Tractability Results for the Consecutive-Ones Property with Multiplicity

Cedric Chauve[1], Ján Maňuch[1,2],
Murray Patterson[2] and Roland Wittler[1,3]

[1]Simon Fraser University, Canada
[2]University of British Columbia, Canada
[3]Universität Bielefeld, Germany

CPM 2011, Palermo, Italia, June 2011

The Consecutive-Ones Property

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

a	b	c	d	e
1	**1**	0	**1**	0
1	0	**1**	0	0
0	**1**	0	**1**	**1**

A non-C1P matrix

f	g	h	i	j
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

a	b	c	d	e
1	**1**	0	**1**	0
1	0	**1**	0	0
0	**1**	0	**1**	**1**

A non-C1P matrix

f	g	h	i	j
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

f	g	h	i	j
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

f	g	h	i	j
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

f	g	h	i	j
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), . . .

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

j	f	g	h	i
0	**1**	**1**	0	**1**
1	**1**	0	0	0
0	**1**	0	**1**	**1**

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

j	f	g	h	i
0	**1**	**1**	0	**1**
1	**1**	0	0	0
0	**1**	0	**1**	**1**

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), ...

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

j	f	i	g	h
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	**1**	**1**	0	**1**

The Consecutive-Ones Property

Definition

- A binary matrix M has the Consecutive Ones-Property (C1P) if its columns can be ordered in such a way that in each row, all 1's are contiguous (A C1P Ordering).
- Classical combinatorial object, used in graph theory (Booth and Lueker 1976), physical mapping (Goldberg et al. 1995), . . .

A C1P matrix

c	a	b	d	e
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	0	**1**	**1**	**1**

A non-C1P matrix

j	f	i	g	h
0	**1**	**1**	**1**	0
1	**1**	0	0	0
0	**1**	**1**	0	**1**

The Consecutive-Ones Property: Important Results

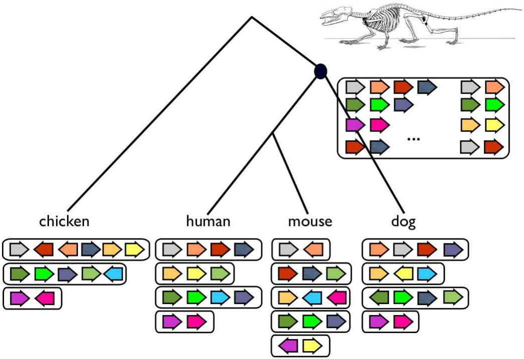
- Introduced by Fulkerson and Gross (1965), motivated by problems in genetics.
- Characterization of non-C1P matrices in terms of forbidden submatrices: Tucker (1972).
- Deciding if a binary matrix M is C1P can be done in polynomial time and all C1P column orderings can be represented in linear space with a PQ-tree: Booth and Lueker (1976).
- Decision algorithm based on partition refinement: Habib et al. (2000).
- Link with PQR-trees and partitive families: Meidanis et al. (1998, 2005), McConnell (2004).
- Algorithmical study of Tucker submatrices: Dom (2008), Blin et al. (2010).

Reconstructing Ancestral Gene Orders

Reconstructing Ancestral Gene Orders (AGOs)

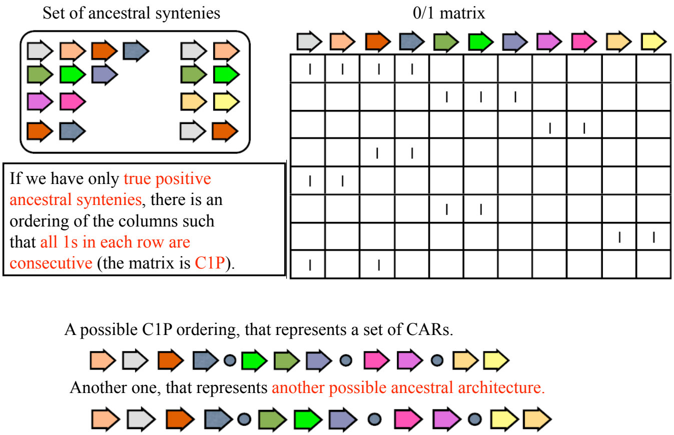
Given a phylogenetic tree on a set of extant (i.e., sequenced) species, we want to infer possible gene orders of an (unknown) ancestor in this tree. We have

1. a set of (orthologous) genomic markers, and
2. a set of ancestral syntenies: groups of markers that are believed to have been contiguous in this ancestral genome.



