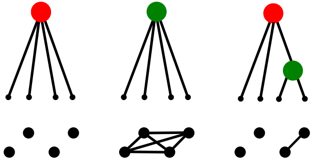
Do all symmetric relations R have a tree-representation?

Relation R over X



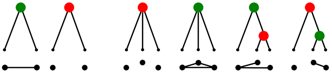
$|X| = 2$ $|X| = 3$

$|X| = 4$



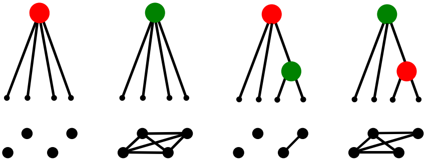
Do all symmetric relations R have a tree-representation?

Relation R over X



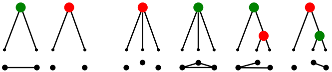
$|X| = 2$ $\qquad$ $|X| = 3$

$|X| = 4$



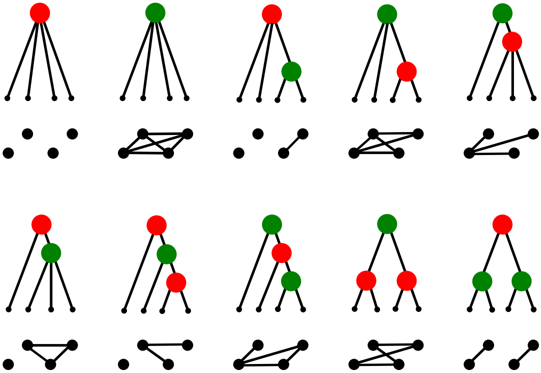
Do all symmetric relations R have a tree-representation?

Relation R over X



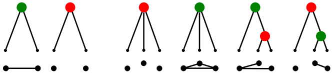
$|X| = 2$ $|X| = 3$

$|X| = 4$

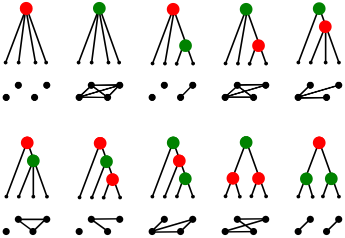


Do all symmetric relations R have a tree-representation?

Relation R over X



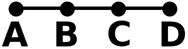
$|X| = 2$ $|X| = 3$



$|X| = 4$

If $1 \leq |X| \leq 3$, then all relations R over X have a tree-representation.

If $|X| = 4$, then all relations R over X have a tree-representation, except:



$$A - B, B - C, C - D \in R$$
$$A - C, A - D, B - D \notin R$$

Theorem (Corneil et al. (1981))

Let R be a symmetric relation over some set X.
Then the following statements are equivalent:

1. *R has a tree-representation.*
2. *The graph-representation of R does not contain induced P_4's* =Cographs

Theorem (Corneil et al. (1981))

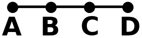
Let R be a symmetric relation over some set X.
Then the following statements are equivalent:

1. *R has a tree-representation.*
2. *The graph-representation of R does not contain induced P_4's =Cographs*

Theorem (Corneil et al. (1981))

Let R be a symmetric relation over some set X.
Then the following statements are equivalent:

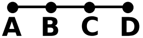
1. *R has a tree-representation.*
2. *The graph-representation of R does not contain induced P_4's* =Cographs

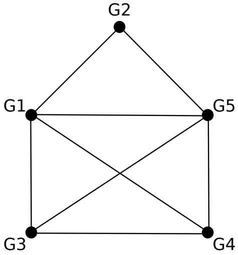


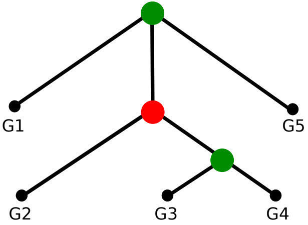
Theorem (Corneil et al. (1981))

Let R be a symmetric relation over some set X.
Then the following statements are equivalent:

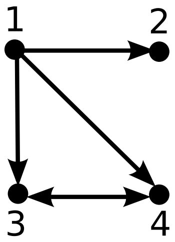
1. *R has a tree-representation.*
2. *The graph-representation of R does not contain induced P_4's* =Cographs

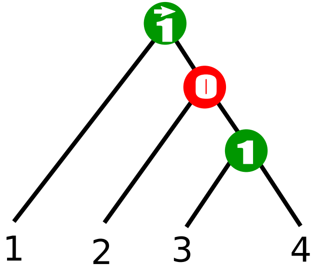






Non-symmetric relations R.





