Colloquium · USF · Tampa Jan'16

Graph Polynomials motivated by Gene Assembly

Hendrik Jan Hoogeboom, Leiden NL
with Robert Brijder, Hasselt B

assembly polynomial of G_w for doc-word w

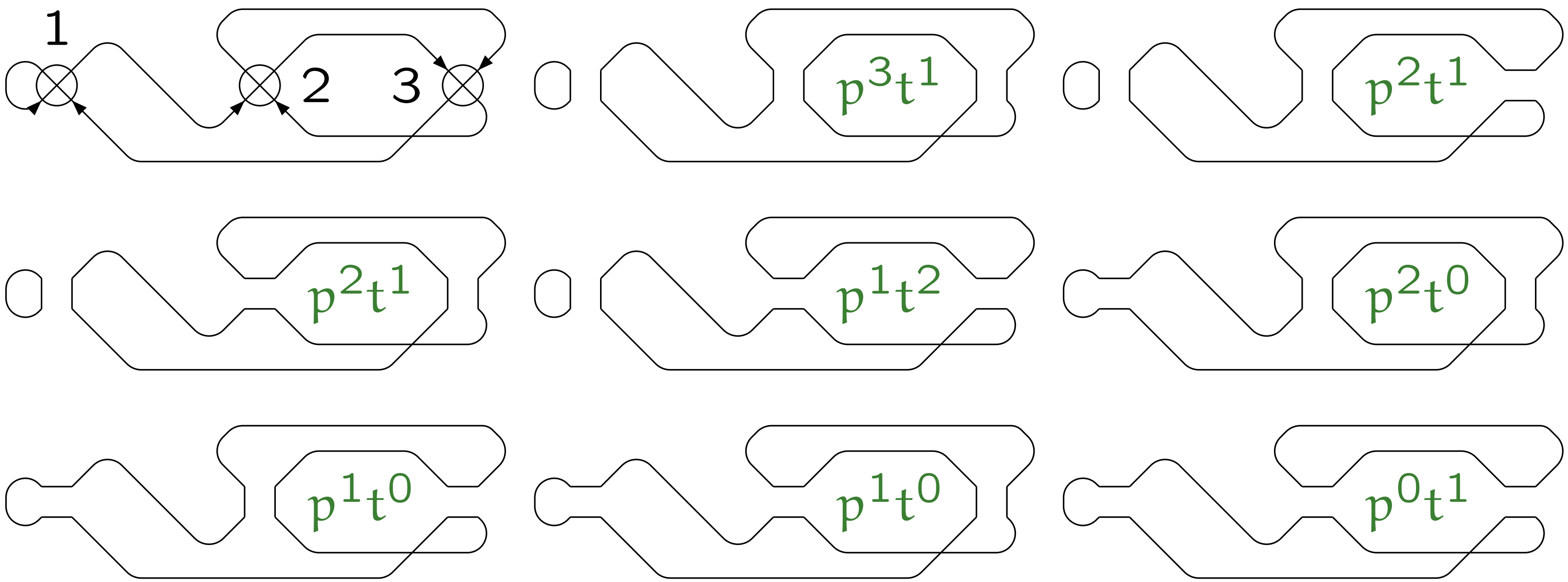
$$S(G_w)(p,t) = \sum_s p^{\pi(s)} t^{c(s)-1},$$

$\otimes$ follow / consistent π / inconsistent

never p

$w = 1\,1\,2\,3\,2\,3$

$p^3t + 2p^2t + p^2 + pt^2 + 2p + t$



Burns, Dolzhenko, Jonoska, Muche, Saito:
Four-regular graphs with rigid vertices associated to DNA recombination (2013)

motivation: models for gene recombination

- ciliates:

  4-regular graphs + Euler tours
- graph polynomials

concept: interlacing $p\,q\,p\,q$

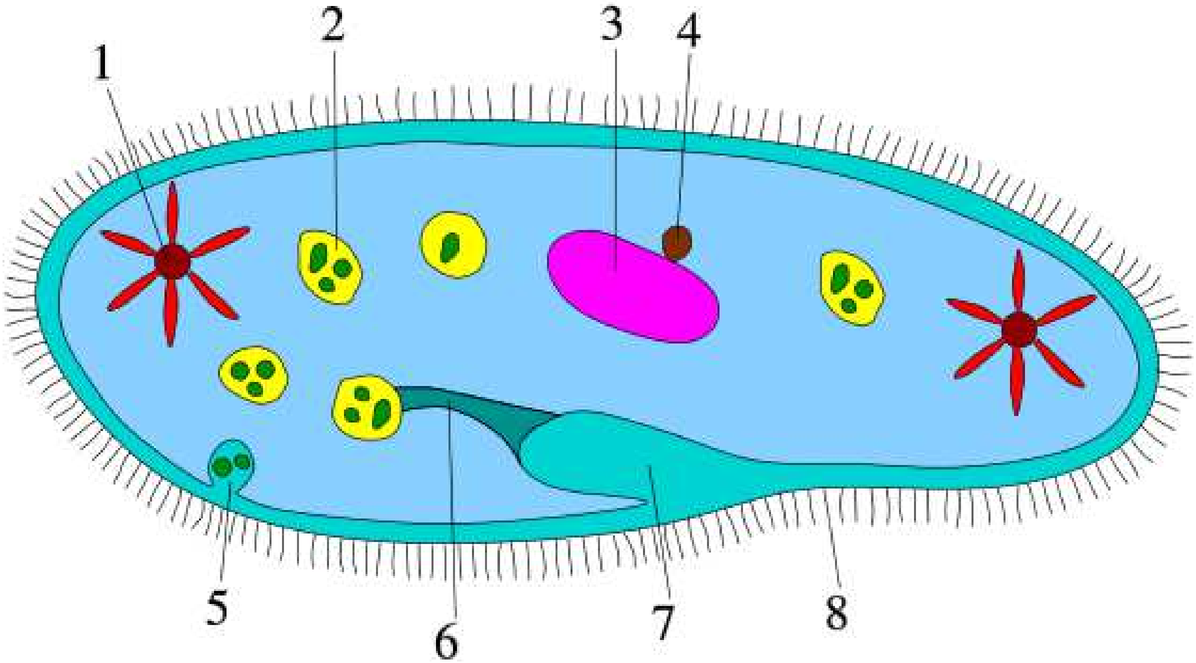
- four worlds, getting more abstract
  - strings – rewriting
  - graphs – combinatorics
  - matrices – linear algebra
  - (set systems) – symm diff ‘XOR’

‘same’ operation (pivot), different tools

Ciliates

cell structure:

3. macronucleous
4. micronucleous
8. cilium



wikipedia Franciscosp2

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*.