Reconstructing AGOs and the C1P

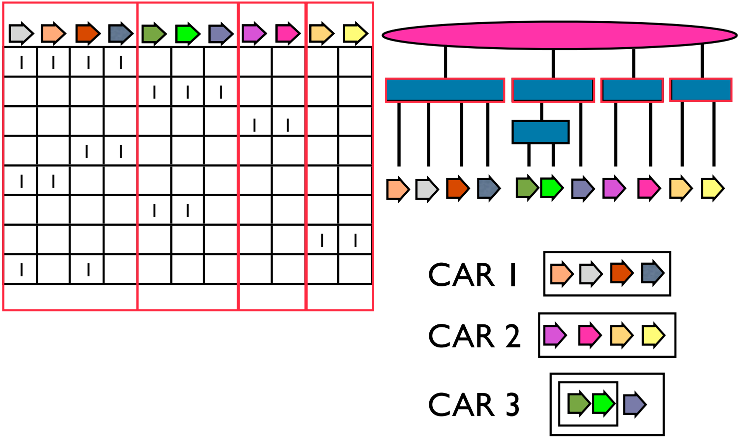
AGOs correspond to C1P orderings of the binary matrix M with rows (columns) corresponding to genomic markers (ancestral syntenies).



Each C1P ordering describes a set of possible Contiguous Ancestral Regions (CARs): Ma et al. (2006), Adam and Sankoff (2007), Chauve and Tannier (2008), ...

Reconstructing AGOs and the C1P

If binary matrix M is C1P, we can represent all C1P orderings, i.e., ancestral gene orders, with a PQ-tree (Booth and Lueker, 1976).

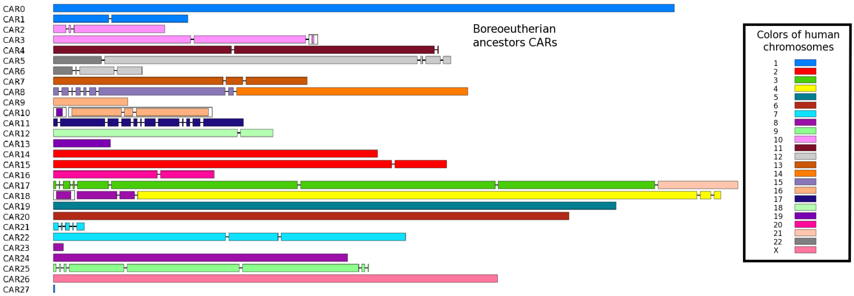


CARs are the children of the root of this PQ-tree

Reconstructing AGOs and the C1P: An Example

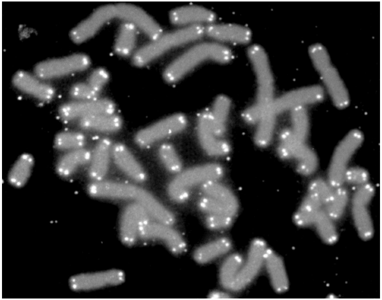
Placental mammals ancestor from 11 extant genomes (Chauve and Tannier, 2008)

- 689 markers (100kb resolution)
- 2326 ancestral syntenies
- well resolved ancestral genome with 28 CARs



Telomeres

A telomere is a region of the DNA sequence at the end of a chromosome, which protects the end of the chromosome from deterioration or from fusion with neighboring chromosomes



A Natural Question

In general, a CAR is an ancestral chromosomal segment, so which CARs are believed to

(a) form a complete ancestral chromosome? or, more generally,

(b) contain an extremity of a chromosome: an ancestral telomere?