A tree with labels 0(•), 1 and $\overrightarrow{1}$ (•) represents a binary relation R, if:

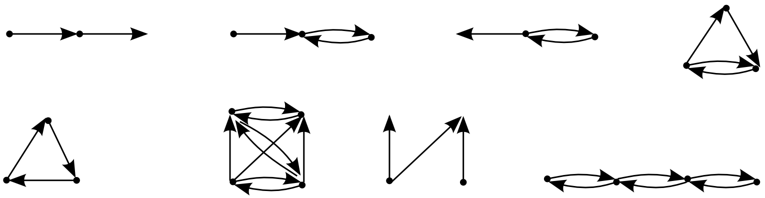
$$\mathrm{lca}(xy) = \begin{cases} 1 & \text{if } (x,y),(y,x) \in R \\ \overrightarrow{1} & \text{if } (x,y) \in R, (y,x) \notin R \text{ and } x \text{ is left from } y \text{ in } T \\ 0 & \text{otherwise} \end{cases}$$

Do all non-symmetric relations R have a tree-representation?

Theorem (Engelfriet et al. (1996))

Let R be an arbitrary relation over some set X.
Then the following statements are equivalent:

1. *R has a tree-representation.*
2. *The graph-representation of R does not contain any of the graphs below as induced subgraph.* =Di-Cographs



k disjoint symmetric relations $R_1, \ldots R_k$

Generalization to sets of symmetric relations

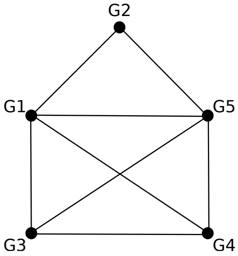
Question: When can disjoint symmetric relations $R_1, R_2, \ldots, R_k$ over X all be represented in a single tree?

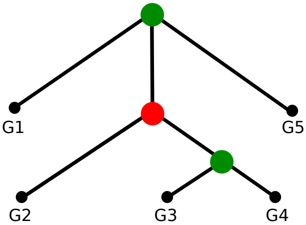
Generalization to sets of symmetric relations

Question: When can disjoint symmetric relations $R_1, R_2, \ldots, R_k$ over X all be represented in a single tree?

For R_1 und $R_1, R_2 = \overline{R_1}$ we simply have:

R_1 must have a tree-representation. (Hence, $R_2 = \overline{R_1}$ has a tree-representation).



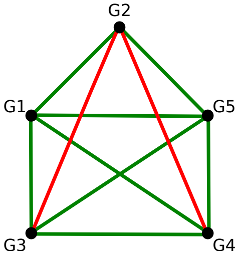


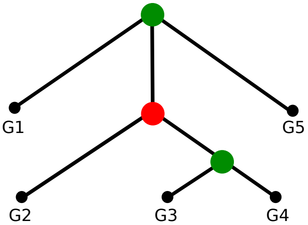
Generalization to sets of symmetric relations

Question: When can disjoint symmetric relations $R_1, R_2, \ldots, R_k$ over X all be represented in a single tree?

For R_1 und $R_1, R_2 = \overline{R_1}$ we simply have:

R_1 must have a tree-representation. (Hence, $R_2 = \overline{R_1}$ has a tree-representation).





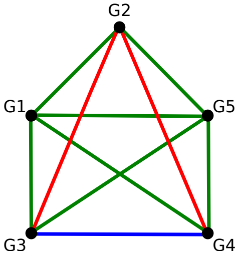
Generalization to sets of symmetric relations

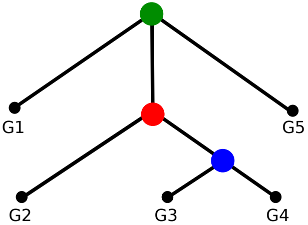
Question: When can disjoint symmetric relations $R_1, R_2, \ldots, R_k$ over X all be represented in a single tree?