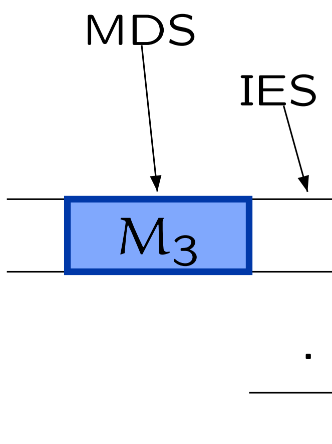
Ciliates: two types of nucleus

gene assembly: splicing and recombination

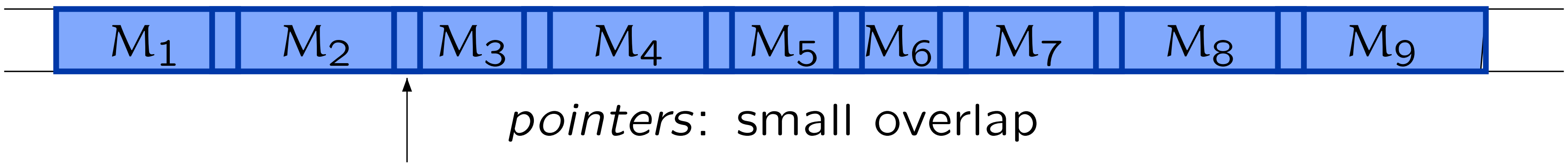
MIC *micronucleus*

MDS macronucleus destined M_i

IES internal eliminated



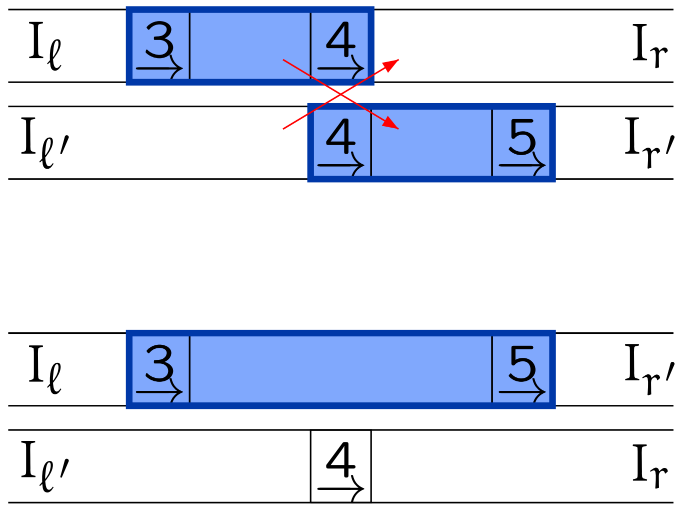
MAC *macronucleus*

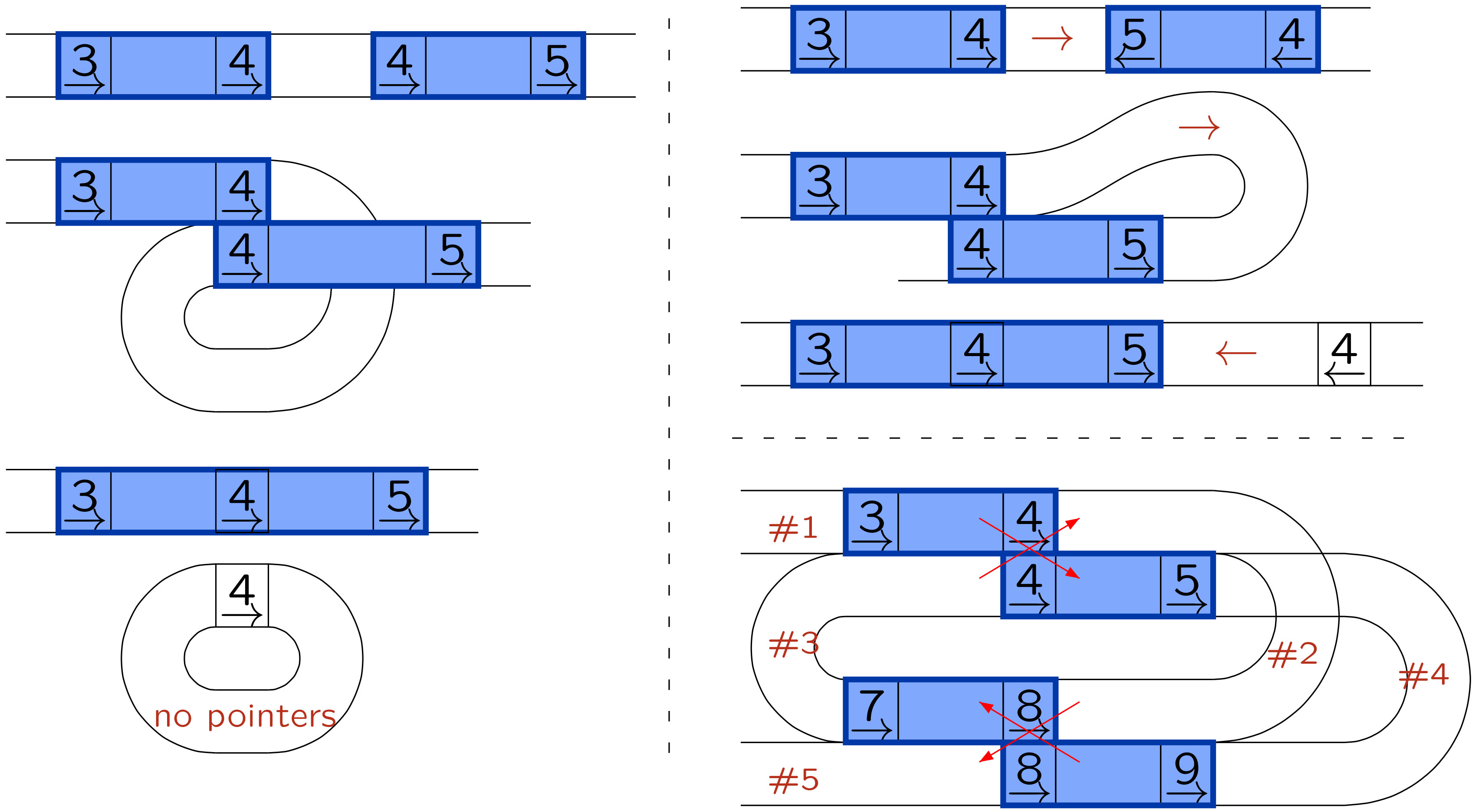


pointers: small overlap

merge consecutive MDS's

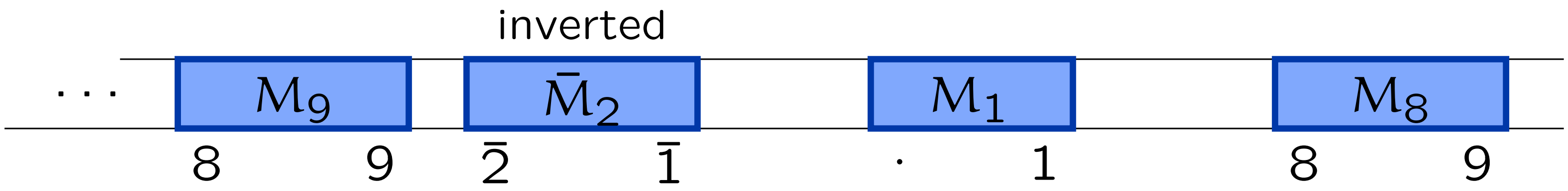
align pointers & swap





Ehrenfeucht, Harju, Petre, Prescott, Rozenberg:
Computation in Living Cells – Gene Assembly in Ciliates (2004)

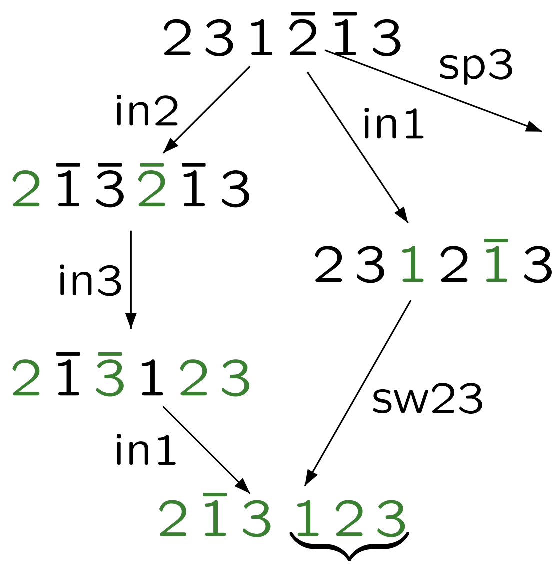
goal: sorting [deleting] pointers



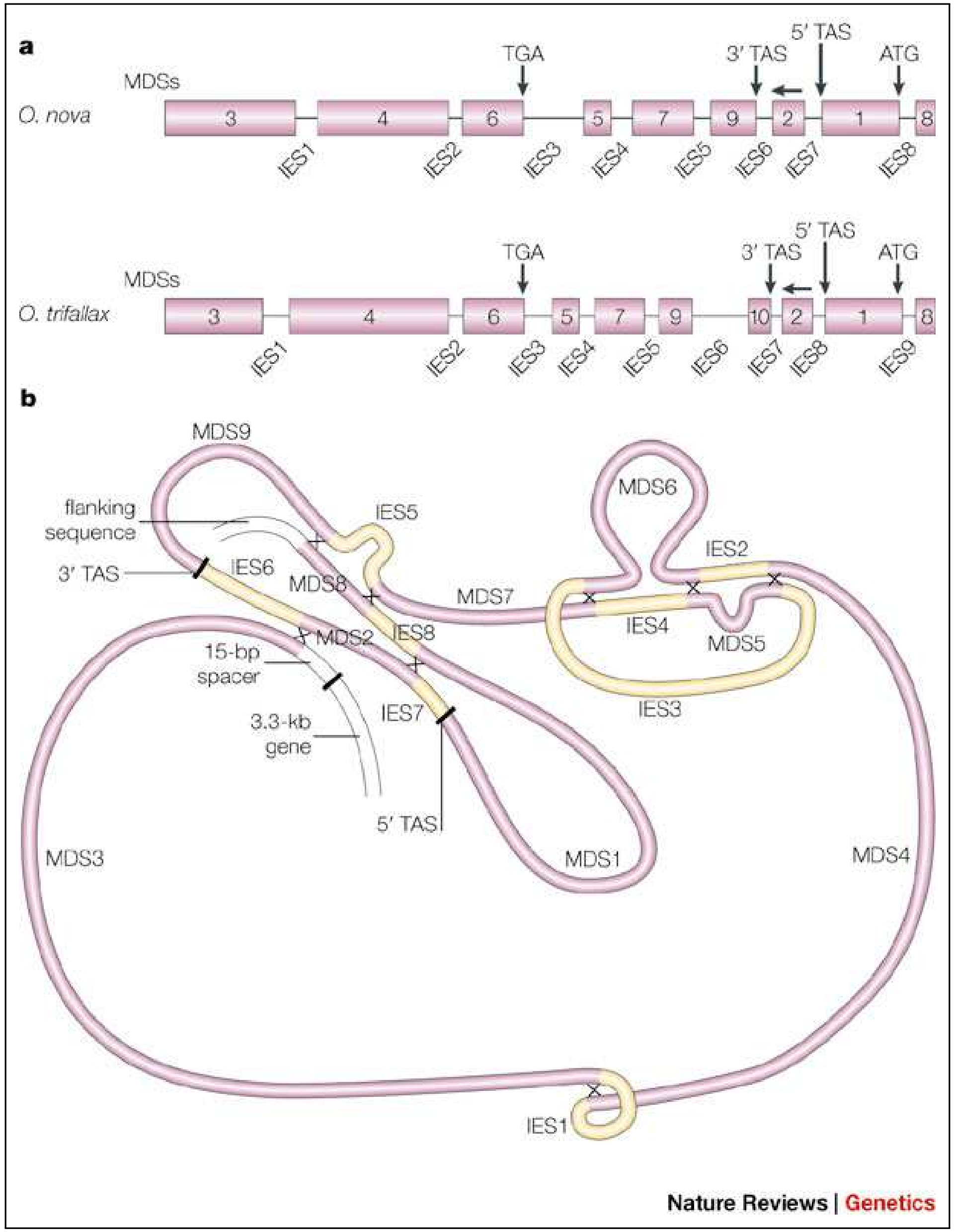
split $u_1\, p\, p\, u_2 \xrightarrow{p} u_1\, p\, p\, u_2$

invert $u_1\, p\, u_2\, \bar{p}\, u_3 \xrightarrow{p} u_1\, p\, \bar{u}_2\, \bar{p}\, u_3$