The C1P with Multiplicity

- Allow each column c of the matrix to appear multiple ($\mathbf{m}(c) \geq 1$) times in any "ordering" S (a sequence) of columns of M
- The question is then to decide if there is an S that is "C1P" (contains each row somewhere as a subsequence) and that each column c satisfies its multiplicity constraint $\mathbf{m}(c)$
- We call such a sequence S an mC1P ordering with multiplicity vector $\mathbf{m}$

A non-C1P matrix

a	b	c	d	e
1	**1**	0	**1**	0
1	0	0	0	**1**
1	0	**1**	**1**	0

mC1P ordering: $\mathbf{m}(a) = 2$ $(\mathbf{m}(b), \ldots, \mathbf{m}(e) = 1)$

e	a	b	d	c	a
0	**1**	**1**	**1**	0	**1**
1	**1**	0	0	0	**1**
0	**1**	0	**1**	**1**	**1**

The C1P with Multiplicity

- Allow each column c of the matrix to appear multiple ($\mathbf{m}(c) \geq 1$) times in any ordering S (a sequence) of columns of M
- The question is then to decide if there is an S that is "C1P" (contains each row somewhere as a subsequence) and that each column c satisfies its multiplicity constraint $\mathbf{m}(c)$
- We call such a sequence S an mC1P ordering with multiplicity vector $\mathbf{m}$

In the literature:

- Even for matrices with 3 ones per row and $\mathbf{m}(c) \leq 2$ for all columns c, this decision problem is NP-hard: Wittler et al. (2009)

Reconstructing AGOs with Telomeres and the mC1P

We model telomeres with a column c' with multiplicity

- Let ancestral synteny abcd contain a marker that is an extremity of an ancestral chromosome (i.e., the synteny is telomeric in two extant decendants of the ancestor)
- abcd is represented in M as follows:

$$
\begin{array}{cccccc}
 & a & b & c & d & & \mathrm{c'} \\
\hline
 & & & \vdots & & & \\
\ldots & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \ldots & \mathbf{1} \\
\ldots & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \ldots & \mathbf{0} \\
 & & & \vdots & & &
\end{array}
$$

- This ensures that if M has the mC1P, then the occurences of c' are located at the extremities of the CARs (o.w. M does not have the mC1P)

Matrices with Matched Multirows: A Polytime Solvable Class of mC1P Instances

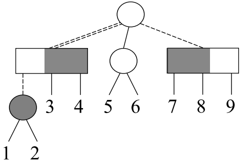
M	1	2	3	4	5	a	b
r_1	1	1	0	0	0	1	1
$\hat{r}_1$	1	1	0	0	0	0	0
r_2	1	1	1	0	0	0	0
r_3	0	0	1	1	1	0	1
$\hat{r}_3$	0	0	1	1	1	0	0
r_4	0	0	0	1	1	0	1
$\hat{r}_4$	0	0	0	1	1	0	0
r_5	1	0	0	1	1	0	0

$\hat{M}$	1	2	3	4	5
r_1	1	1	0	0	0
r_2	1	1	1	0	0
r_3	0	0	1	1	1
r_4	0	0	0	1	1
r_5	1	0	0	1	1

Left: Binary matrix M, with matched multirows. Let $\mathbf{m}(1) = \cdots = \mathbf{m}(5) = 1$ and $\mathbf{m}(a) = \mathbf{m}(b) = 2$: a and b are multicolumns and r_1, r_3 and r_4 are multirows.
Right: The corresponding matrix $\hat{M}$. Since in $\hat{M}$, by definition $\hat{r}_i = r_i$ for all multirows r_i, the matched multirows are discarded.

