AMS 2012 Spring Southeastern Section Meeting
University of South Florida, Tampa, FL
Special Session on Discrete Models in Molecular Biology
march 11, 2012

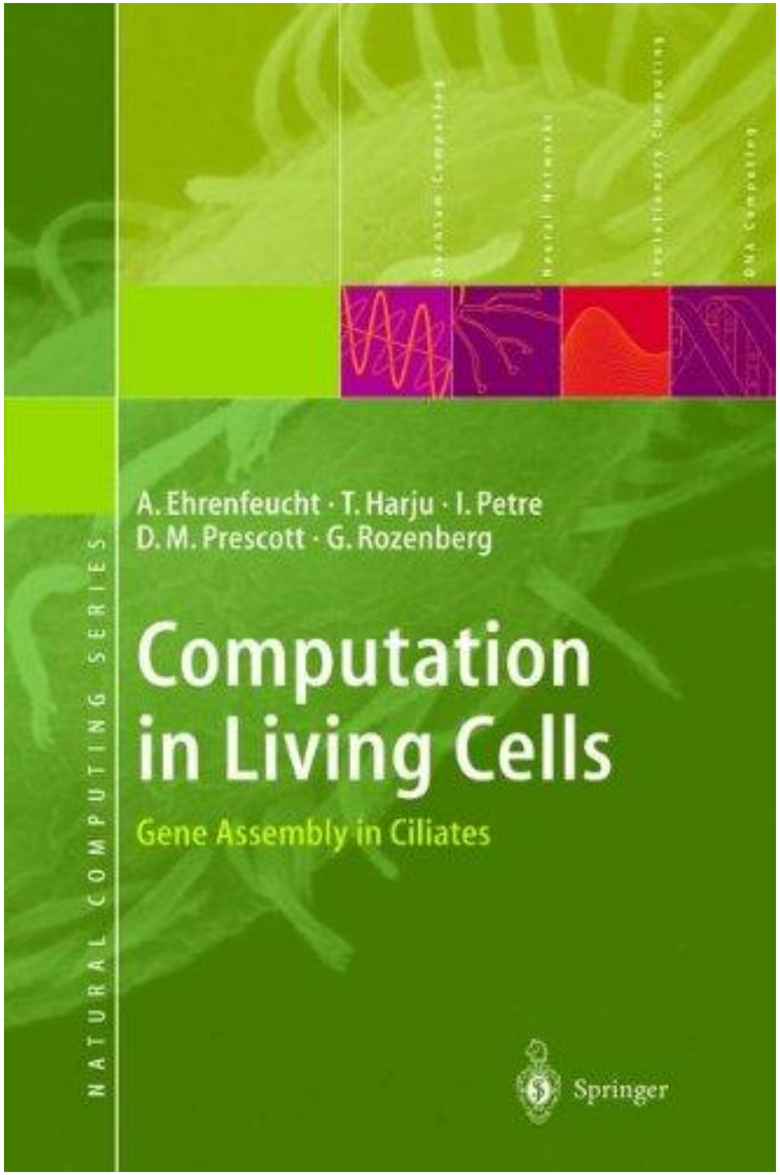
The Algebra of Ciliates

Hendrik Jan Hoogeboom
Leiden

with
Robert Brijder
Hasselt

image: http://www.depts.ttu.edu/hillcountry/research/protozoa.php

the book



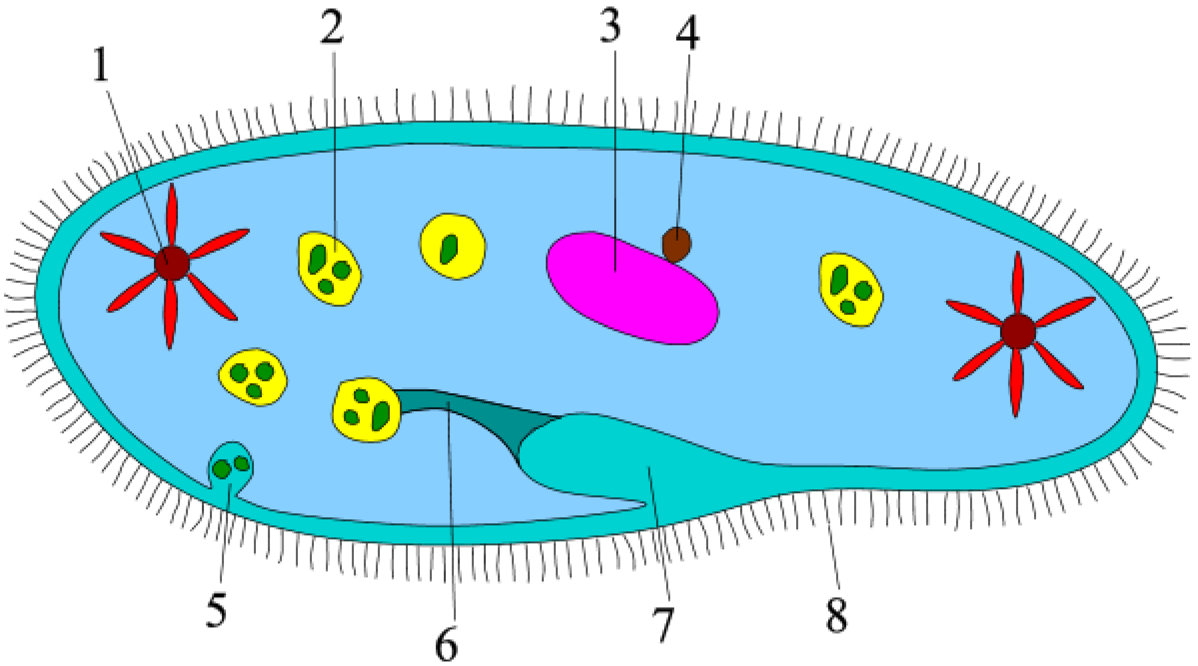
Computation in Living Cells
– Gene Assembly in Ciliates
A.Ehrenfeucht, T.Harju, I.Petre,
D.M.Prescott, G.Rozenberg
Springer, 2004

the overview

R.Brijder, M.Daley, T.Harju, N.Jonoska, I.Petre, G.Rozenberg: Computational nature of gene assembly in ciliates.

in: G. Rozenberg, T.H.W. Bäck, J.N. Kok, editors, *Handbook of Natural Computing,* Springer, 2012. Expected

ciliates: micro and macro



cell structure:

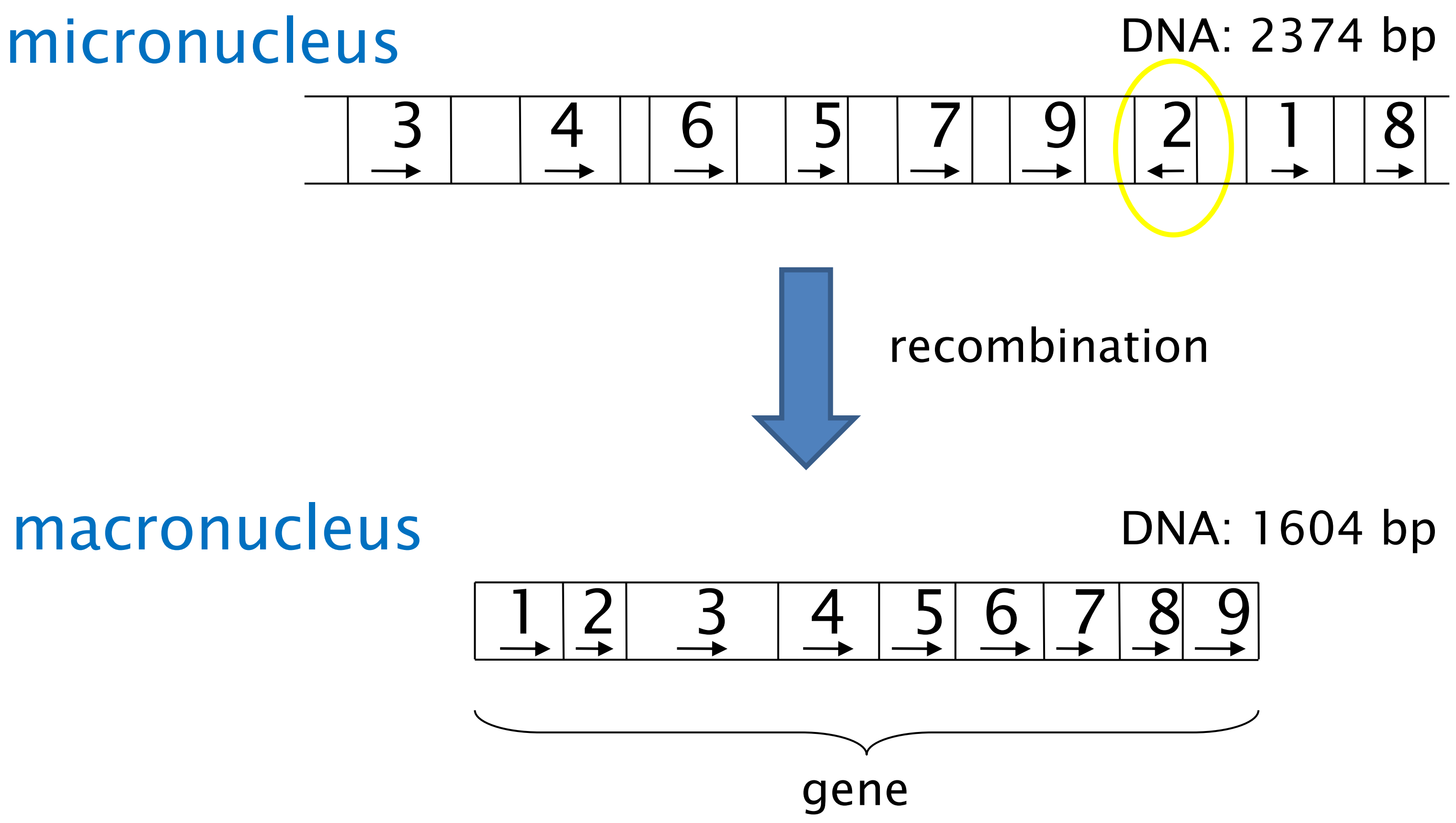
3. macronucleous
4. micronucleous
8. cilium

Unlike most other eukaryotes, ciliates have two different sorts of nuclei: a small, diploid **micronucleus** (reproduction), and a large, polyploid **macronucleus** (general cell regulation). The latter is generated from the micronucleus by amplification of the genome and heavy editing.