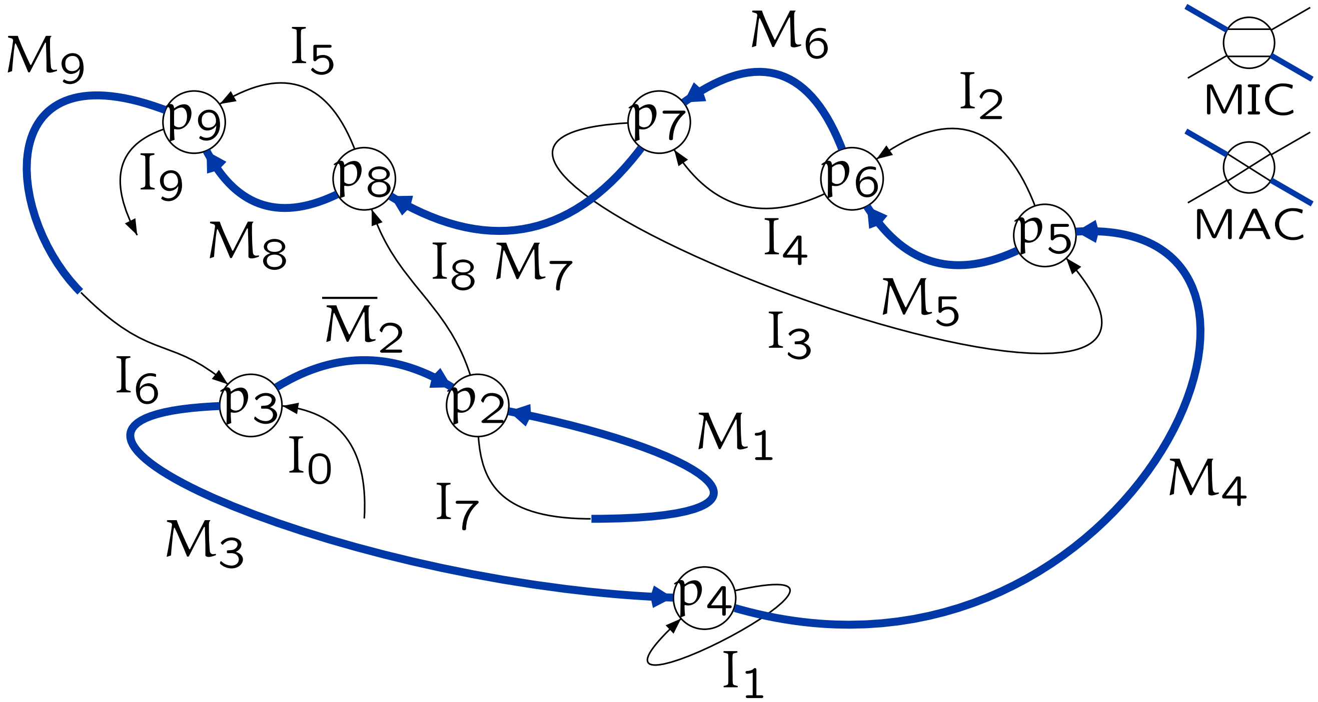
swap $u_1\, p\, u_2\, q\, u_3\, p\, u_4\, q\, u_5 \xrightarrow{p,q} u_1\, p\, u_4\, q\, u_3\, p\, u_2\, q\, u_5$



▷ result depends on operations?



David M. Prescott. Genome gymnastics: unique modes of dna evolution and processing in ciliates. Nature Reviews Genetics (December 2000)



MIC $\mathrm{I}_0\, \mathrm{M}_3\, \mathrm{I}_1\, \mathrm{M}_4\, \mathrm{I}_2\, \mathrm{M}_6\, \mathrm{I}_3\, \mathrm{M}_5\, \mathrm{I}_4\, \mathrm{M}_7\, \mathrm{I}_5\, \mathrm{M}_9$

$\mathrm{I}_6\, \overline{\mathrm{M}}_2\, \mathrm{I}_7\, \mathrm{M}_1\, \mathrm{I}_8\, \mathrm{M}_8\, \mathrm{I}_9$

MAC $\bar{\mathrm{I}}_9 \bar{\mathrm{I}}_5 \bar{\mathrm{I}}_8 \mathrm{I}_7 \underbrace{\mathrm{M}_1 \mathrm{M}_2 \cdots \mathrm{M}_8 \mathrm{M}_9} \mathrm{I}_6 \bar{\mathrm{I}}_0$, I_1 and $\mathrm{I}_2\, \mathrm{I}_4\, \mathrm{I}_3$

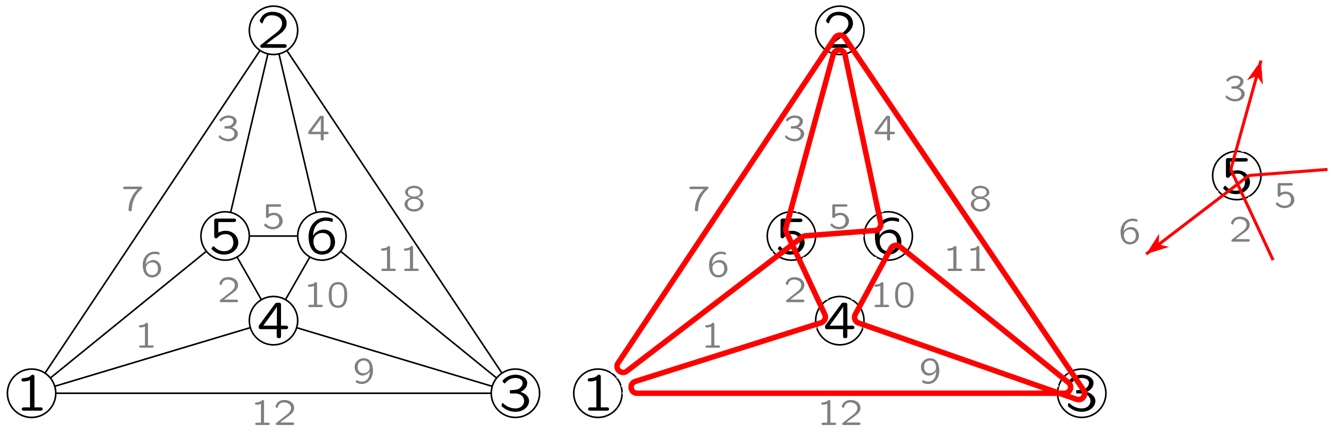
4-regular graph with Euler circuit

double occurrence string w defines

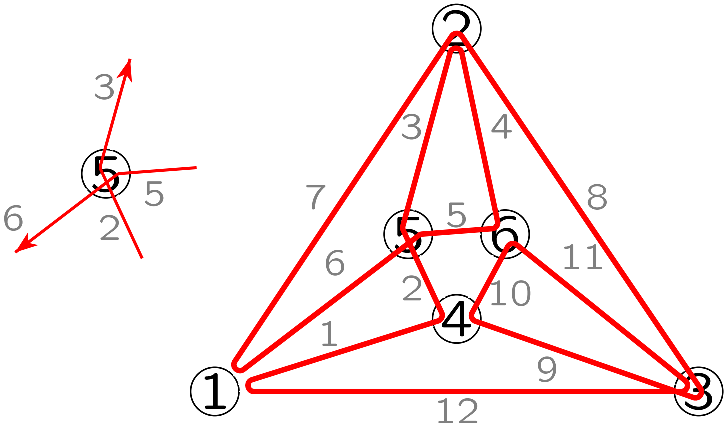
4-regular graph G_w + Euler circuit C_w

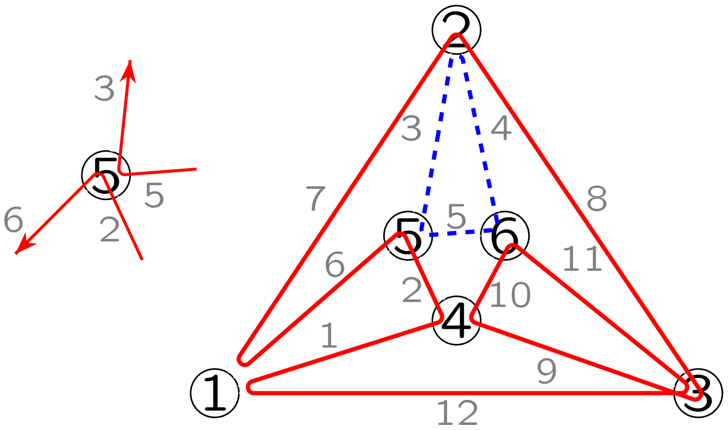
or 2-in 2-out graph + directed circuit

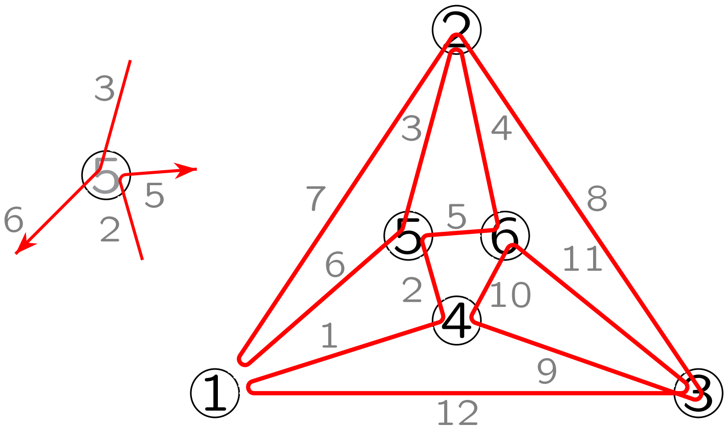
$w = 1\,4\,5\,2\,6\,5\,1\,2\,3\,4\,6\,3$