Idea of the Approach

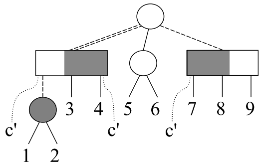
	1	2	3	4	5	6	7	8	9	c'
r_1	1	1	0	0	0	0	0	0	0	1
$\hat{r}_1$	1	1	0	0	0	0	0	0	0	0
r_2	1	1	1	0	0	0	0	0	0	0
r_3	0	0	1	1	0	0	0	0	0	1
$\hat{r}_3$	0	0	1	1	0	0	0	0	0	0
r_4	0	0	0	0	0	0	1	1	0	1
$\hat{r}_4$	0	0	0	0	0	0	1	1	0	0
r_5	0	0	0	0	0	0	0	1	1	0
r_6	0	0	0	0	1	1	0	0	0	0



Left: Binary matrix M, with matched multirows. Let $\mathbf{m}(c') = 2$.
Right: PQ-tree for $\hat{M}$. P-nodes are represented by circular nodes and Q-nodes by rectangular nodes.

Idea of the Approach

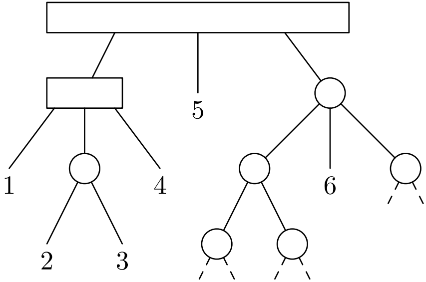
	1	2	3	4	5	6	7	8	9	c'
r_1	1	1	0	0	0	0	0	0	0	1
$\hat{r}_1$	1	1	0	0	0	0	0	0	0	0
r_2	1	1	1	0	0	0	0	0	0	0
r_3	0	0	1	1	0	0	0	0	0	1
$\hat{r}_3$	0	0	1	1	0	0	0	0	0	0
r_4	0	0	0	0	0	0	1	1	0	1
$\hat{r}_4$	0	0	0	0	0	0	1	1	0	0
r_5	0	0	0	0	0	0	0	1	1	0
r_6	0	0	0	0	1	1	0	0	0	0



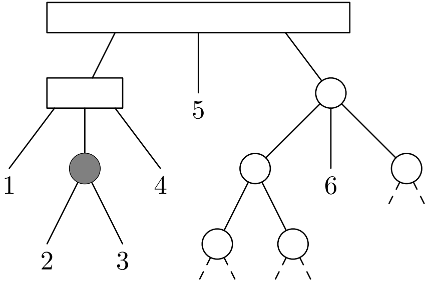
Left: Binary matrix M, with matched multirows. Let $\mathbf{m}(c') = 2$.
Right: PQ-tree for $\hat{M}$. P-nodes are represented by circular nodes and Q-nodes by rectangular nodes.

An example of a valid mC1P-ordering is $c'\,1\,2\,3\,4\,c'\,7\,8\,9\,5\,6$ which is obtained by inserting two copies of c' into the corresponding positions. Notice that inserting c' between 2 and 3 would break row r_2.