http://en.wikipedia.org/wiki/Ciliate

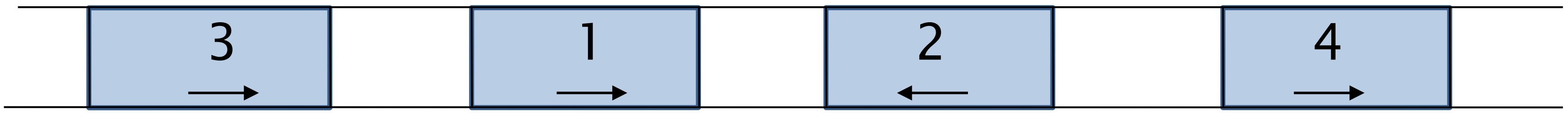
from micro to macro

http://oxytricha.princeton.edu/cgi-bin/get_MDS_IES_Info.cgi?num=38

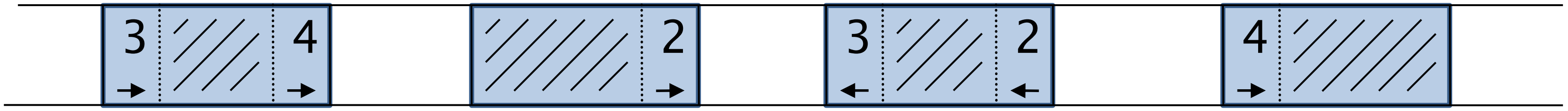


Greslin, Prescott etal. Reordering of nine exons is necessary to form a functional actin gene in Oxytricha nova. PNAS 86, 6264–6268, Aug 1989.

pointers

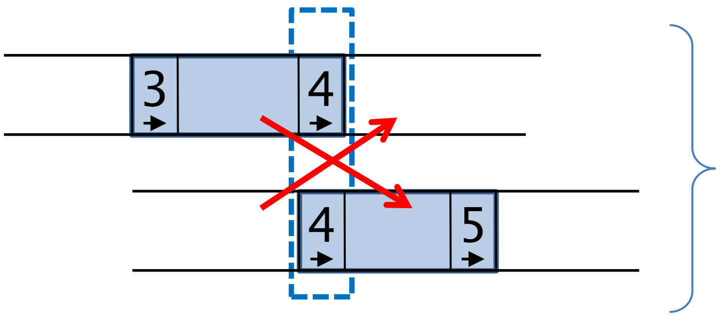


pointers – overlapping segments (for glueing)

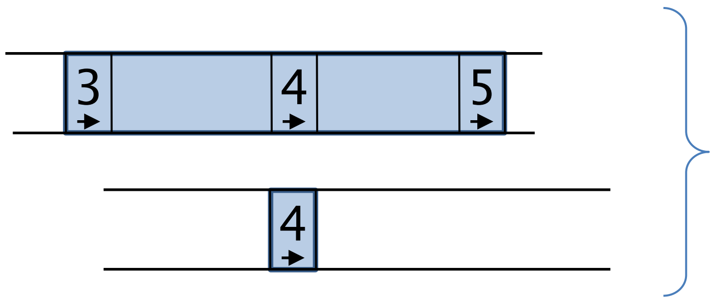


e.g., pointer 5 of actin gene: 13 bp

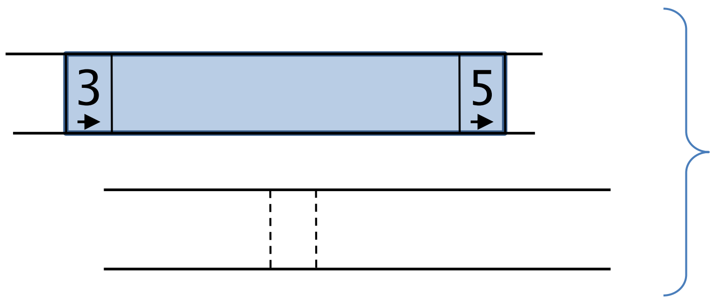
rc_4 recombination on pointer 4 'generic'



before



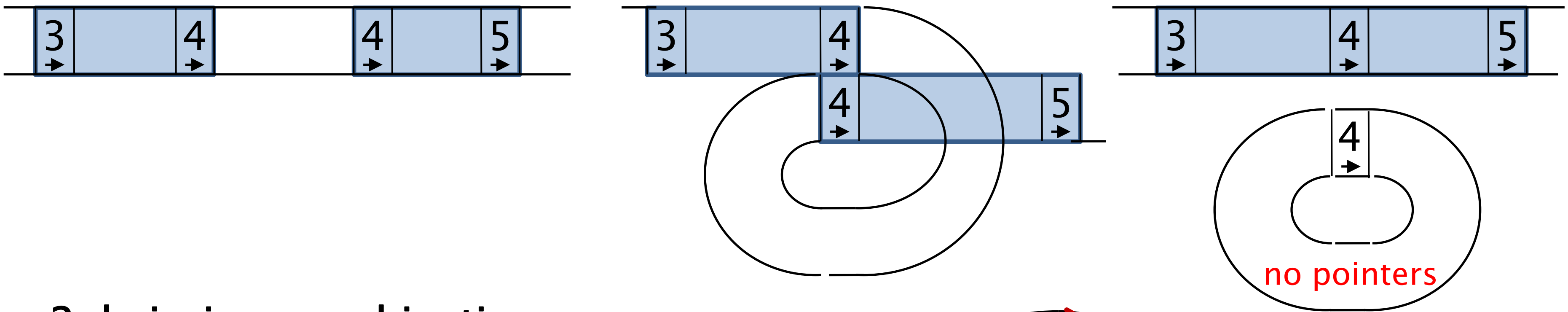
after 'math view'



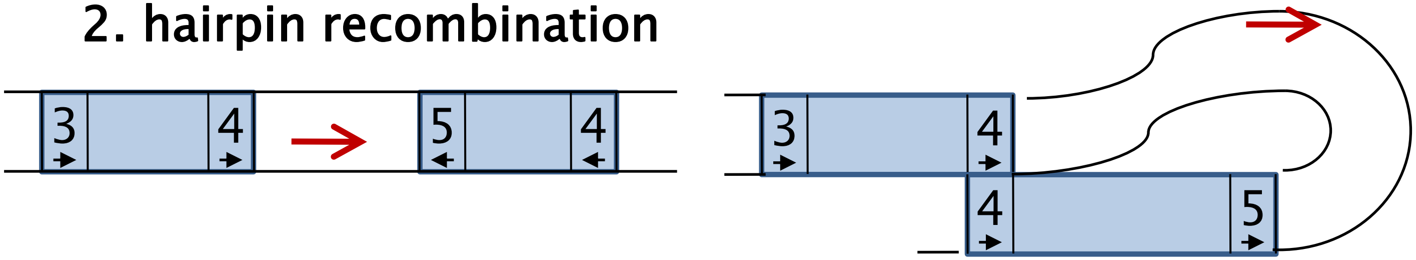
after 'ciliate view'