$w = 1\,4\,\mathbf{5}\,2\,6\,\mathbf{5}\,1\,2\,3\,4\,6\,3$







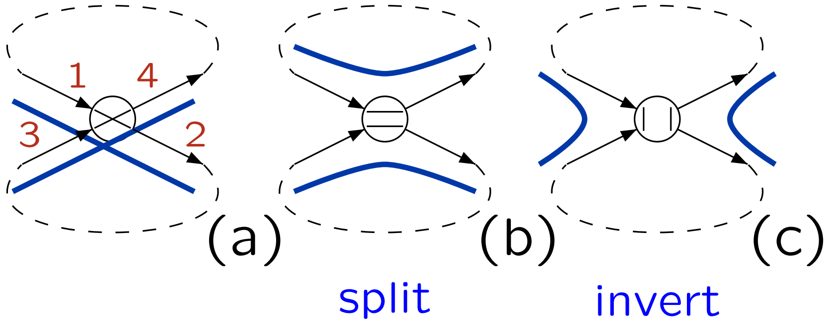
segment split

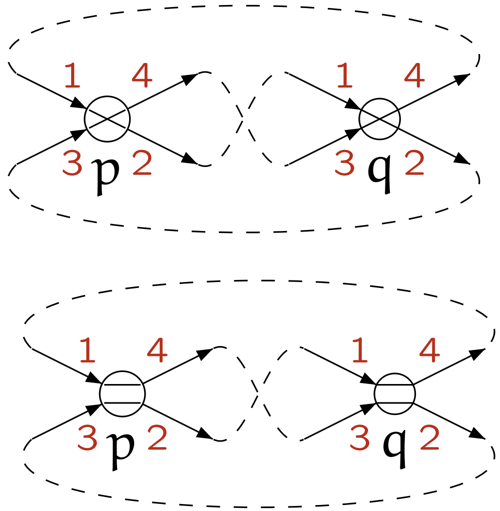
segment inverted

1 4 **5** 1 2 3 4 6 3 & **5** 2 6

1 4 **5** $\underline{6\;2}$ **5** 1 2 3 4 6 3

(a) follows C
(b) orientation consistent
(c) orientation inconsistent





interlaced $\ldots\underbrace{{}^{1}\mathrm{p}^{2}\ldots{}^{1}\mathrm{q}^{2}}\ldots\underbrace{{}^{3}\mathrm{p}^{4}\ldots{}^{3}\mathrm{q}^{4}}\ldots$

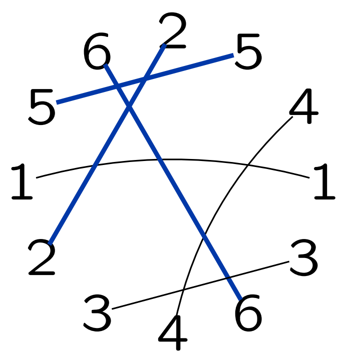
$\ldots\underbrace{{}^{1}\mathrm{p}^{4}\ldots{}^{3}\mathrm{q}^{2}}\ldots\underbrace{{}^{3}\mathrm{p}^{2}\ldots{}^{1}\mathrm{q}^{4}}\ldots$

segments are swapped

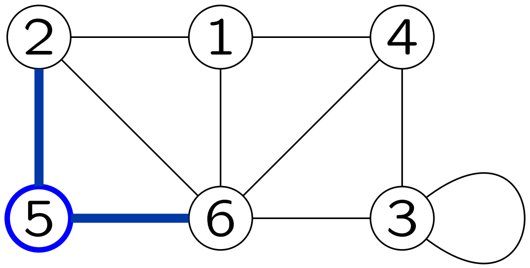
Kotzig. Eulerian lines in finite 4-valent graphs (1966) rearrangements should not break genome

$$\ldots \underbrace{p \ldots q} \ldots \underbrace{p \ldots q} \ldots$$

interlace graph $\mathrm{I}(\mathrm{C}_w)$

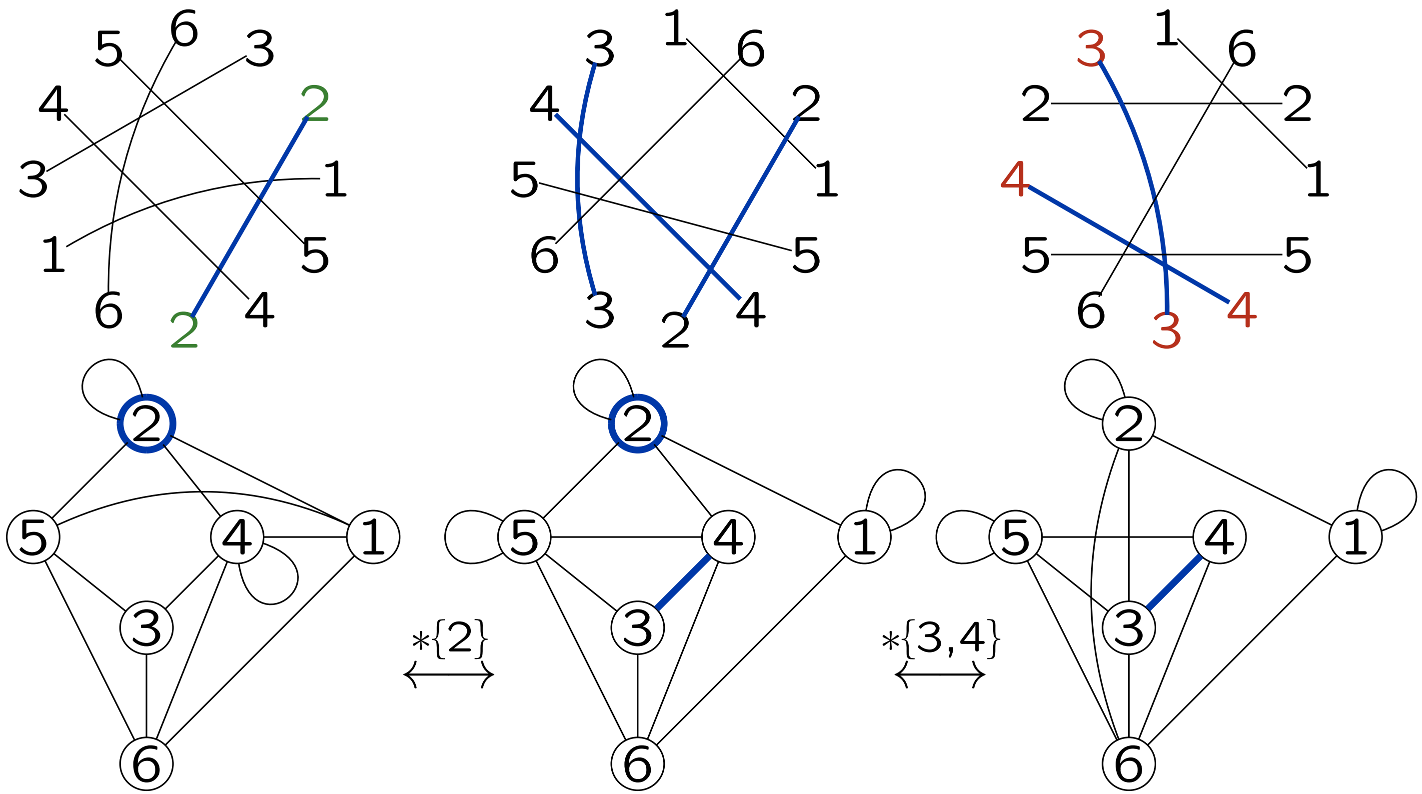


$w = 1\,4\,\underline{\mathbf{5}\,2\,6\,\mathbf{5}}\,1\,2\,\bar{3}\,4\,6\,3$



circle graph (bar + loop for orientation) circular string

interlace



doc string

$w * \{2\}$ $\quad$ $w = 1\,2\,6\,\bar{1}\,3\,4\,\bar{5}\,6\,3\,\bar{2}\,4\,5$ $\quad$ $w * \{3,4\}$

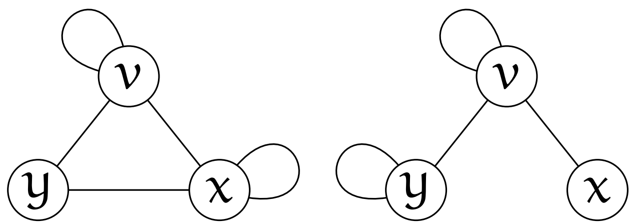
$1\,2\,\bar{3}\,\bar{6}\,5\,\bar{4}\,\bar{3}\,1\,\bar{6}\,\bar{2}\,4\,5$ $\quad$ $1\,2\,6\,\bar{1}\,3\,\bar{2}\,4\,\bar{5}\,6\,3\,4\,5$

invert $\quad$ swap

$$\begin{array}{rcl} \text{4-regular graph} & & \text{(circle graph)} \\ + \text{ Euler cycle} & & \text{interlace graph} \\ C & \longrightarrow & I(C) \\ \text{invert} \downarrow & & \downarrow \text{local complement} \\ C * v & \longrightarrow & I(C) * v \end{array}$$

$\dots \bar{x} \dots v \overrightarrow{\dots y \dots x \dots} \bar{v} \dots y$ interleaved

$\dots \bar{x} \dots v \overleftarrow{\dots \bar{x} \dots \bar{y} \dots} v \dots y$ not interleaved