$$R_1 = \{G1 - G2, G1 - G3, G1 - G4, G1 - G5, G2 - G5, G3 - G4, G3 - G5, G4 - G5\} = \text{“all green edges”}$$

$$R_2 = \{G2 - G3, G2 - G4\} = \text{“all red edges”}$$

$$R_3 = \{G3 - G4\} = \text{“all blue edges”}$$





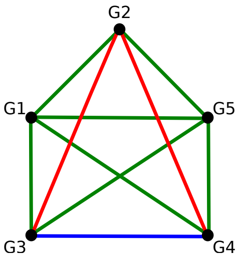
Generalization to sets of symmetric relations

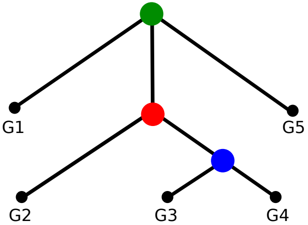
Question: When can disjoint symmetric relations $R_1, R_2, \ldots, R_k$ over X all be represented in a single tree?

Theorem (Böcker und Dress (1999), H. et. al (2014))

Disjoint symmetric relationen $R_1, R_2, \ldots, R_k$ over X can be represented in a single tree, if and only if both conditions are satisfied:

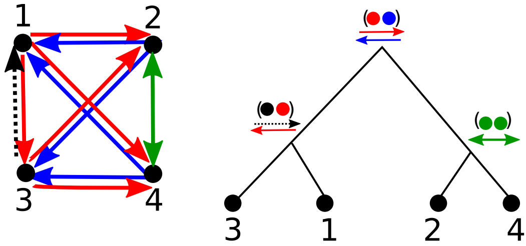
1. *[Cograph] Each R_i has a tree-representation, that is, the graph-representation of each R_i does not contain induced P_4's;*
2. *[Δ-condition] No triangle in the graph-representation of $\cup_{i=1}^{k} R_i$ (= edge-colored complete graph) has 3 distinct colors.*





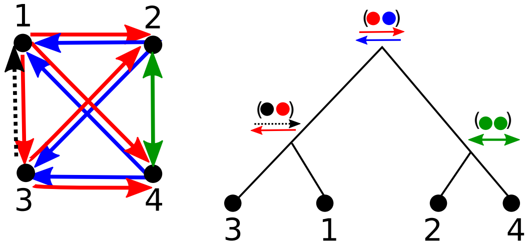
k disjoint relation $R_1, \ldots, R_k$

Sets of non-symmetric disjoint relations



Wlog. let $R_1, \ldots, R_k$ be relations s.t. $\cup_i R_i = X \times X_{|\mathrm{irr}}$.

Sets of non-symmetric disjoint relations

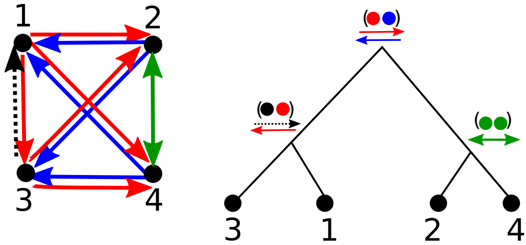


Wlog. let $R_1, \dots, R_k$ be relations s.t. $\cup_i R_i = X \times X_{|\mathrm{irr}}$.

A tree-representation of relations $R_1, \dots, R_k$ over X is a tree with leaf set X and event-labels (i,j), $i,j \in \{1, \dots, k\}$ s.t.:

$$\mathrm{lca}(xy) = \begin{cases} (i,i) & \text{if } (x,y),(y,x) \in R_i \\ (i,j) & \text{if } (x,y) \in R_i, (y,x) \in R_j, i \neq j \text{ AND } x \text{ is left from } y \text{ in } T \end{cases}$$

Sets of non-symmetric disjoint relations



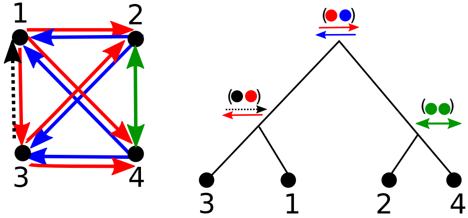
Theorem (Engelfriet et al. (1996))