recombination on pointers

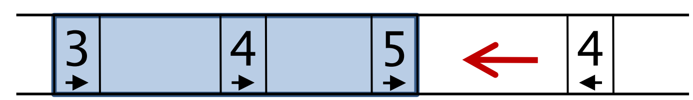
1. loop recombination

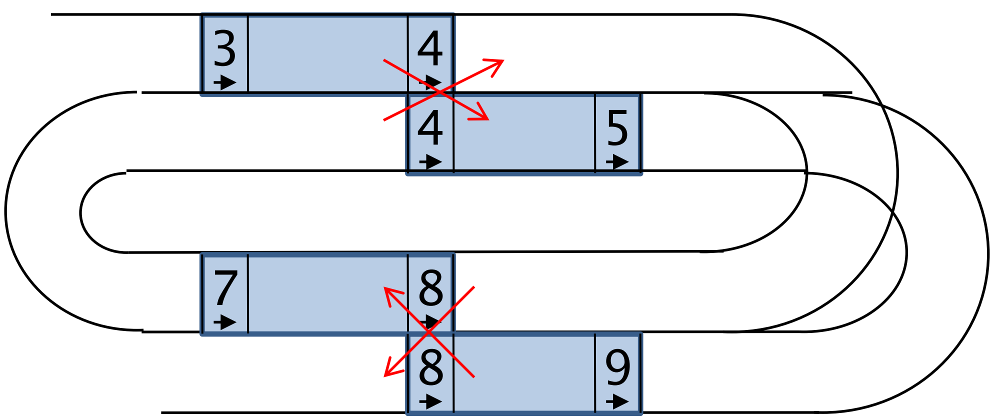


2. hairpin recombination



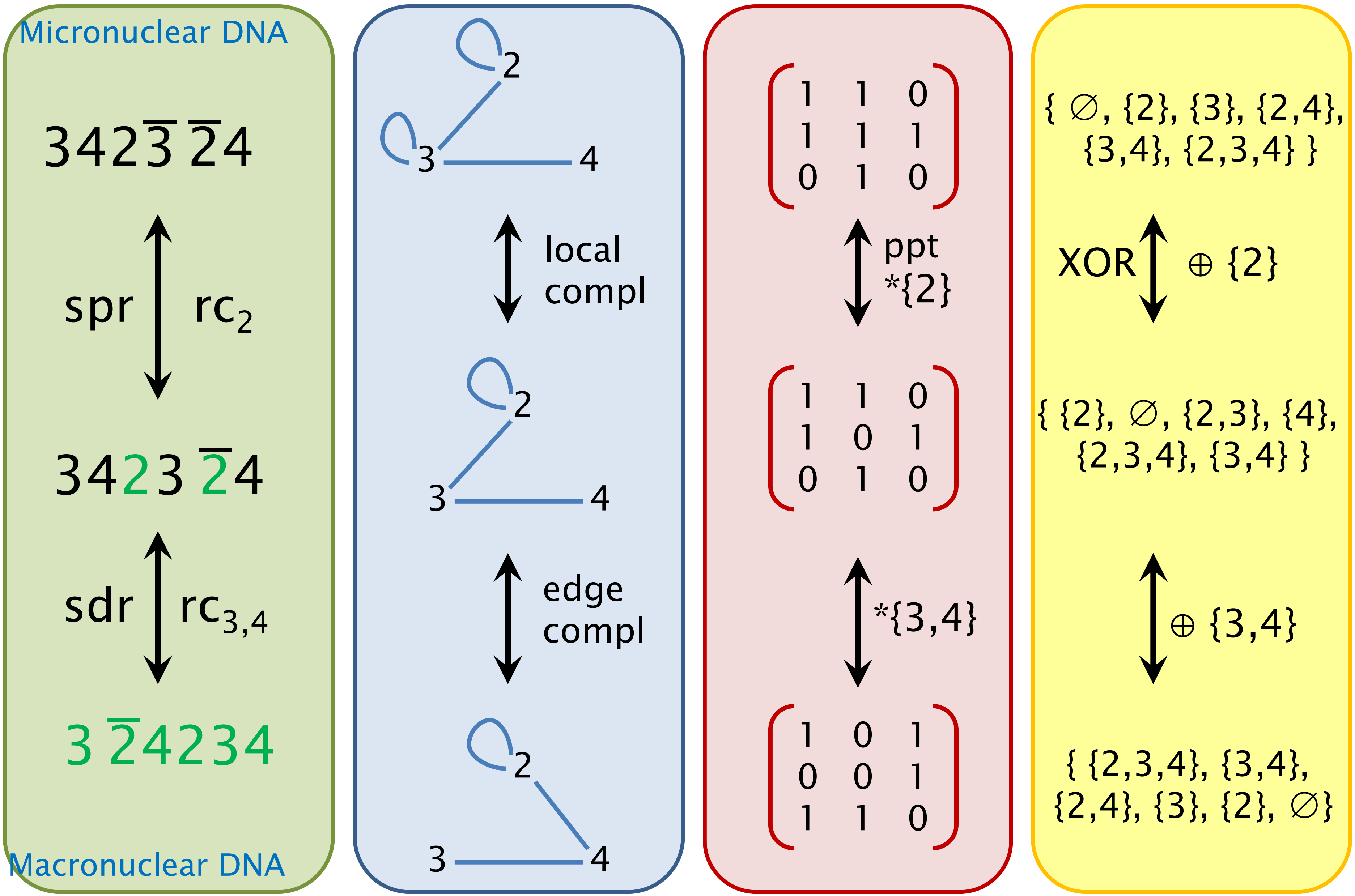
3. double-loop recombination





pointers between
two copies of 4

four worlds

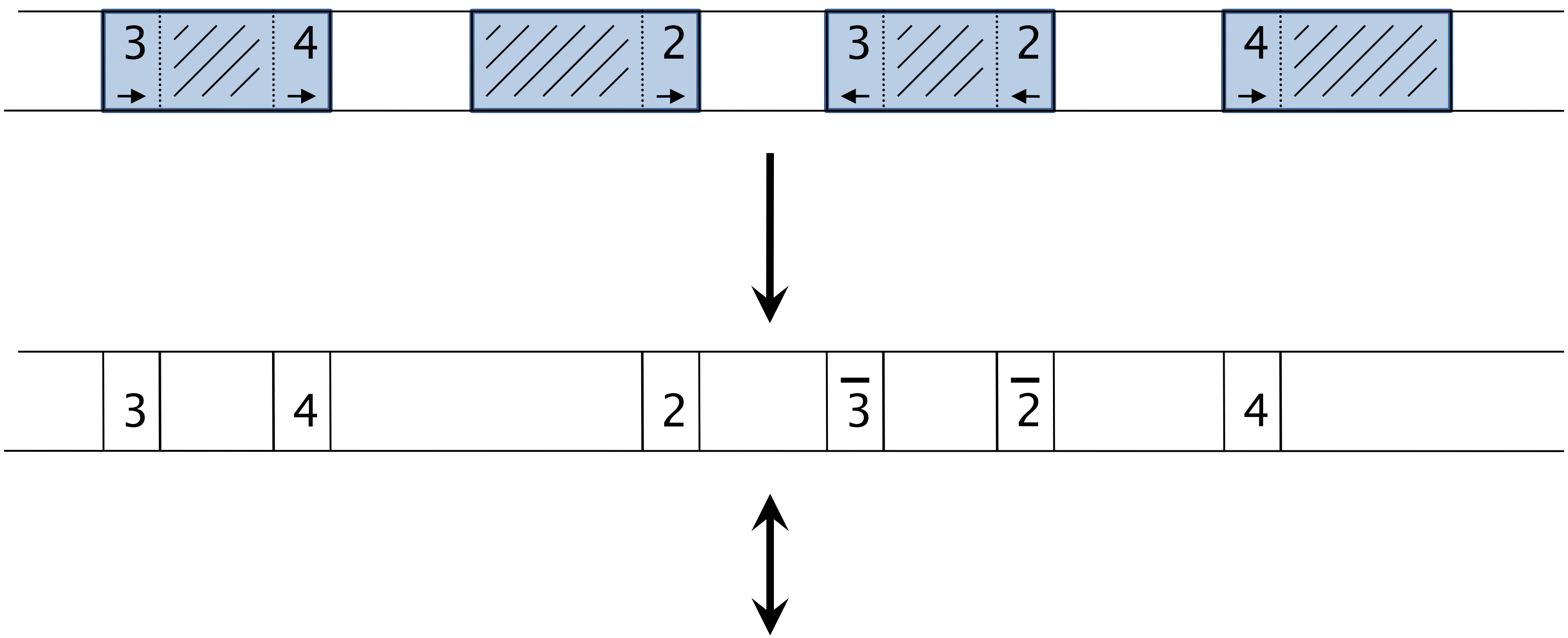


Micronuclear DNA
$342\overline{3}\,\overline{2}4$
spr
rc_2
$3423\,\overline{2}4$
sdr
$rc_{3,4}$
$3\,\overline{2}4234$
Macronuclear DNA
2
3
4
local compl
2
3
4
edge compl
2
3
4
$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
ppt
*{2}
$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
*{3,4}
$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$
{ ∅, {2}, {3}, {2,4}, {3,4}, {2,3,4} }
XOR
⊕ {2}
{ {2}, ∅, {2,3}, {4}, {2,3,4}, {3,4} }
⊕ {3,4}
{ {2,3,4}, {3,4}, {2,4}, {3}, {2}, ∅}