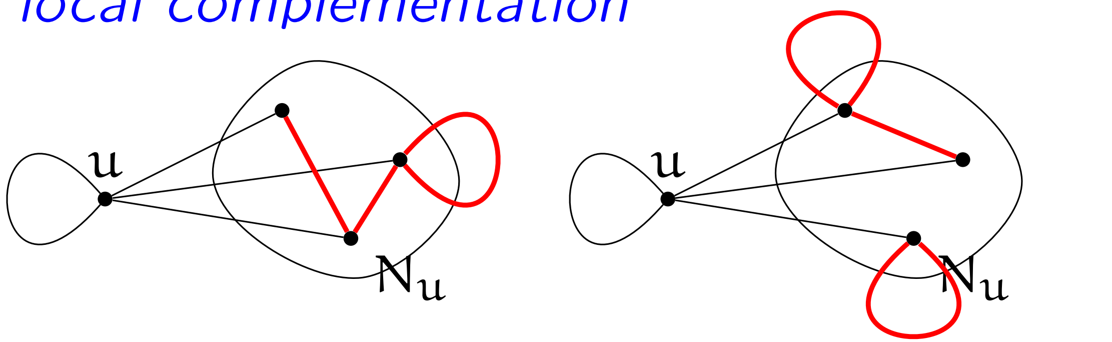


$G \mapsto G * u$

local complementation

looped vertex u

$N_u = N_G(u) \setminus \{u\}$



$G \mapsto G * \{u, v\}$

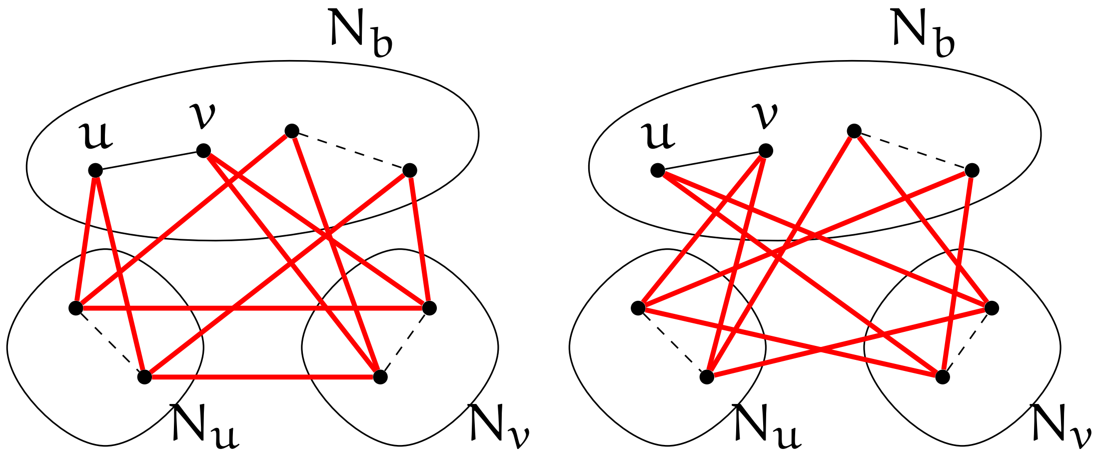
edge complementation

unlooped edge $\{u, v\}$

$N_u = N_G(u) \setminus N_G(v)$

$N_v = N_G(v) \setminus N_G(u)$

$N_b = N_G(u) \cap N_G(v)$

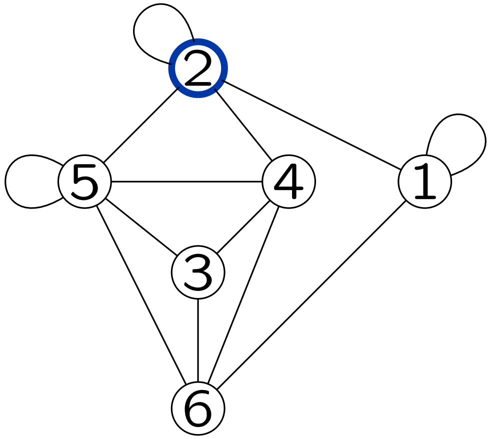


invert $I(C * u) = I(C) * v$

swap $I(C * \{u, v\}) = I(C) * \{u, v\}$ when defined how do these operations interact?

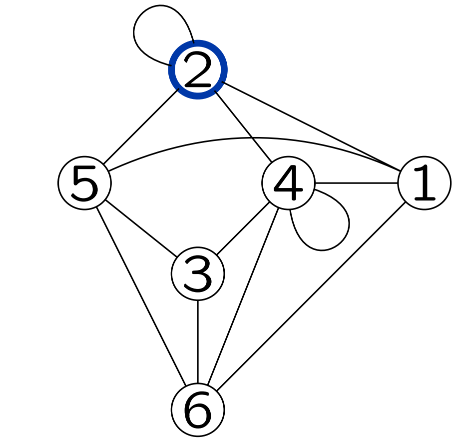
dependent on (order) operations chosen?

what are the intermediate products?



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

$\downarrow *\{2\}$ $\qquad$ $\downarrow$?? PPT



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

$$A = \begin{array}{c c} & \begin{array}{cc} X & V\backslash X \end{array} \\ \begin{array}{c} X \\ V\backslash X \end{array} & \begin{pmatrix} P & Q \\ R & S \end{pmatrix} \end{array}$$

$$A * X = \begin{array}{c c} & \begin{array}{cc} X & V\backslash X \end{array} \\ \begin{array}{c} X \\ V\backslash X \end{array} & \begin{pmatrix} P^{-1} & -P^{-1}Q \\ RP^{-1} & S - RP^{-1}Q \end{pmatrix} \end{array}$$

partial inverse

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

PPT matches edge & local complement

$$P = \begin{array}{c c} & p \\ p & \begin{pmatrix} 1 \end{pmatrix} \end{array}, \; P = \begin{array}{c c} & \begin{array}{cc} p & q \end{array} \\ \begin{array}{c} p \\ q \end{array} & \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \end{array}$$

Geelen. A Generalization of Tutte's Characterization of Totally Unimodular Matrices. JCTB (1997)

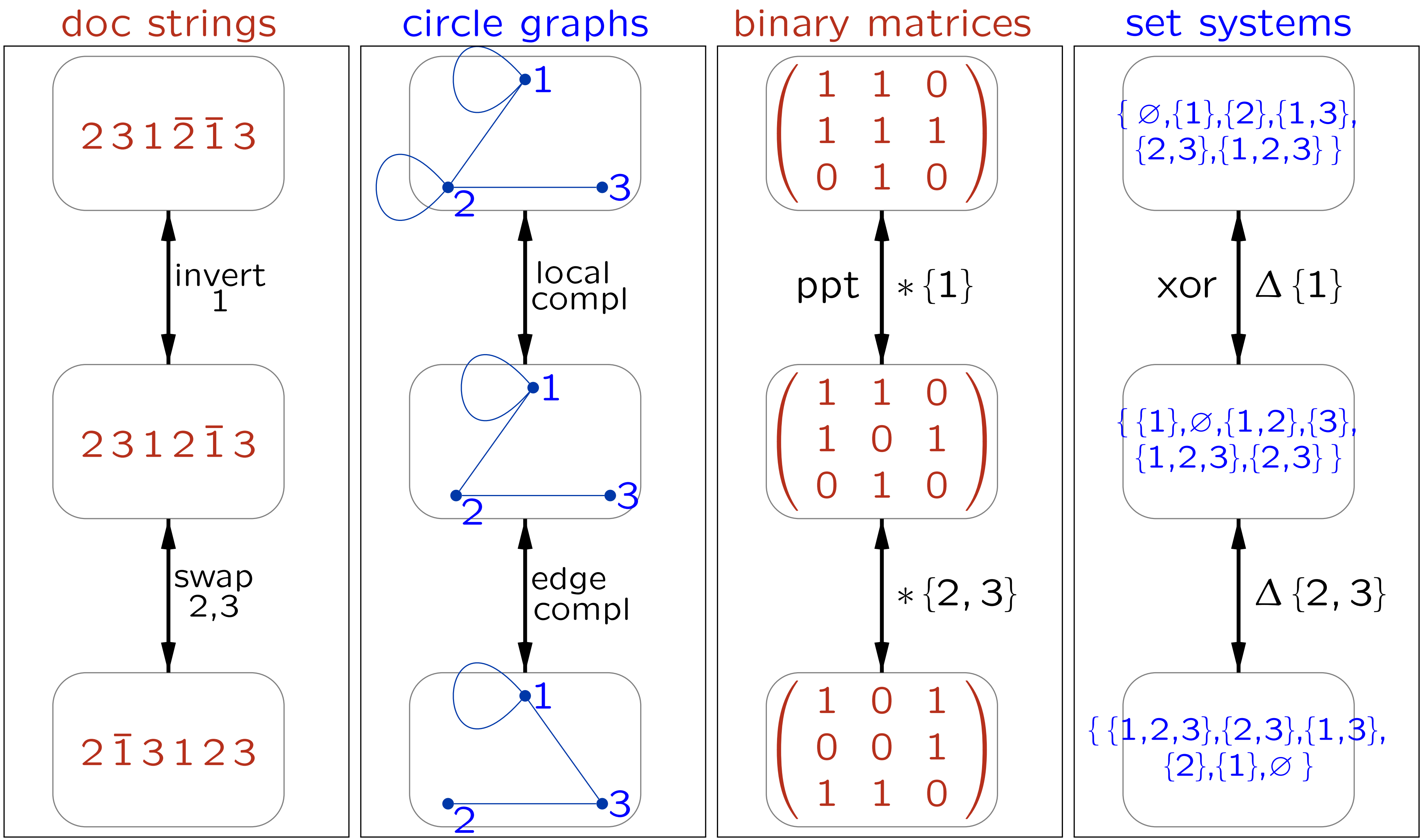
(partial inverse) $A\begin{pmatrix} x_1 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_2 \\ y_2 \end{pmatrix}$ iff $A * X\begin{pmatrix} x_2 \\ y_1 \end{pmatrix} = \begin{pmatrix} x_1 \\ y_2 \end{pmatrix}$