Let $R_1, \ldots, R_k$ be disjoint relations over X. Then the following statements are equivalent:

1. *$R_1, \ldots, R_k$ can be represented in a single tree.*
2. *The graph-representation of $\cup_{i=1}^{k} R_i$ (= arc-colored complete di-graph) is a uniformly non-prime (unp.) 2-structure*

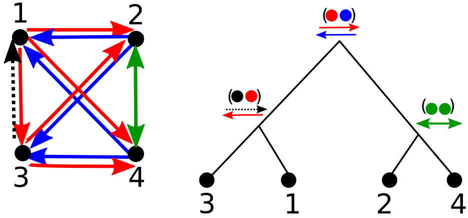
What are unp. 2-structures? - They are defined in terms of modules (omitted here)

Constructive Characterization



Since $\cup_i R_i = X \times X_{|\mathrm{irr}}$, for each distinct vertices $x, y \in X$:

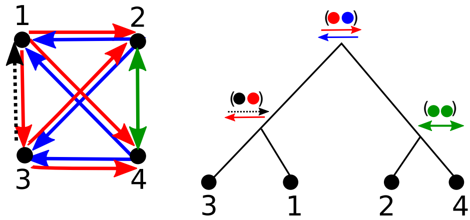
Constructive Characterization



Since $\cup_i R_i = X \times X_{|\mathrm{irr}}$, for each distinct vertices $x, y \in X$:

Either $(xy), (yx) \in R_i$ or $(xy) \in R_i$ and $(yx) \in R_j$, $j \neq i$.

Constructive Characterization

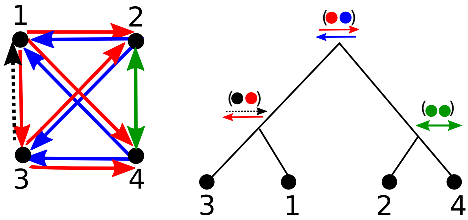


Since $\cup_i R_i = X \times X_{|\mathrm{irr}}$, for each distinct vertices $x, y \in X$:

Either $(xy), (yx) \in R_i$ or $(xy) \in R_i$ and $(yx) \in R_j$, $j \neq i$.

Hence we have on the arcs (xy) and (yx) either one color i or two colors i, j.

Constructive Characterization



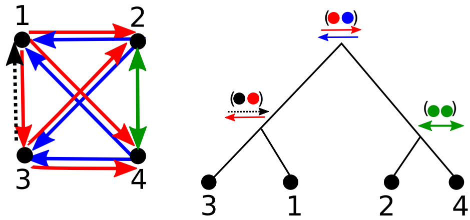
Since $\cup_i R_i = X \times X_{|\mathrm{irr}}$, for each distinct vertices $x, y \in X$:

Either $(xy), (yx) \in R_i$ or $(xy) \in R_i$ and $(yx) \in R_j$, $j \neq i$.

Hence we have on the arcs (xy) and (yx) either one color i or two colors i, j.

Define $D_{xy} := \{i, j \mid (x, y) \text{ has color } i, (y, x) \text{ has color } j\}$

Constructive Characterization



Since $\cup_i R_i = X \times X_{|\mathrm{irr}}$, for each distinct vertices $x, y \in X$:

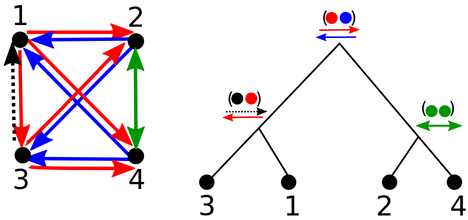
Either $(xy),(yx) \in R_i$ or $(xy) \in R_i$ and $(yx) \in R_j$, $j \neq i$.

Hence we have on the arcs (xy) and (yx) either one color i or two colors i, j.