quest for the "right" model

- strings + graphs
- matrices
- set systems

abstraction: pointers



$3\,4\,2\,\bar{3}\,\bar{2}\,4$ ‘legal’ string

... 4774 ... realistic strings vs. generalizations

string pointer reduction systems

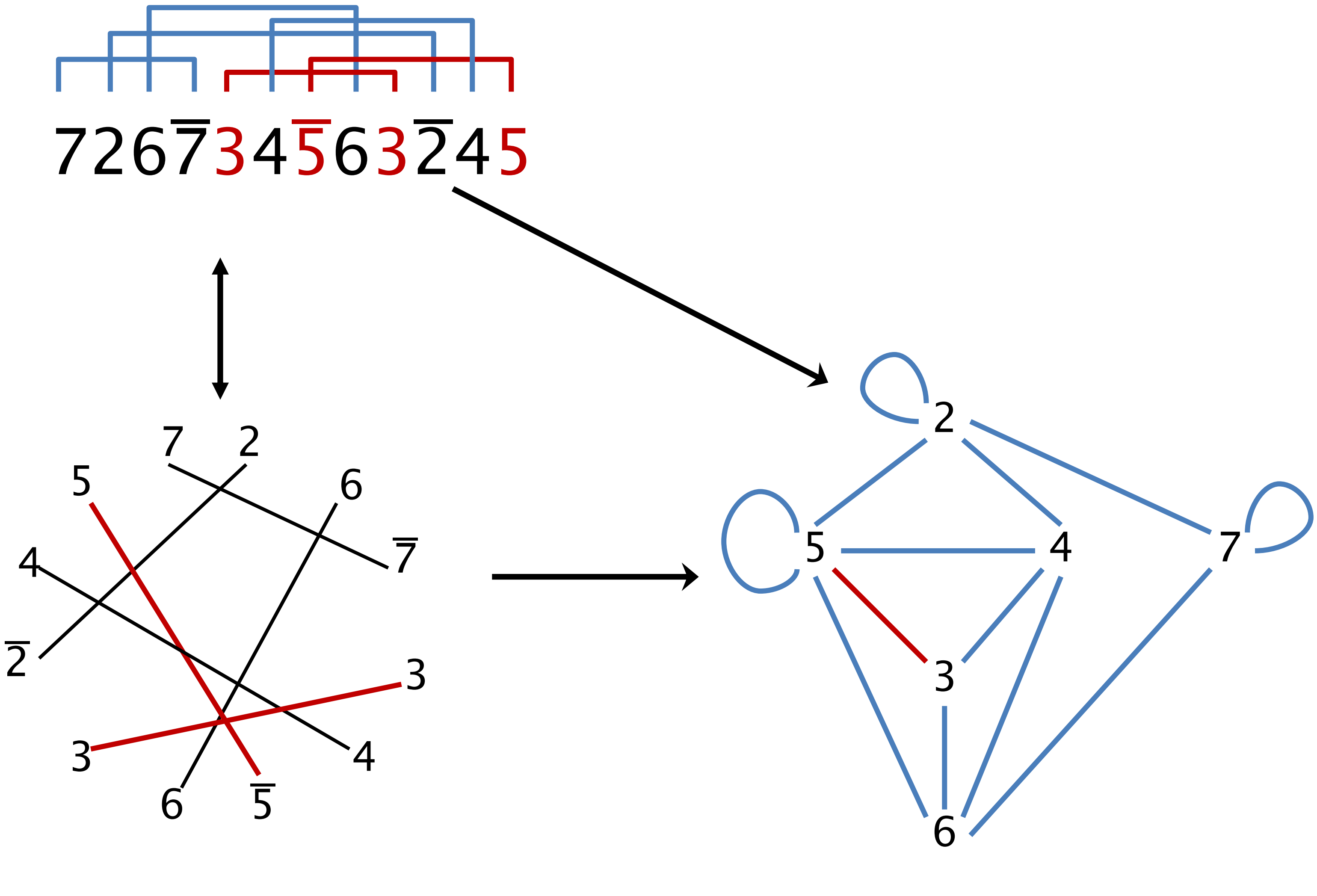
$$rc_p(\ u_1 p p u_2\ ) = u_1 p p u_2$$

no rearrangement
excision circular molecule

$$rc_p(\ u_1 p u_2 \bar{p} u_3\ ) = u_1 p \bar{u}_2 \bar{p} u_3$$

$$rc_{p,q}(\ u_1 p u_2 q u_3 p u_4 q u_5\ ) = u_1 p u_4 q u_3 p u_2 q u_5$$

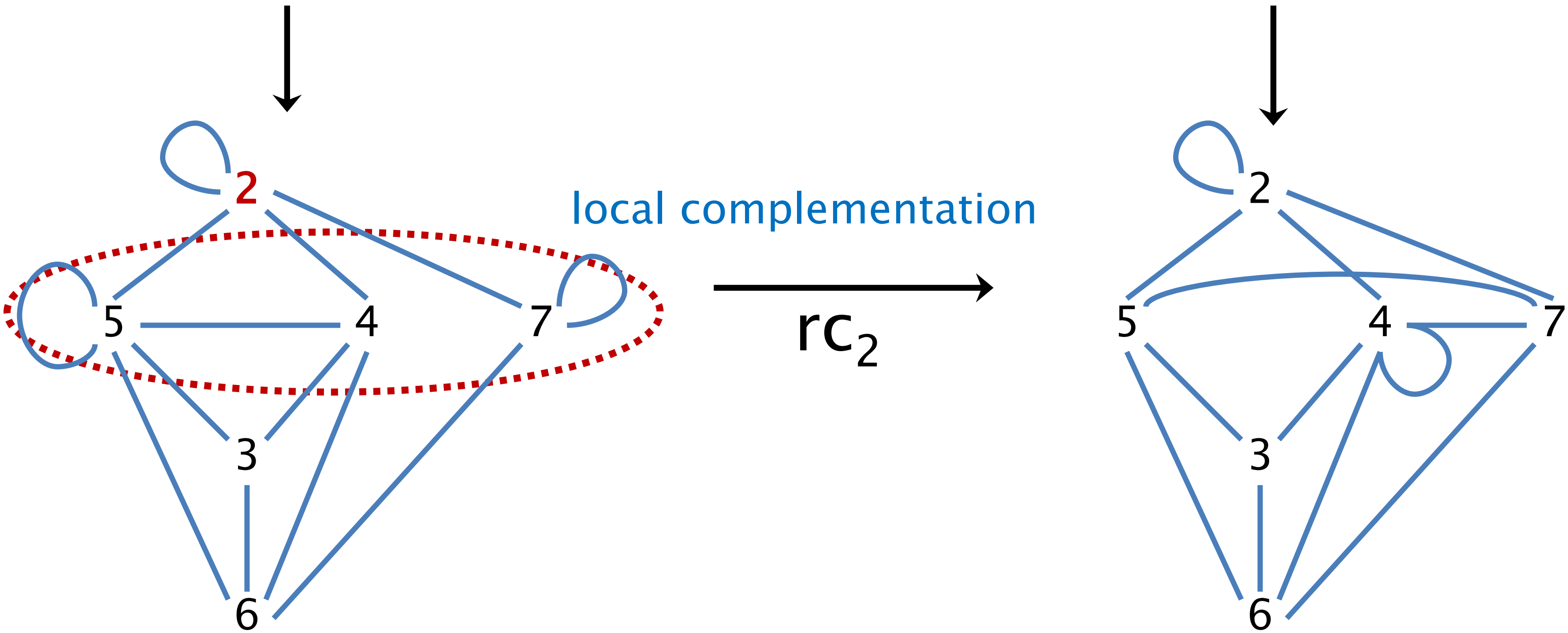
circle & interval / overlap graph



string to overlap graph

real generalization

$$7\,2\,6\,\overline{7}\,3\,4\,\overline{5}\,6\,3\,\overline{2}\,4\,5 \xrightarrow{\;rc_2\;} 7\,2\,\overline{3}\,\overline{6}\,5\,\overline{4}\,\overline{3}\,7\,\overline{6}\,\overline{2}\,4\,5$$

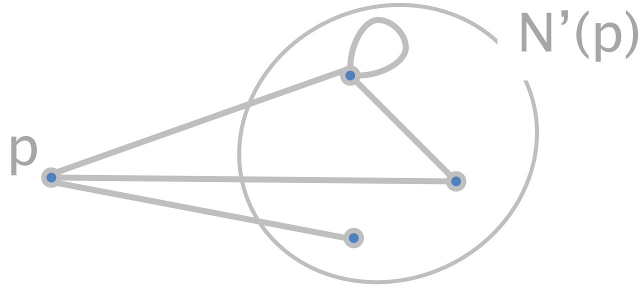


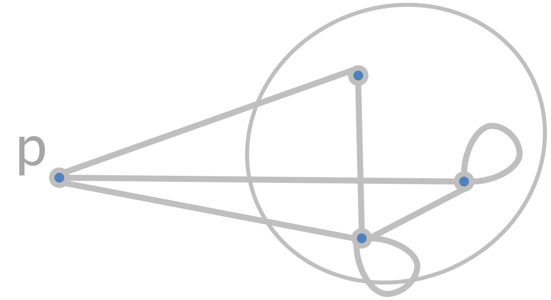
Ehrenfeucht etal, *Theor. Comp. Sci.*, 2003

(for signed graphs instead of looped graphs)

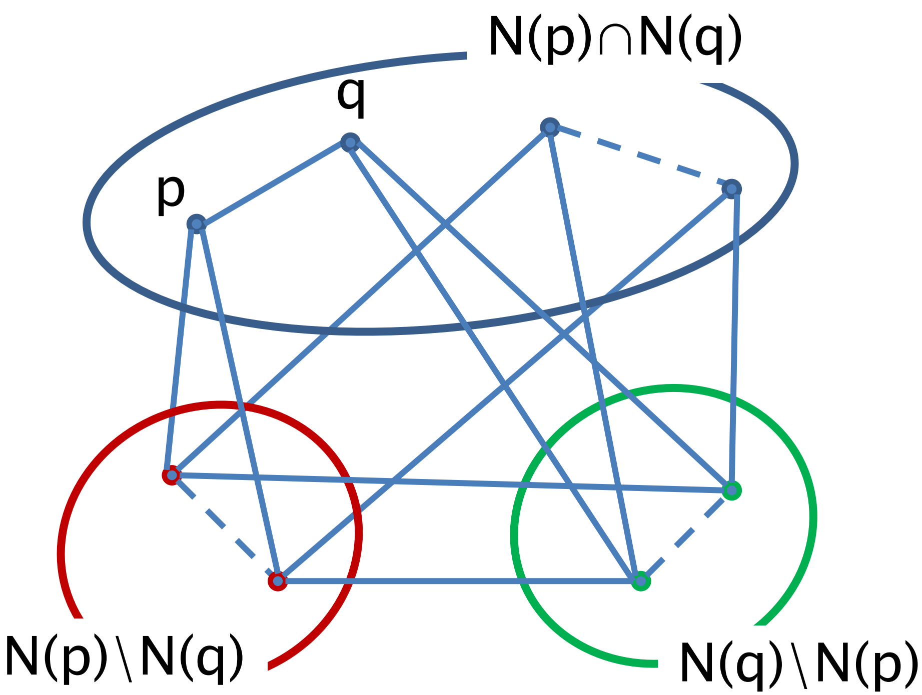
graph operations

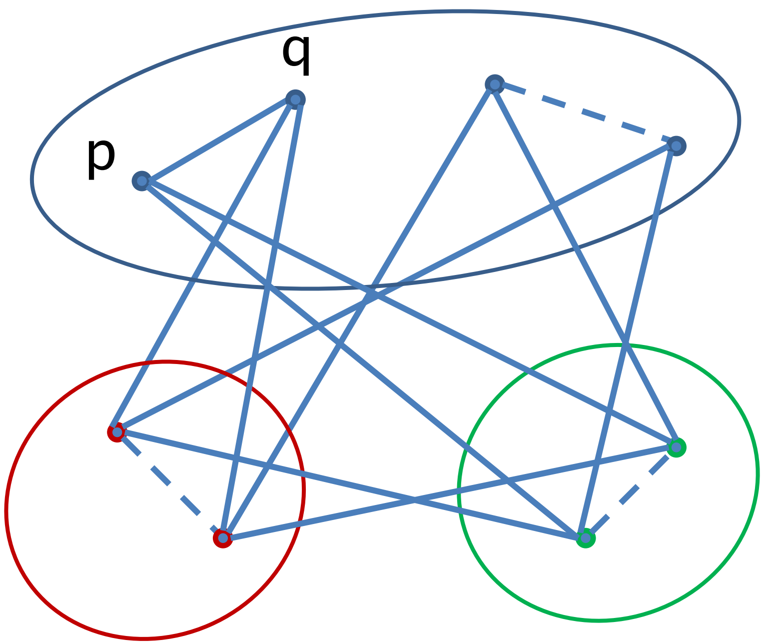
rc_p *local complementation* looped vertex p





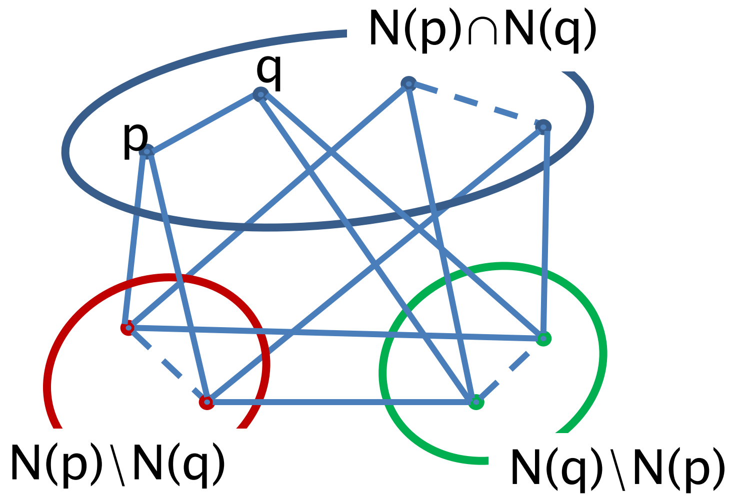
$rc_{p,q}$ *edge complementation* unlooped edge pq