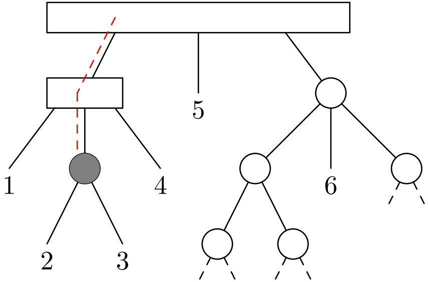
Consistency Check: The Four Cases



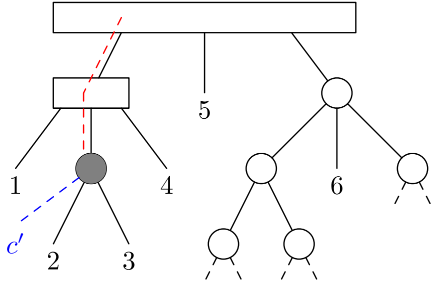
Consistency Check: Case 1



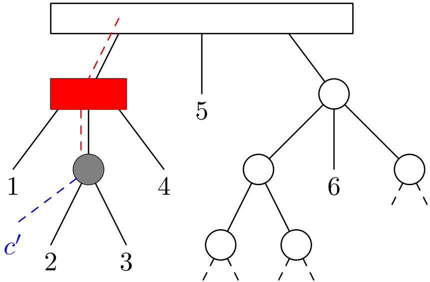
Consistency Check: Case 1



Consistency Check: Case 1

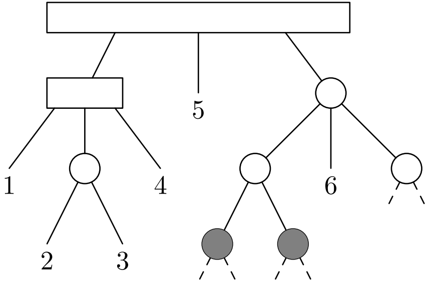


Consistency Check: Case 1

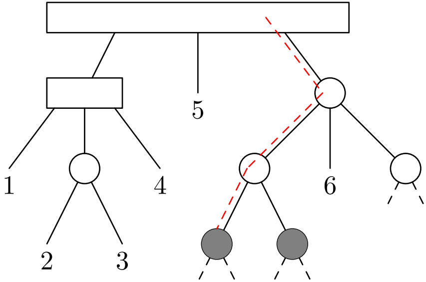


Here, insertion of c' would break either row 123 or row 234.

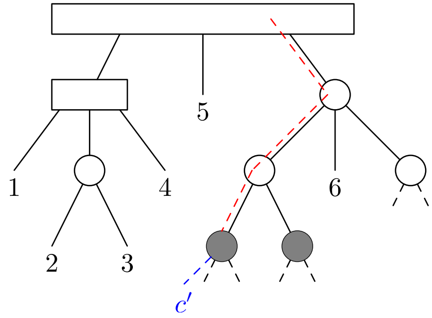
Consistency Check: Case 2



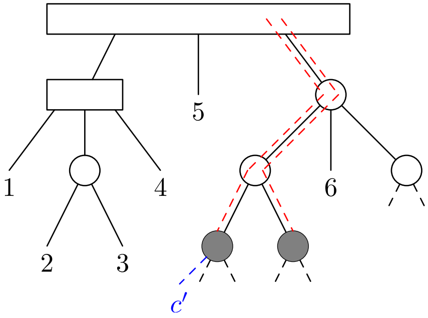
Consistency Check: Case 2



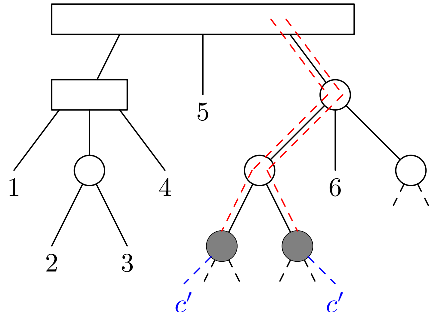
Consistency Check: Case 2



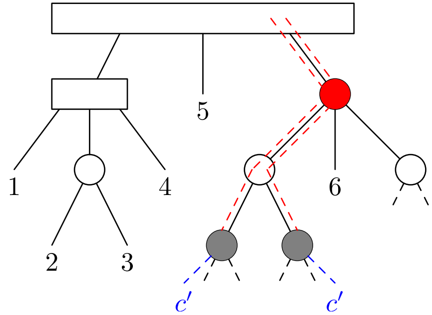
Consistency Check: Case 2



Consistency Check: Case 2

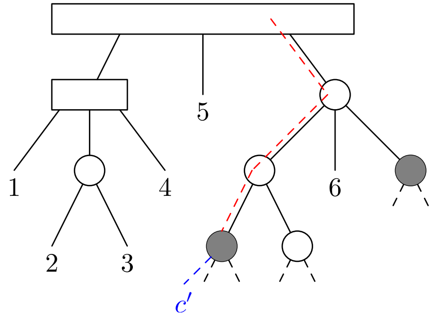


Consistency Check: Case 2

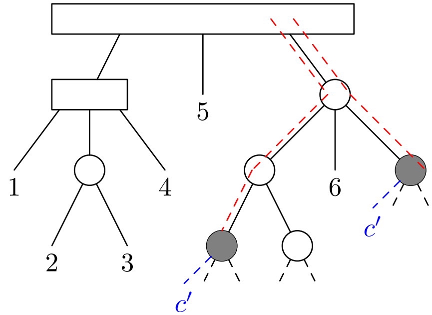


Here, insertion of c' would break one of the rows associated with this node.