Define $D_{xy} := \{i, j \mid (x,y) \text{ has color } i, (y,x) \text{ has color } j\}$

Exmpl.: $D_{14} = D_{34} = \{\bullet, \bullet\}, D_{13} = \{\bullet, \bullet\}, D_{24} = \{\bullet\}$

Constructive Characterization



$$D_{xy} = \{i,j \mid (x,y) \text{ has color } i, (y,x) \text{ has color } j\}$$

Theorem (2016)

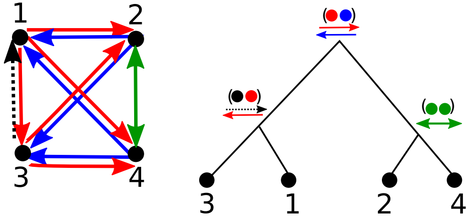
Disjoint symmetric relationen $R_1, R_2, \ldots, R_k$ over X can be represented in a single tree, if and only if both conditions are satisfied:

1. *[Di-Cograph] Each R_i has a tree-representation, that is, the graph-representation of each R_i is a di-cograph;*
2. *[Δ-condition] For all distinct $x, y, z \in X$ it holds that*

$$|\{D_{xy}, D_{xz}, D_{yz}\}| \leq 2$$

Sloppy: "No triangle has 3 distinct pairs of colors."

Constructive Characterization



$$|\{D_{13}, D_{14}, D_{34}\}| = |\{\{\bullet, \bullet\}, \{\bullet, \bullet\}\}| = 2$$

Theorem (2016)

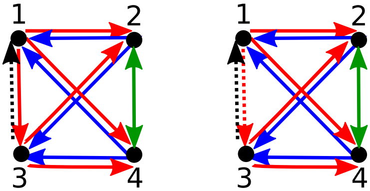
Disjoint symmetric relationen $R_1, R_2, \ldots, R_k$ over X can be represented in a single tree, if and only if both conditions are satisfied:

1. *[Di-Cograph] Each R_i has a tree-representation, that is, the graph-representation of each R_i is a di-cograph;*
2. *[Δ-condition] For all distinct $x, y, z \in X$ it holds that*

$$|\{D_{xy}, D_{xz}, D_{yz}\}| \leq 2$$

Sloppy: "No triangle has 3 distinct pairs of colors."

Constructive Characterization

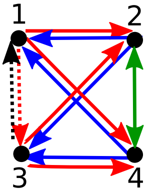


Given set of relation $R_1, \ldots, R_k$
(= colored complete di-graph G with colors $c\colon E \to \{1, \ldots, k\}$)

Reversible refinement:
Define new relations $R'_1, \ldots, R'_l$ by setting new colors in G via

$$c_{new}(xy) = c_{new}(ab) \quad \Leftrightarrow \quad c(xy) = c(ab) \text{ AND } c(yx) = c(ba)$$

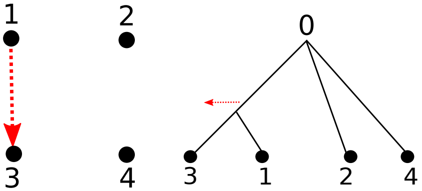
Constructive Characterization



For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label " $\rightarrow$ "

Constructive Characterization

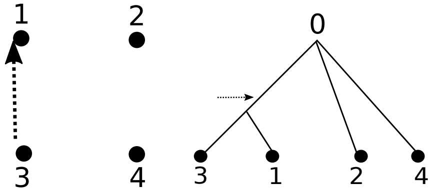


For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label " $\rightarrow$ "

$$\mathscr{C}^1 = \{\{1,3\}\}$$

Constructive Characterization

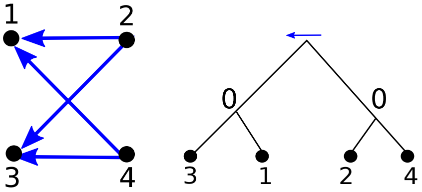


For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label "$\rightarrow$"

$$\mathscr{C}^1 = \{\{1,3\}\}$$

Constructive Characterization

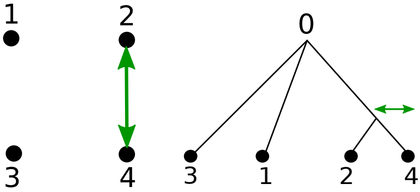


For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label " $\rightarrow$ "

$$\mathscr{C}^1 = \{\{1,3\},\{1,2,3,4\}\}$$

Constructive Characterization

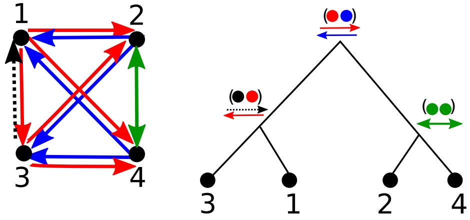


For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label " $\rightarrow$ "

$$\mathscr{C}^1 = \{\{1,3\},\{1,2,3,4\},\{2,4\}\}$$

Constructive Characterization



For each single relation R_i of $R_1, \ldots, R_k$
(= mono-chromatic subgraph with color i = di-cographs)