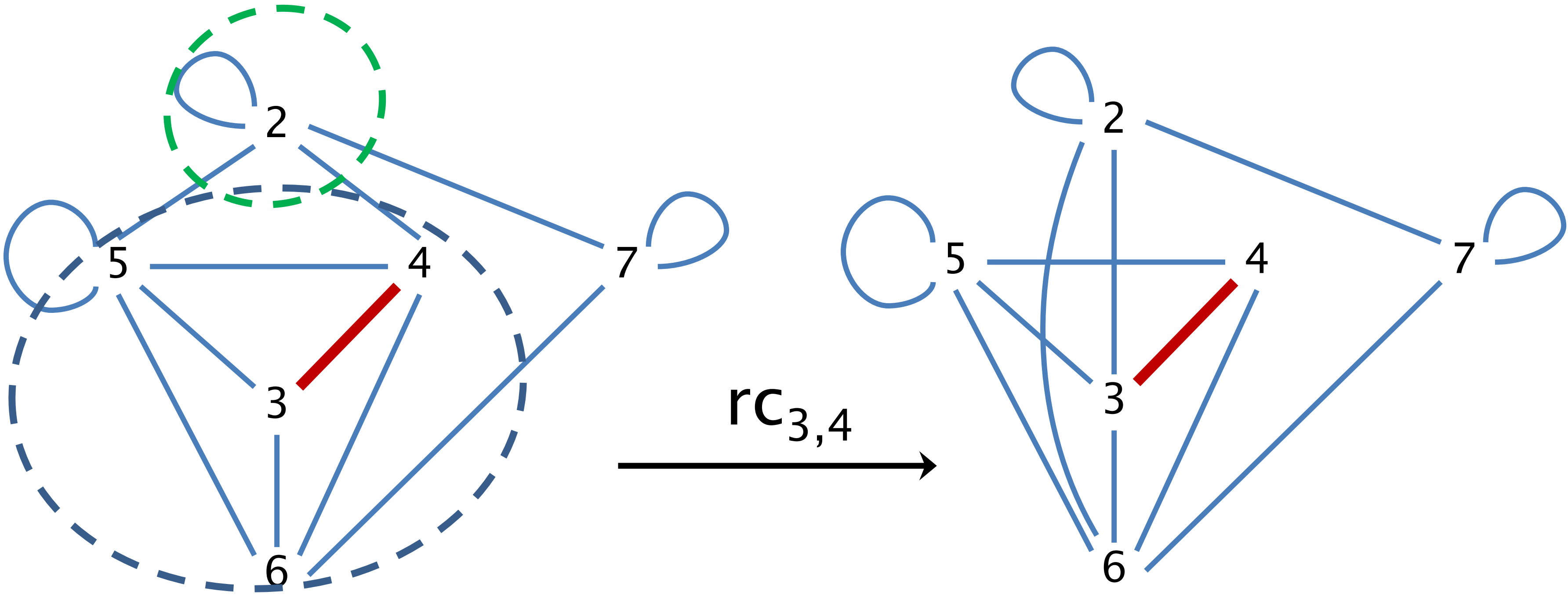


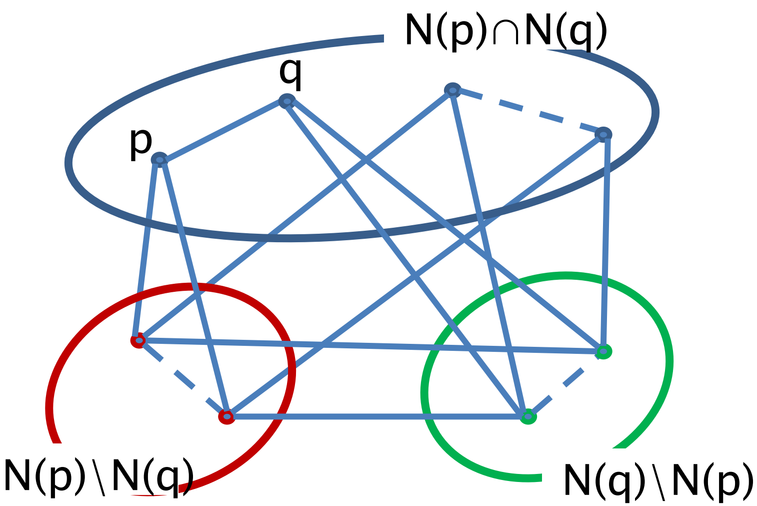


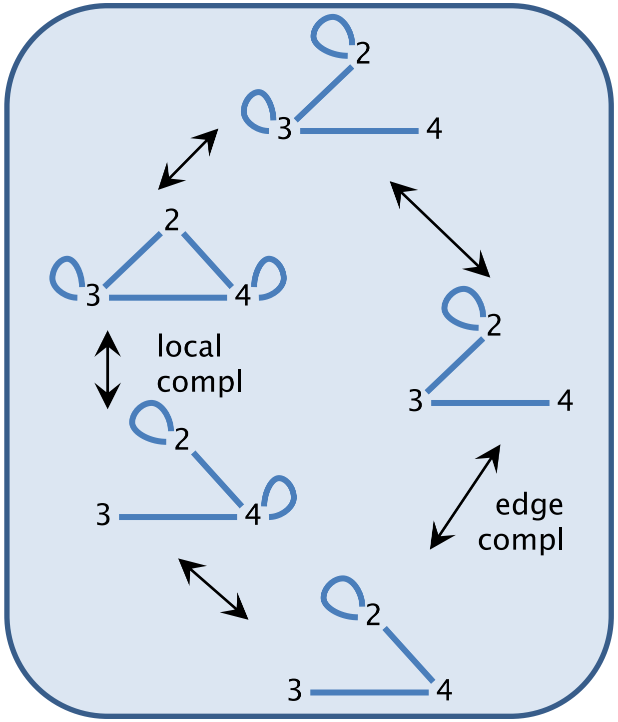
example edge complementation

$rc_{3,4}$ on edge 3,4









questions:

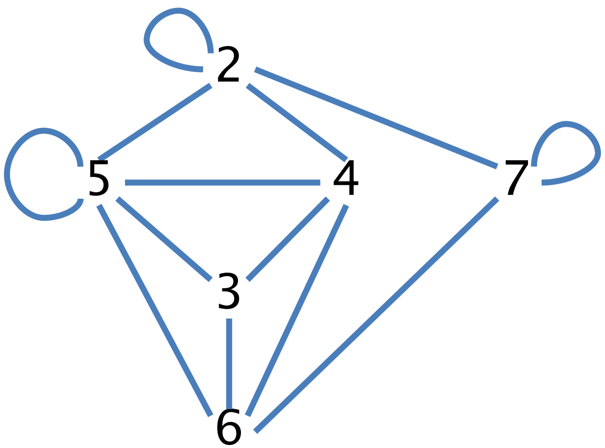
how do $rc_{p,q}$ and $rc_{p'}$ or $rc_{p',q'}$ interact ?

is the result of reductions dependent on (order) operations chosen?

quest for the “right” model

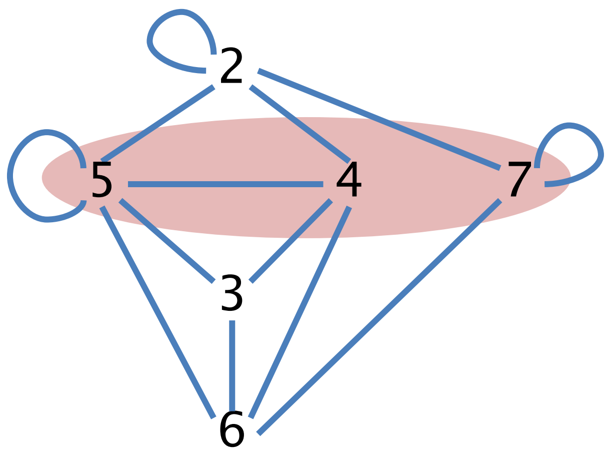
- strings
- graphs $\Leftrightarrow$ matrices
- set systems

graphs and matrices



$$
\begin{array}{c c}
 & \begin{array}{cccccc} 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\
\begin{array}{c} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} &
\begin{pmatrix}
1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
$$

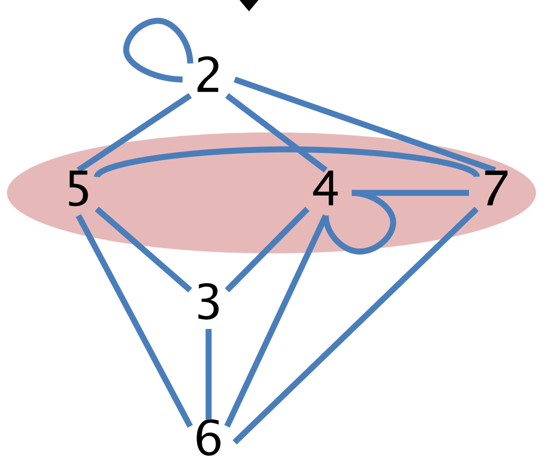
reconsider local/edge complementation



$\longleftrightarrow$

$$
\begin{array}{c|c}
 & \begin{matrix} 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\
\begin{matrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} &
\begin{pmatrix}
1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 0 & 1 & 1 & 0 \\
1 & 1 & 1 & 1 & 1 & 0 \\
0 & 1 & 1 & 1 & 0 & 1 \\
1 & 0 & 0 & 0 & 1 & 1
\end{pmatrix}
\end{array}
$$

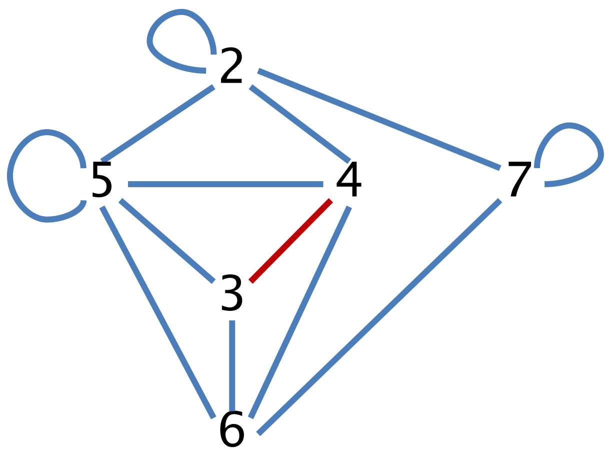
$\downarrow rc_2$ $\qquad$ $\downarrow rc_2$



$\longleftrightarrow$

$$
\begin{array}{c|c}
 & \begin{matrix} 2 & 3 & 4 & 5 & 6 & 7 \end{matrix} \\
\begin{matrix} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{matrix} &
\begin{pmatrix}
1 & 0 & 1 & 1 & 0 & 1 \\
0 & 0 & 1 & 1 & 1 & 0 \\
1 & 1 & 1 & 0 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 \\
0 & 1 & 1 & 1 & 0 & 1 \\
1 & 0 & 1 & 1 & 1 & 0
\end{pmatrix}
\end{array}
$$