Thm. $(A * X) * Y = A * (X \,\Delta\, Y)$ when defined

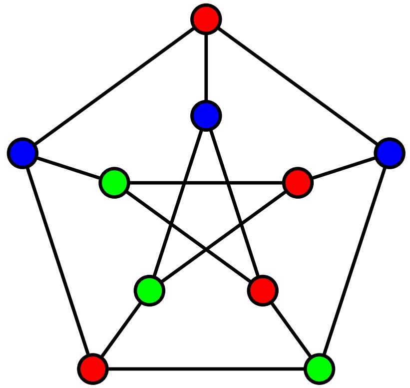
symmetric difference

any sequence involving all pointers:
$A * \{p_1, p_2\} \cdots * \{p_n\} = A * V = A^{-1}$

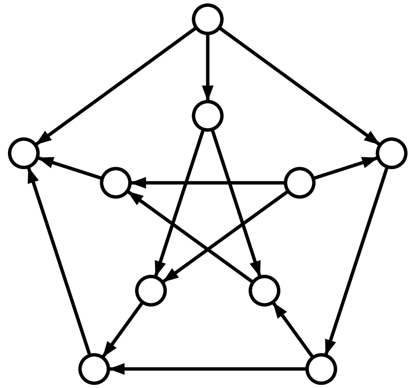
Cor. does not depend on order of operations (!)


doc strings
circle graphs
binary matrices
set systems
2 3 1 $\bar{2}$ $\bar{1}$ 3
invert 1
2 3 1 2 $\bar{1}$ 3
swap 2,3
2 $\bar{1}$ 3 1 2 3
1
2
3
local compl
edge compl
$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 1 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
ppt
$* \{1\}$
$\begin{pmatrix} 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{pmatrix}$
$* \{2, 3\}$
$\begin{pmatrix} 1 & 0 & 1 \\ 0 & 0 & 1 \\ 1 & 1 & 0 \end{pmatrix}$
$\{ \varnothing,\{1\},\{2\},\{1,3\},\{2,3\},\{1,2,3\} \}$
xor
$\Delta \{1\}$
$\{ \{1\},\varnothing,\{1,2\},\{3\},\{1,2,3\},\{2,3\} \}$
$\Delta \{2, 3\}$
$\{ \{1,2,3\},\{2,3\},\{1,3\},\{2\},\{1\},\varnothing \}$

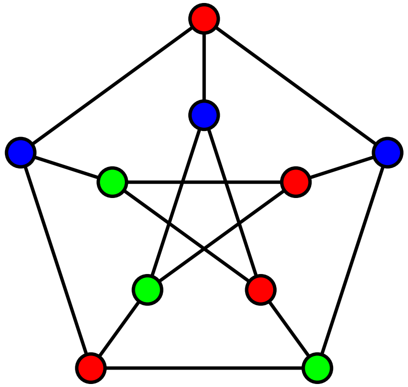
Graph Polynomials

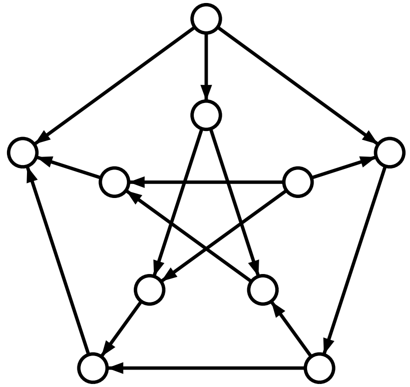


3 colours



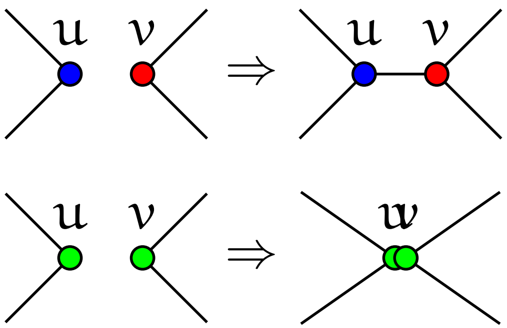
16680 acyclic orientations





chromatic polynomial

$$\chi_G(t) = t(t-1)(t-2) \cdot (t^7 - 12t^6 + 67t^5 \cdots - 230t^4 + 529t^3 - 814t^2 + 775t - 352)$$



$$\chi_G(t) = \chi_{G+uv}(t) + \chi_{G/uv}(t)$$

$$\chi_G(t) = \chi_{G-e}(t) - \chi_{G/e}(t)$$

deletion & contraction

acyclic orientations

$$16680 = (-1)^{|V_G|}\chi_G(-1)$$

definitions

recursive / closed form

combinatorial / algebraic

evaluations

combinatorial interpretation

polynomials

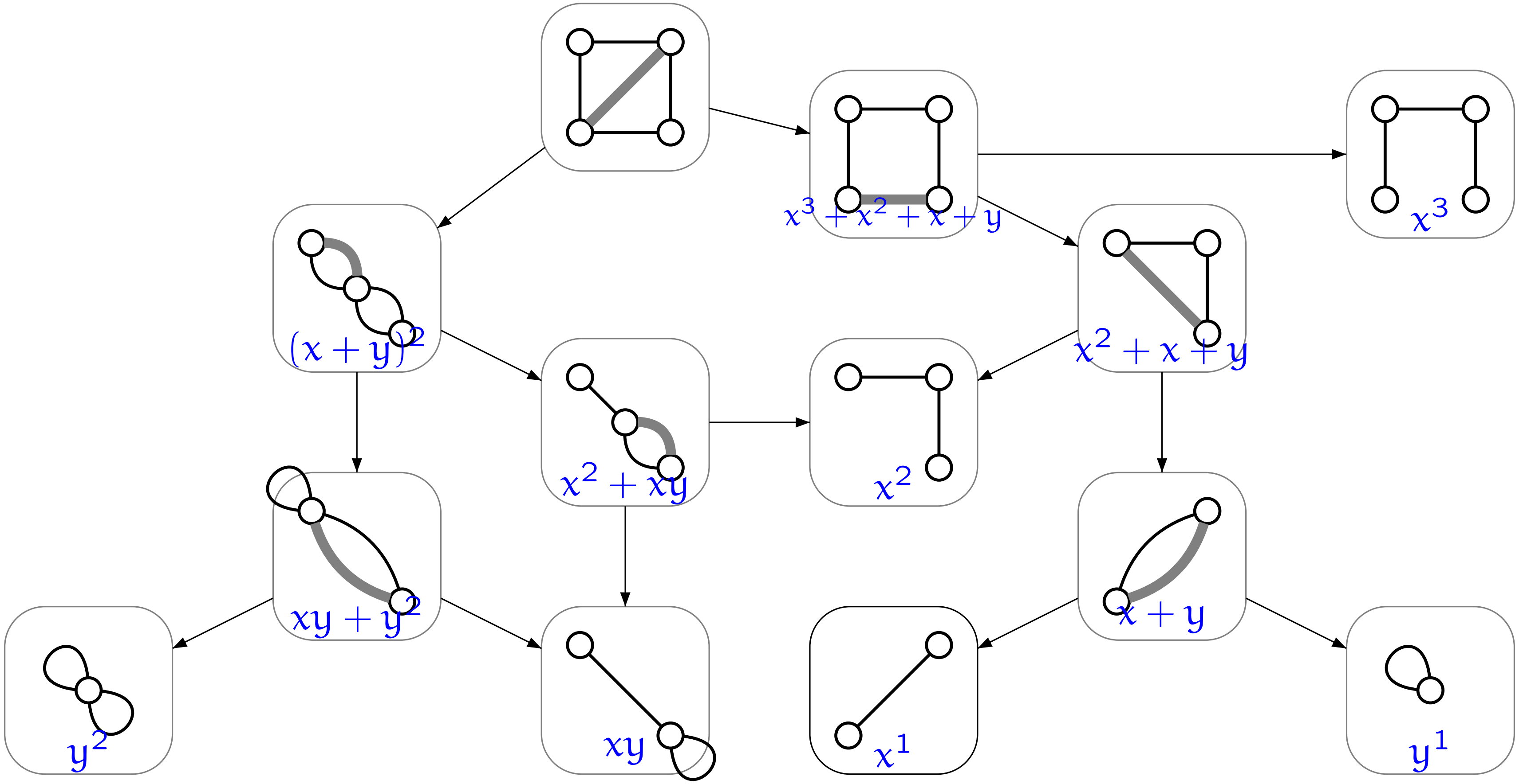
- Tutte (*one to rule them all...*)
- Martin (on 2-in 2-out graphs)
- assembly (Ciliates) ‘transition pol’
- interlace (DNA reconstruction)

and their relations!

loops and parallel edges

diamond graph:

$T_D = x^3 + 2x^2 + 2xy + x + y^2 + y$



$$G = (V, E),\ G[A] = (V, A)$$

$k(A)$ connected components in $G[A]$

$$T_G(x, y) = \sum_{A \subseteq E} (x-1)^{k(A)-k(E)} (y-1)^{k(A)+|A|-|V|}$$

rank nullity 'circuit rank'

$$T_G(x, y) = \begin{cases} 1 & \text{no edges} \\ x\, T_{G/e}(x, y) & \text{bridge } e \\ y\, T_{G-e}(x, y) & \text{loop } e \\ T_{G-e}(x, y) + T_{G/e}(x, y) & \text{other} \end{cases}$$

deletion & contraction

$$T_G(x, y) = x^i y^j \qquad \text{with } i \text{ bridges and } j \text{ loops}$$

extended to matroids

recipe theorem:

deletion-contraction implies Tutte evaluation (Tutte-Grothendieck invariant)

$$\chi_G(t) = (-1)^{|V|-c(G)}\, t^{c(G)}\, T_G(1-t, 0)$$

$$T_D = x^3 + 2x^2 + 2xy + x + y^2 + y$$
$$\chi_D(t) = t(t-1)(t-2)^2$$

(check Wolfram alpha)

evaluations:

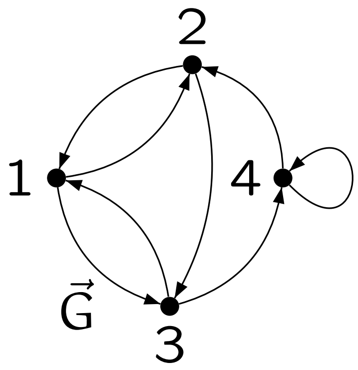
$T_G(2,0)$ acyclic orientations $= (-1)^{|V|}\chi_G(-1)$

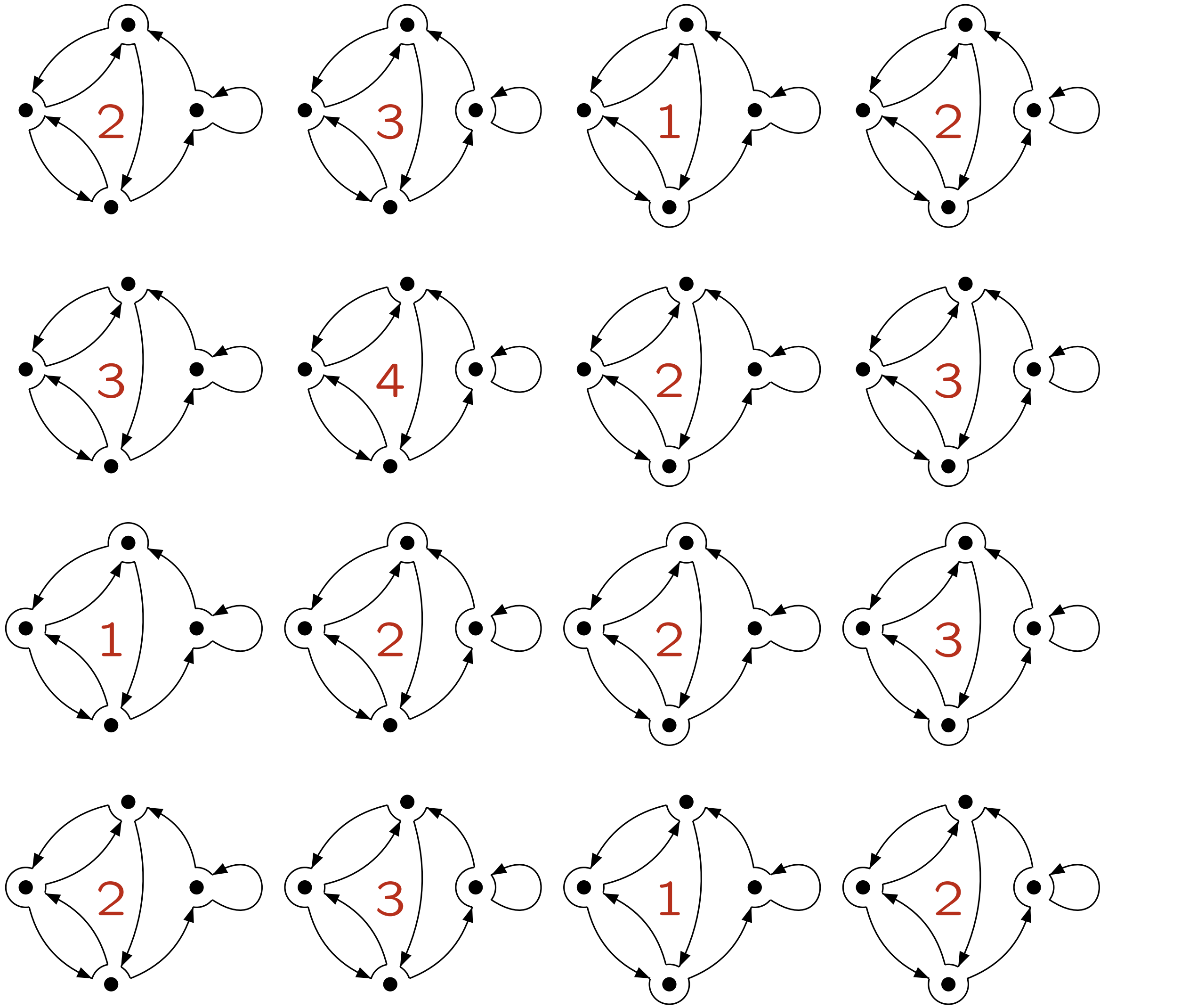
$T_G(2,1)$ forests

$T_G(1,1)$ spanning forests

$T_G(1,2)$ spanning subgraphs

$T_G(0,2)$ strongly connected orientations



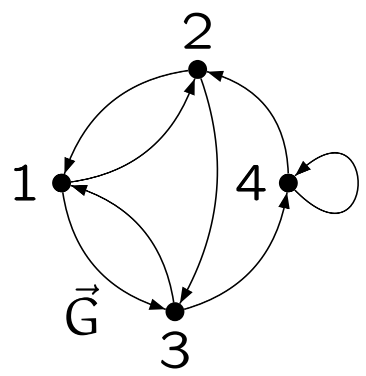


counting components: 3×1, 7×2, 5×3, 1×4.

transition system $\mathcal{T}(\vec{G})$ (graph state)
half-edge connections at vertices

Martin polynomial of 2-in 2-out digraph $\vec{G}$

$$m(\vec{G}; y) = \sum_{T \in \mathcal{T}(\vec{G})} (y-1)^{k(T) - c(\vec{G})}$$



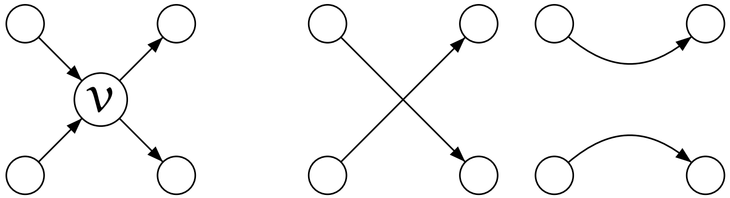
$c(\vec{G})$ components
$k(T)$ circuits for transition system T

$k\,:\; 3 \times 1,\ 7 \times 2,\ 5 \times 3,\ 1 \times 4.$

$3(y-1)^0 + 7(y-1)^1 + 5(y-1)^2 + 1(y-1)^3$

Thm. (1) recursive form (2) Tutte connection
(3) evaluations

Pierre Martin, Enumérations eulériennes dans les multigraphes et invariants de Tutte-Grothendieck, PhD thesis, 1977