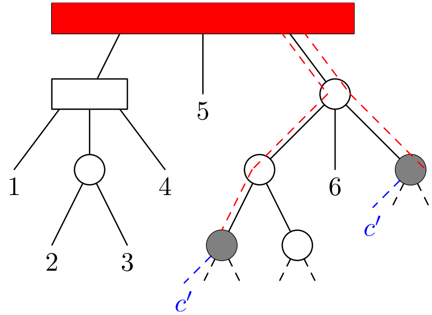
Consistency Check: Case 3



Consistency Check: Case 3

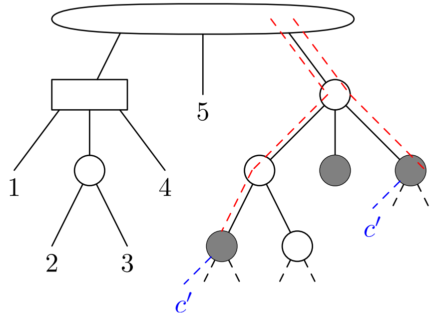


Consistency Check: Case 3

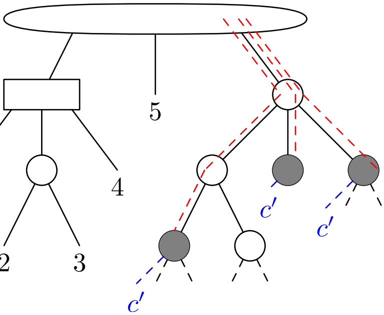


Here, insertion of c' would break one of the rows associated with the root node.

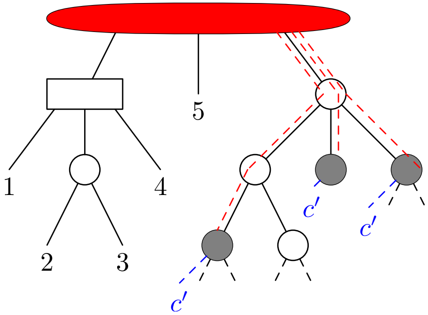
Consistency Check: Case 4



Consistency Check: Case 4



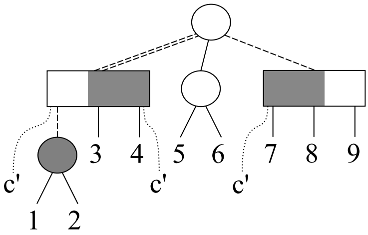
Consistency Check: Case 4



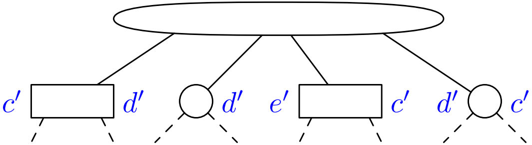
Here, insertion of c' would break one of the rows associated with the root node.

Multiplicity Check

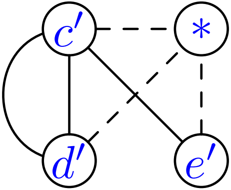
- If the consistency check succeeds for each row, we simply have to ensure that the PQ-tree satisfies the multiplicity requirement



Case with Several Multicolumns



This corresponds to an Eulerian cycle in the following multigraph



Conclusion

Here we extend the domain of tractable instances of deciding the C1P with multiplicity. Several questions remain open:

- Is this the largest class of tractable instances of the mC1P?
- Is there structure analgous to the PQ-tree that could encode all mC1P-orderings of a matrix that satisfies this property? (Note that our data structure does not incorporate the multiplicity constraint)
- Our algorithm takes time $O(mn)$ where m (n) is the number of rows (columns). It is open whether there is an $O(m + n + \ell)$-time algorithm where ℓ is the number of entries 1 in M