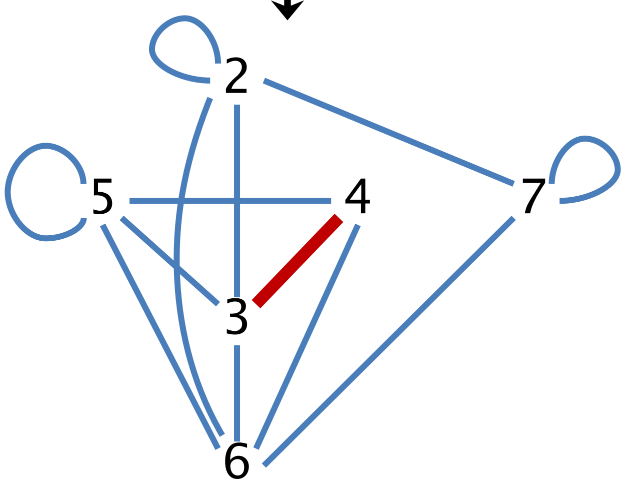
reconsider edge complementation



$$\longleftrightarrow$$

$$\begin{array}{c|cccccc} & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 2 & 1 & 0 & 1 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 & 1 & 1 & 0 \\ 4 & 1 & 1 & 0 & 1 & 1 & 0 \\ 5 & 1 & 1 & 1 & 1 & 1 & 0 \\ 6 & 0 & 1 & 1 & 1 & 0 & 1 \\ 7 & 1 & 0 & 0 & 0 & 1 & 1 \end{array}$$

$$\downarrow rc_{3,4} \qquad\qquad \downarrow rc_{3,4}$$



$$\longleftrightarrow$$

$$\begin{array}{c|cccccc} & 2 & 3 & 4 & 5 & 6 & 7 \\ \hline 2 & 1 & 1 & 0 & 0 & 1 & 1 \\ 3 & 1 & 0 & 1 & 1 & 1 & 0 \\ 4 & 0 & 1 & 0 & 0 & 1 & 1 \\ 5 & 0 & 1 & 0 & 1 & 1 & 0 \\ 6 & 1 & 1 & 1 & 1 & 0 & 1 \\ 7 & 1 & 0 & 1 & 0 & 1 & 1 \end{array}$$

principal pivot transform

$A * X$ is defined iff $A[X]$ is invertible/nonsingular/det $\neq 0$

$$A = \begin{array}{cc} & X \\ X & \left(\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right) \end{array}$$

$$A * X = \left(\begin{array}{c|c} P^{-1} & -P^{-1} Q \\ \hline R P^{-1} & S - R P^{-1} Q \end{array}\right)$$

partial inverse

$$A \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \quad \text{iff} \quad A{*}X \begin{pmatrix} x_2 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$$

$\leftarrow$ X pointers

$\leftarrow$ other

M.J. Tsatsomeros. Principal pivot transforms: properties and applications.
Linear Algebra and its Applications, 307(1-3):151-165, 2000

principal pivot transform

$A * X$ is defined iff $A[X]$ is invertible/nonsingular/det $\neq 0$

$$A = \begin{array}{cc} & X \\ X & \left(\begin{array}{c|c} P & Q \\ \hline R & S \end{array}\right) \end{array} \qquad A * X = \left(\begin{array}{c|c} P^{-1} & -P^{-1} Q \\ \hline R P^{-1} & S - R P^{-1} Q \end{array}\right)$$

$$\begin{array}{c c} & \begin{array}{cccccc} 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \left(\begin{array}{cccccc} 1 & 0 & 1 & 1 & 0 & 1 \\ 0 & 0 & 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 0 \\ 0 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 & 1 & 1 \end{array}\right) \end{array} \xrightarrow{rc_{3,4}} \begin{array}{c c} & \begin{array}{cccccc} 2 & 3 & 4 & 5 & 6 & 7 \end{array} \\ \begin{array}{c} 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \end{array} & \left(\begin{array}{cccccc} 1 & 1 & 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 1 & 1 & 0 \\ 0 & 1 & 0 & 0 & 1 & 1 \\ 0 & 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 0 & 1 & 1 \end{array}\right) \end{array}$$

J.F. Geelen, A generalization of Tutte's characterization of totally unimodular matrices, Journal of Combinatorial Theory, Series B 70 (1997) 101–117.

principal pivot transform

$$A \begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix} \quad \text{iff} \quad A{*}X \begin{pmatrix} x_2 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$$

using partial inversion

$$\boxed{(A * X) * Y = A * (X \overset{\text{xor}}{\oplus} Y)}$$

(when defined)

$$\boxed{\underbrace{A * \{p_1, p_2\} \ldots * p_n}_{\text{any sequence involving all pointers}} = A * V = A^{-1}}$$

(all pointers)

this shows

- how the rc_p and $rc_{p,q}$ interact
- result does not depend on order of operations

what *is* happening?

$34 2 \bar{3} \bar{2} 4$ $\qquad\qquad$ $3 \bar{2} 4 2 3 4$

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array} \xrightarrow{rc_3\ rc_4\ rc_2} \begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 0 & 1 \\ 3 & 0 & 0 & 1 \\ 4 & 1 & 1 & 0 \end{array}$$

multiply (over the binary numbers)

$$\underbrace{\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}}_{\text{micro}} \underbrace{\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}}_{\text{macro}} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

$+$	xor $\oplus$	$1+1=0$
$*$	and $\wedge$	

sorting DNA = computing the inverse

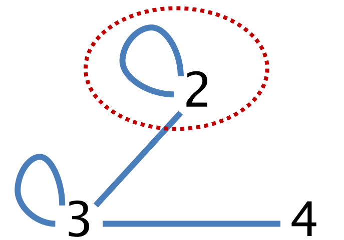
ppt is (partial) inverse

applicability

A * X is defined iff A[X] is invertible

rc_2

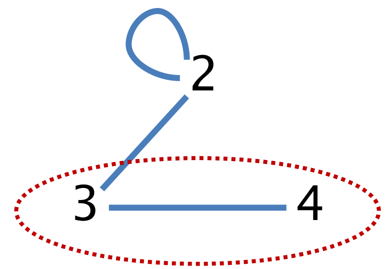
$3 4 2 \bar{3} \bar{2} 4$



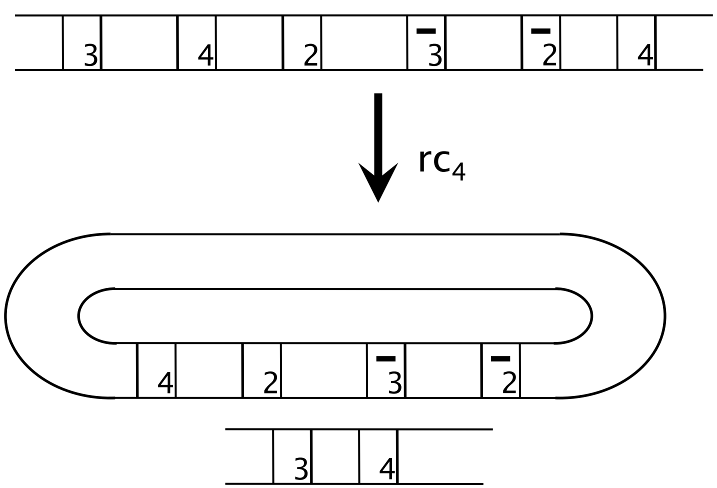
$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$

$rc_{3,4}$

$3 4 2 3 \bar{2} 4$

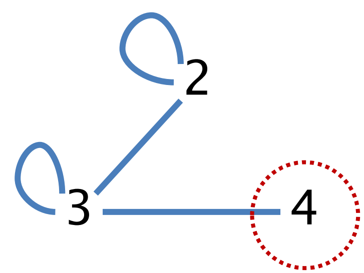


$$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$$



$3 4 2 \bar{3} \bar{2} 4$

undefined



$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & 1 & 1 & 0 \\ 3 & 1 & 1 & 1 \\ 4 & 0 & 1 & 0 \end{array}$$

conclusion (for now)

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt)
- we can use results about ppt to know more about gene assembly
  - independent order operations
  - interaction operations

quest for the “right” model

- strings
- graphs
- matrices
- set systems

the most elegant model was hidden

4

applicable sets

A * X is defined iff A[X] is invertible

$$\begin{array}{c|ccc} & 2 & 3 & 4 \\ \hline 2 & \multicolumn{3}{c}{\multirow{3}{*}{$\begin{pmatrix} 1 & 1 & 0 \\ 1 & \boxed{1} & \boxed{1} \\ 0 & \boxed{1} & \boxed{0} \end{pmatrix}$}} \\ 3 & & & \\ 4 & & & \end{array} \longleftrightarrow$$

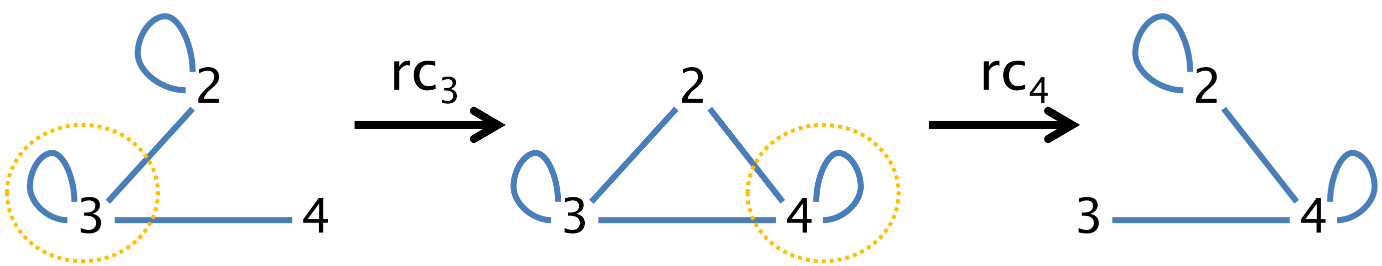
set system

$V = \{2,3,4\}$

$D = \{ \emptyset, \{2\}, \{3\}, \{2,4\}, \boxed{\{3,4\}}, \{2,3,4\} \}$

A[{3,4}]

*{3,4} is not $rc_{3,4}$



graphs $\subseteq$ set systems (strict)

$3\,4\,2\,\bar{3}\,\bar{2}\,4$

$$\begin{array}{c c} & \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{array}{c} 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix} \end{array}$$