1. Build the respective tree-representation
2. compute "1-clusters" $\mathscr{C}^1$ = set of leaves that are descendants of vertices with label " $\rightarrow$ "

$$\mathscr{C}^1 = \{\{1,3\},\{1,2,3,4\},\{2,4\}\}$$

Constructive Characterization

Theorem (2016)

Disjoint symmetric relationen $R_1, R_2, \ldots, R_k$ over X can be represented in a single tree,

$\Leftrightarrow$

1. *[Di-Cograph] Each R_i has a tree-representation, that is, the graph-representation of each R_i is a di-cograph;*
2. *[Δ-condition] For all distinct $x, y, z \in X$ it holds that*

$$|\{D_{xy}, D_{xz}, D_{yz}\}| \leq 2$$

Sloppy: "No triangle has 3 distinct pairs of colors."

$\Leftrightarrow$

1. *[Di-Cograph]*

2'. $\mathscr{C}^1$ *in rver. refinment is tree-like (no elements overlap)*

Based on the latter characterization, we have designed an $O(|X|^2)$-recognition algorithm to test whether there is a tree-representation, and if so, construct it – **ask Nic for the fancy details ;)**

Summary and Outlook

1. Tree-representable sets of disjoint relations
2. From the “Constructive Characterization” we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:

Summary and Outlook

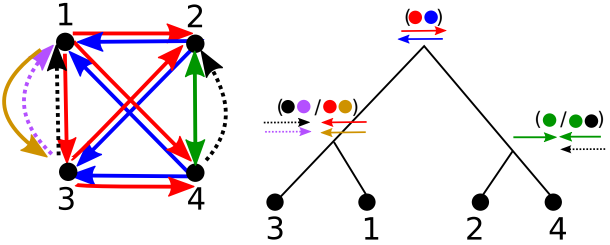
1. Tree-representable sets of disjoint relations
2. From the “Constructive Characterization” we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:

Summary and Outlook

1. Tree-representable sets of disjoint relations
2. From the "Constructive Characterization" we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:

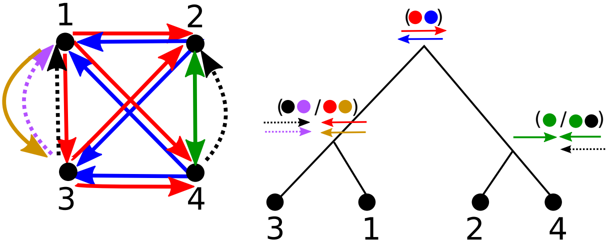
Summary and Outlook

1. Tree-representable sets of disjoint relations
2. From the "Constructive Characterization" we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:



Summary and Outlook

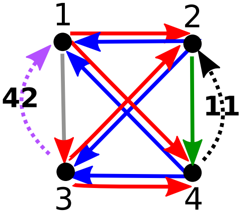
1. Tree-representable sets of disjoint relations
2. From the "Constructive Characterization" we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:

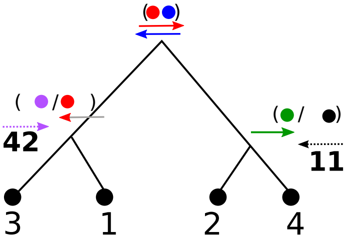


color	1	2	3	...	k	int
bitV	0	1	0	...	1	42

Summary and Outlook

1. Tree-representable sets of disjoint relations
2. From the "Constructive Characterization" we get for free an $O(|X|^2)$-time recognition algorithm and a good hint for possible heuristics to clean up estimates of sets of relations.
3. NP-completeness of Editing-Problem
4. Generalizations to sets of NON-disjoint relation = colored multi-di-graphs:





color	1	2	3	...	k	int
bitV	0	1	0	...	1	42

THANK YOU!