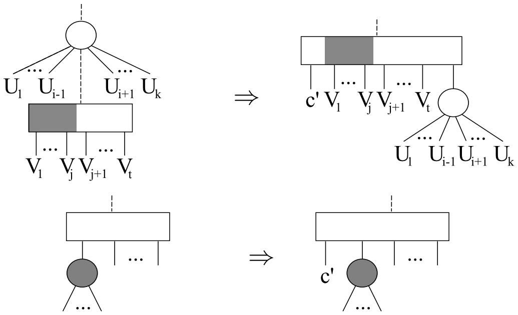
Acknowledgements

- Éric Tannier for suggesting the idea of using the mC1P to model telemeres in this setting
- NSERC discovery grant

Thanks!

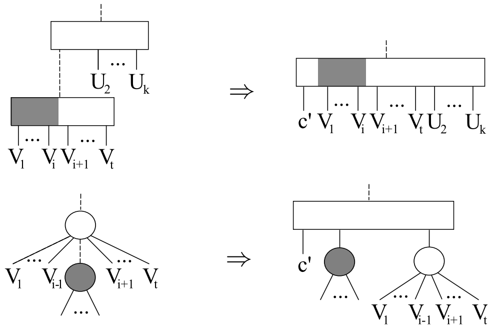
Any Questions or Comments?

Transformation Rules



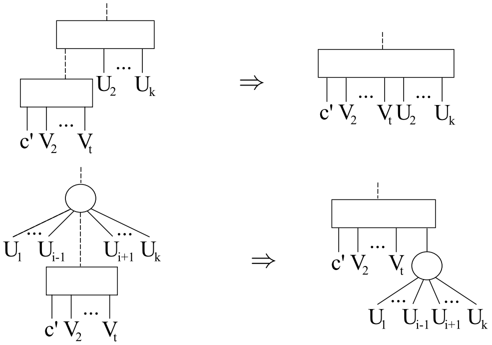
Transformation rules for the LCAs to construct an augmented PQ-tree. An LCA and its parent node are replaced by the nodes shown on the right. The LCA (or the segment of an LCA, respectively) are highlighted in gray.

Transformation Rules



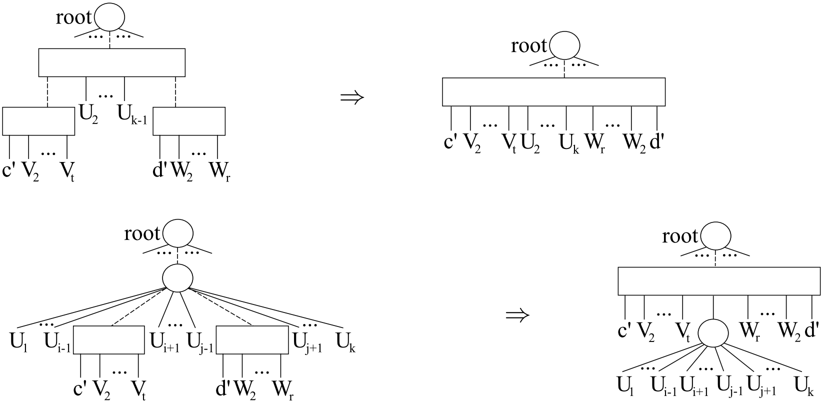
Transformation rules for the LCAs to construct an augmented PQ-tree. An LCA and its parent node are replaced by the nodes shown on the right. The LCA (or the segment of an LCA, respectively) are highlighted in gray.

Transformation Rules



Transformation rules for bottom-up iteration to construct an augmented PQ-tree. A newly created Q-node and its parent node are replaced by the nodes shown on the right.

Transformation Rules



Special transformation rules for bottom-up iteration to construct an augmented PQ-tree. A newly created Q-node two levels below the root node and its parent node are replaced by the nodes shown on the right.