$V = \{2,3,4\}$

$D = \{\ \emptyset, \{2\}, \{3\}, \{2,4\}, \{3,4\}, \{2,3,4\}\ \}$

rc_3

$\oplus 3$ xor 3

$3\,\bar{2}\,\bar{4}\,\bar{3}\,\bar{2}\,4$

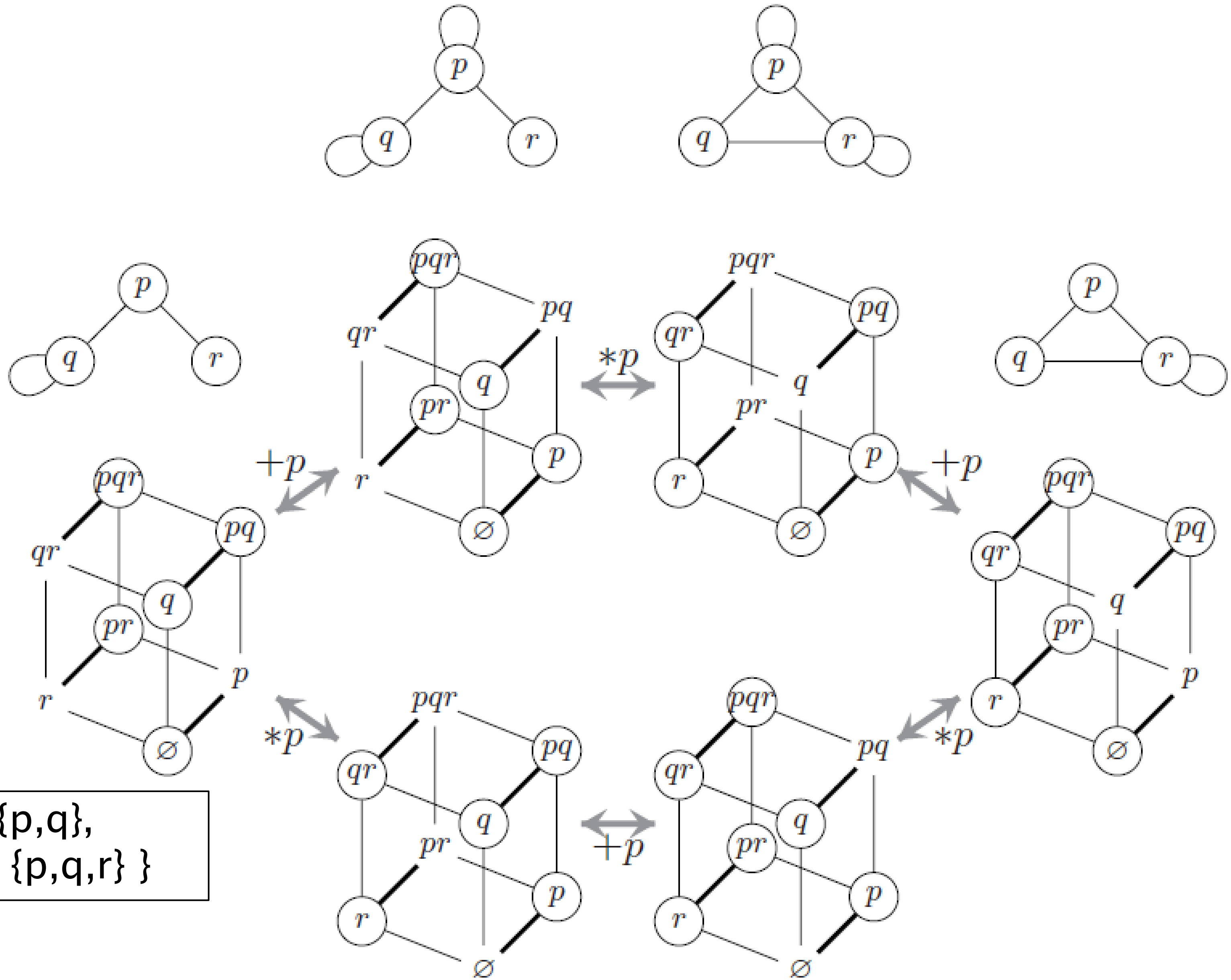
$$\begin{array}{c c} & \begin{array}{ccc} 2 & 3 & 4 \end{array} \\ \begin{array}{c} 2 \\ 3 \\ 4 \end{array} & \begin{pmatrix} 0 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{pmatrix} \end{array}$$

$V = \{2,3,4\}$

$D' = \{\ \{3\}, \{2,3\}, \emptyset, \{2,3,4\}, \{4\}, \{2,4\}\ \}$

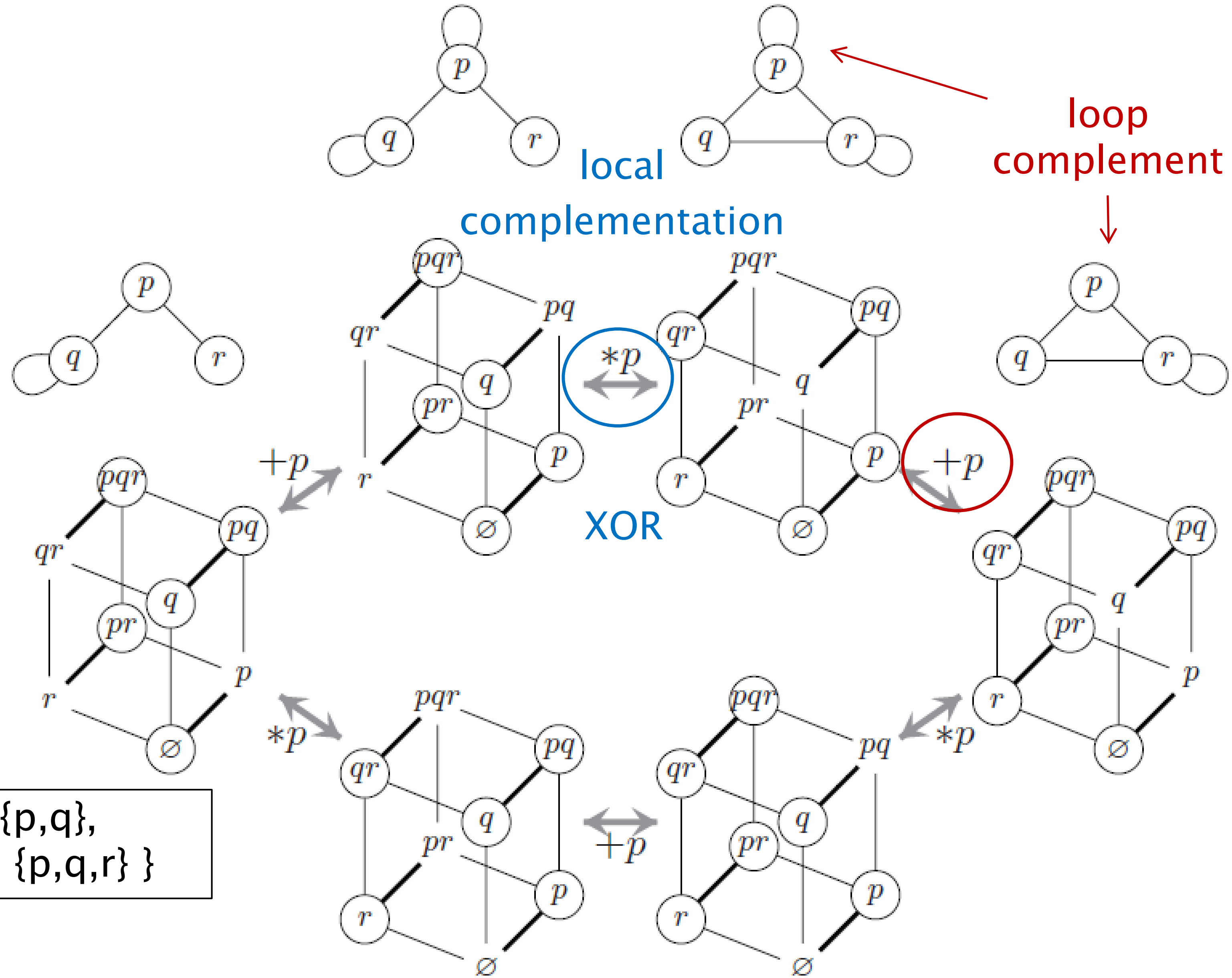
applicability (!) XOR {4} is defined,
while rc_4 is not, nb. {4} not in D

algebra of set systems



{ ∅, {q}, {p,q},
{p,r}, {p,q,r} }

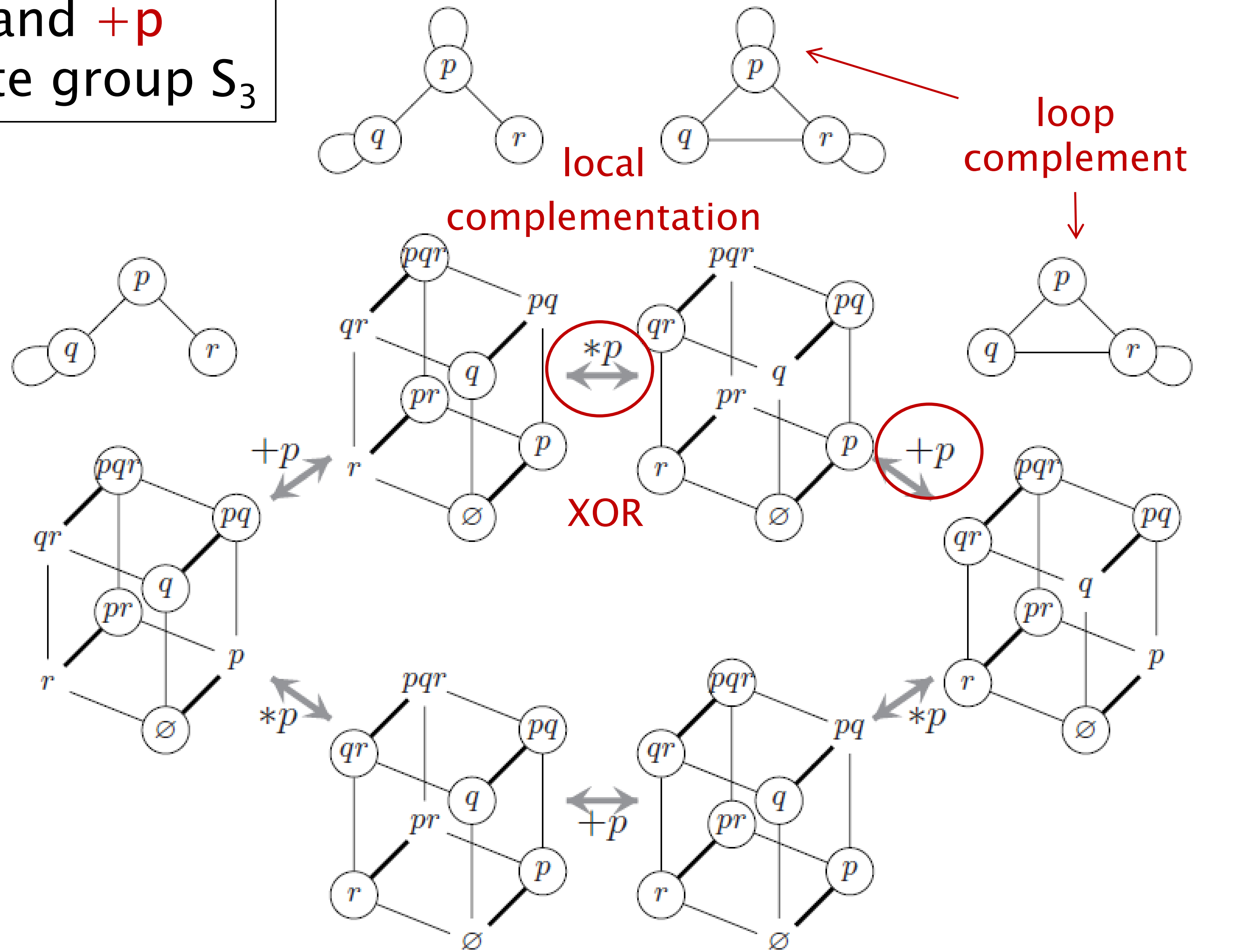
algebra of set systems



{ ∅, {q}, {p,q},
{p,r}, {p,q,r} }

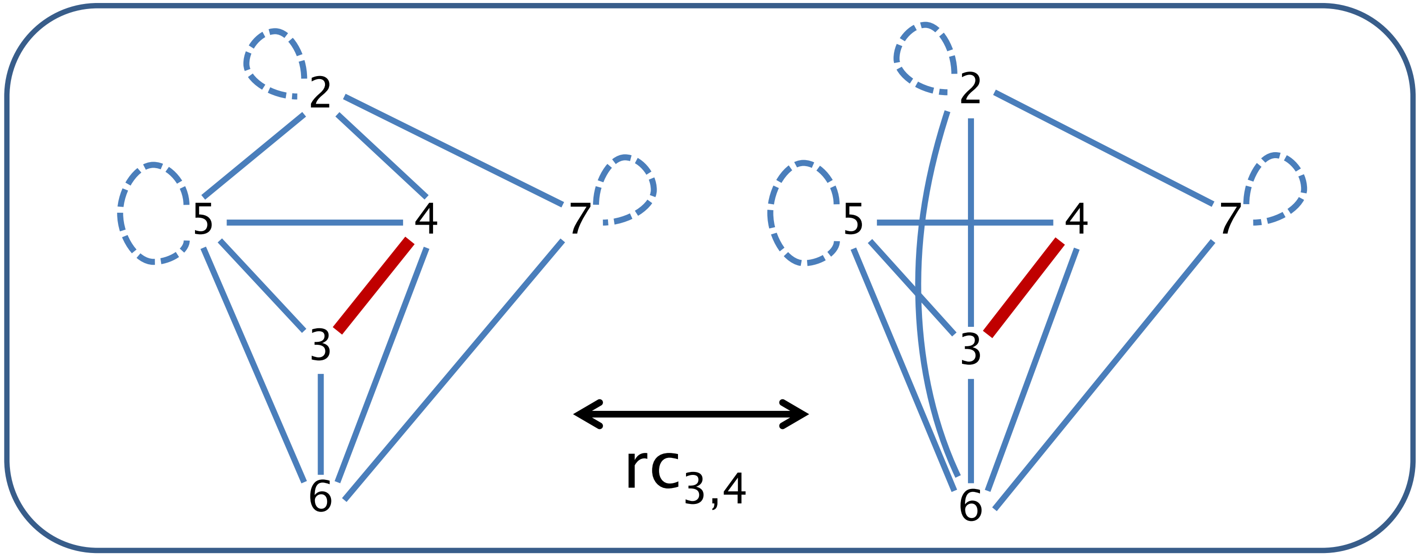
algebra of set systems

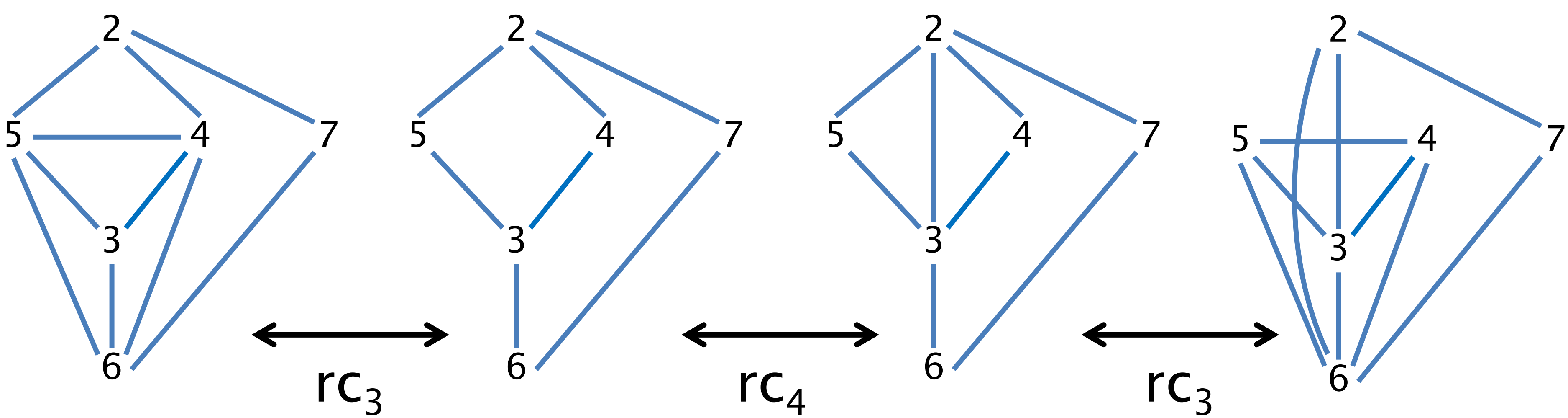
*p and +p
generate group S_3



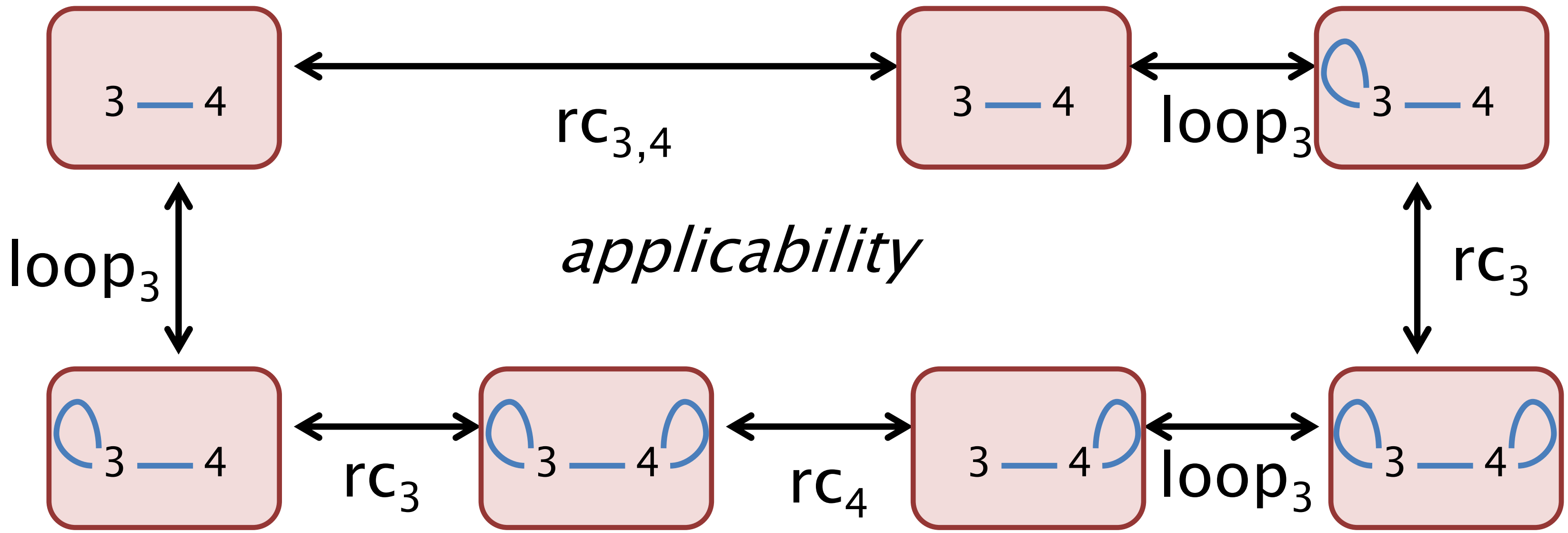
edge complement vs. local complement

ignoring loops





edge complement vs. local complement



basic algebra S_3

$*3 \; *4 = *4 \; *3$

$*3 \; *3 = \mathrm{id} = +3 \; +3$

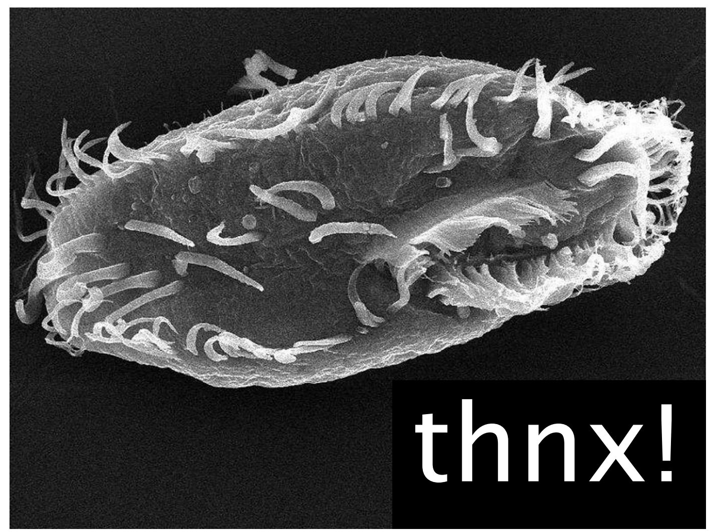
$+3 \; *3 \; +3 = *3 \; +3 \; *3$

$+3 \; *3 \; *4 \; +3 \; *3 \; +3 =$
$+3 \; *3 \; +3 \; *3 \; +3 \; *4 =$
$+3 \; *3 \; *3 {+} 3 \; *3 \; *4 =$
$+3 \; +3 \; *3 \; *4 =$
$*3 \; *4 =$
$*\{3,4\}$

- by careful modeling we find that gene assembly is *actually* principal pivot transform (ppt) and XOR
- we can use results about ppt (on matrices) and XOR (on set systems) to know more about gene assembly
- but also inspiration the other way around …

however …

- parallellism
- ‘simple’ operations



references (to self)

recombination is ppt

R. Brijder, T. Harju, H.J. Hoogeboom, Pivots, determinants, and perfect matchings of graphs (2008) [arXiv:0811.3500] finally to appear.

(independently) Nathan Pflueger: Graph reductions, binary rank, and pivots in gene assembly. DAM 159 (2011) 2117–2134.

*p +p algebra of operations

R. Brijder, H.J. Hoogeboom. The Group Structure of Pivot and Loop Complementation on Graphs and Set Systems. Eur.J.Comb. (2011).

general

R. Brijder, H.J. Hoogeboom. Reality-and-Desire in Ciliates. In: *Algorithmic Bioprocesses* (Condon etal, eds.), Natural Computing Series, Springer (2009) pp.99–115.

A. Ehrenfeucht, T. Harju, I. Petre, D. Prescott, G. Rozenberg, *Computation in Living Cells: Gene Assembly in Ciliates*, Natural Computing Series, Springer (2004).