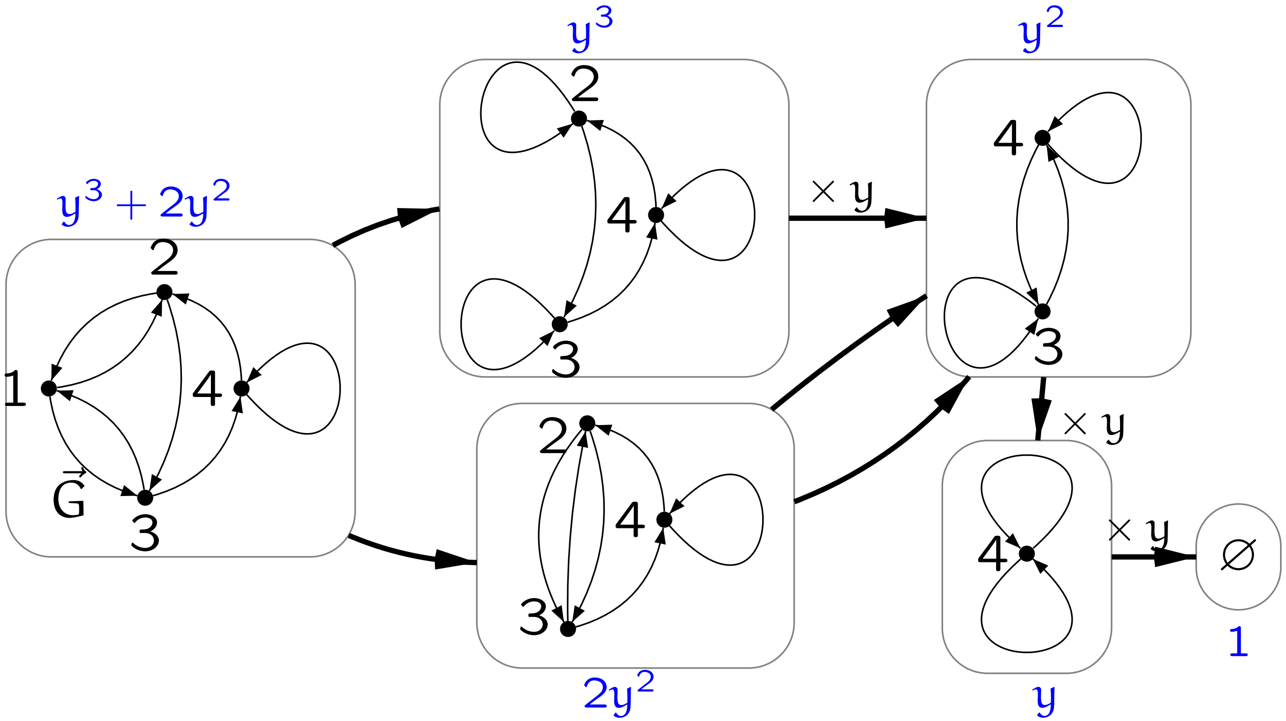


graph reductions: glueing edges

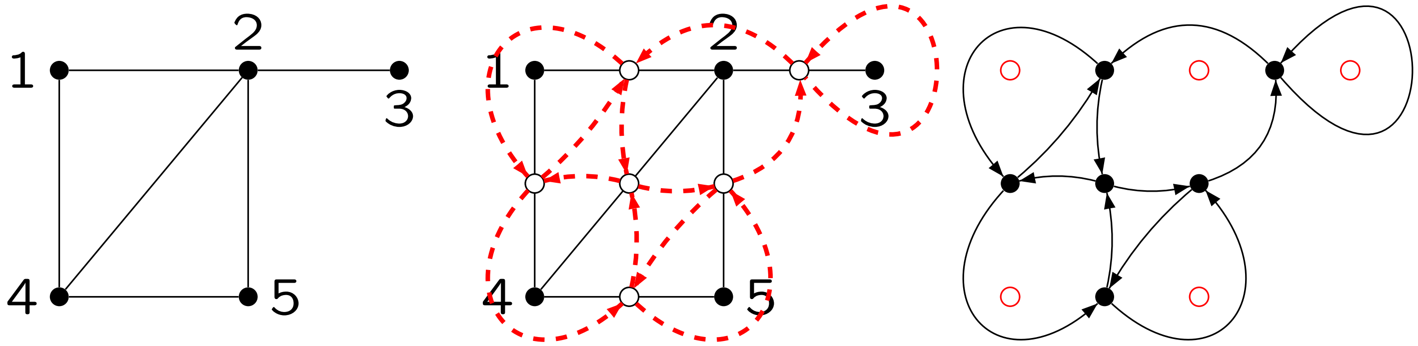
Def. $m(\vec{G}; y) = 1$ for $n = 0$

$m(\vec{G}; y) = y\, m(\vec{G}\,'; y)$ cut vertex

$m(G; y) = m(\vec{G}'_v; y) + m(\vec{G}''_v; y)$ without loops



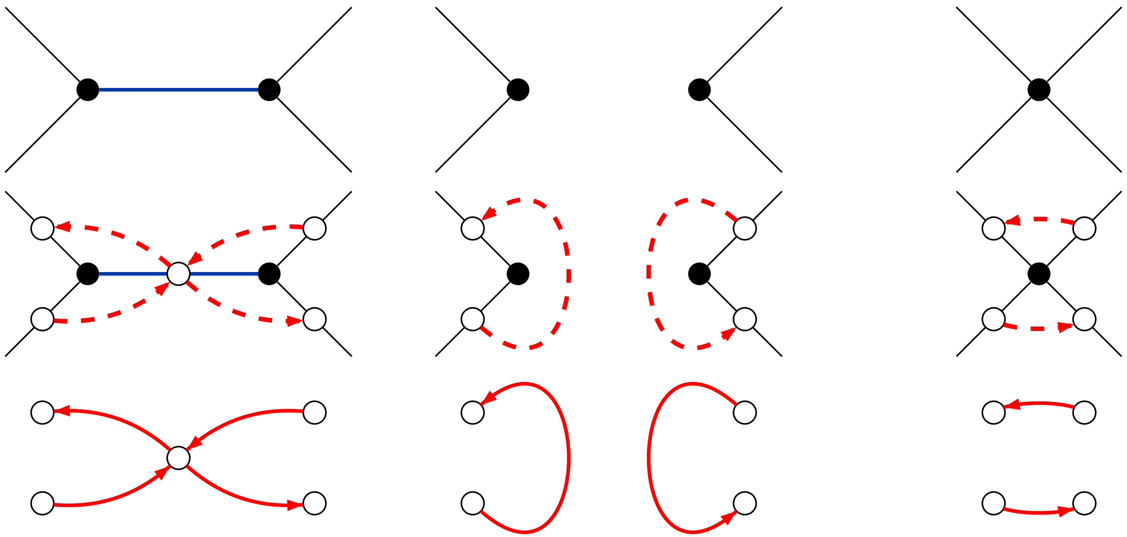
$3(y-1)^0 + 7(y-1)^1 +$
$5(y-1)^2 + 1(y-1)^3$



plane graph G, with medial graph $\vec{G}_m$

Thm. $m(\vec{G}_m; y) = T(G; y, y)$

proof: deletion-contraction $\equiv$ glueing edges



$m(\vec{G};y) = \sum_{T\in\mathcal{T}(\vec{G})} (y-1)^{k(T)-c(\vec{G})}$

$\vec{G}$ 2-in 2-out digraph and $n = |V(\vec{G})|$

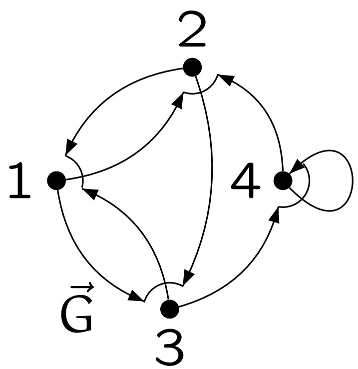
Thm. $m(\vec{G};-1) = (-1)^n(-2)^{a(\vec{G})-1}$ 'third connection'

$m(\vec{G};0) = 0$, when $n > 0$

$m(\vec{G};1)$ number of Eulerian systems

$m(\vec{G};2) = 2^n$

$m(\vec{G};3) = k|m(\vec{G};-1)|$ for odd k



$a(\vec{G})$ anti circuits

$m(\vec{G};y) = y^3 + 2y^2$

$m(\vec{G};-1) = 1$

Rearrangement Polynomials

Assembly polynomial
Interlace polynonial
connected to Martin polynomial

interlace graph
local and edge complement

assembly polynomial of G_w for doc-word w

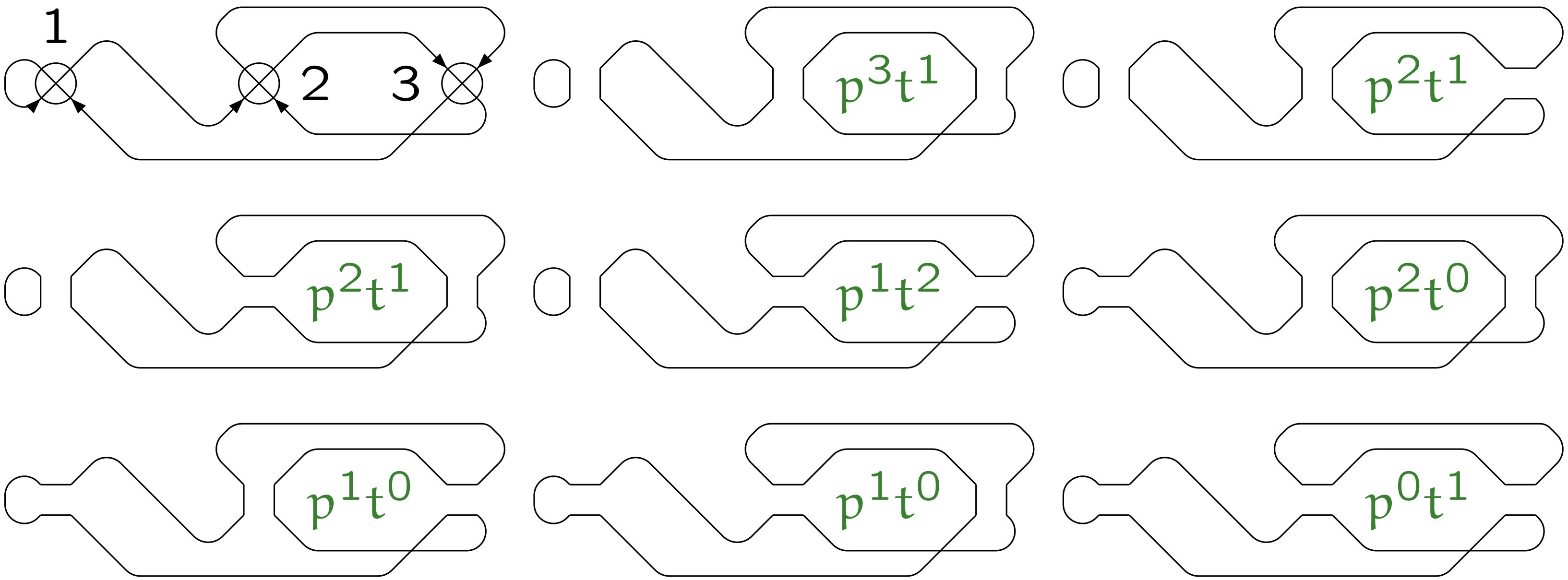
$$S(G_w)(p,t) = \sum_s p^{\pi(s)} t^{c(s)-1},$$

follow / consistent π / inconsistent

never p

$w = 1\,1\,2\,3\,2\,3$

$p^3t + 2p^2t + p^2 + pt^2 + 2p + t$



Burns, Dolzhenko, Jonoska, Muche, Saito:
Four-regular graphs with rigid vertices associated to DNA recombination (2013)

transition polynomials $W = (a, b, c)$

transition T defines partition V_1, V_2, V_3

eg wrt fixed cycle

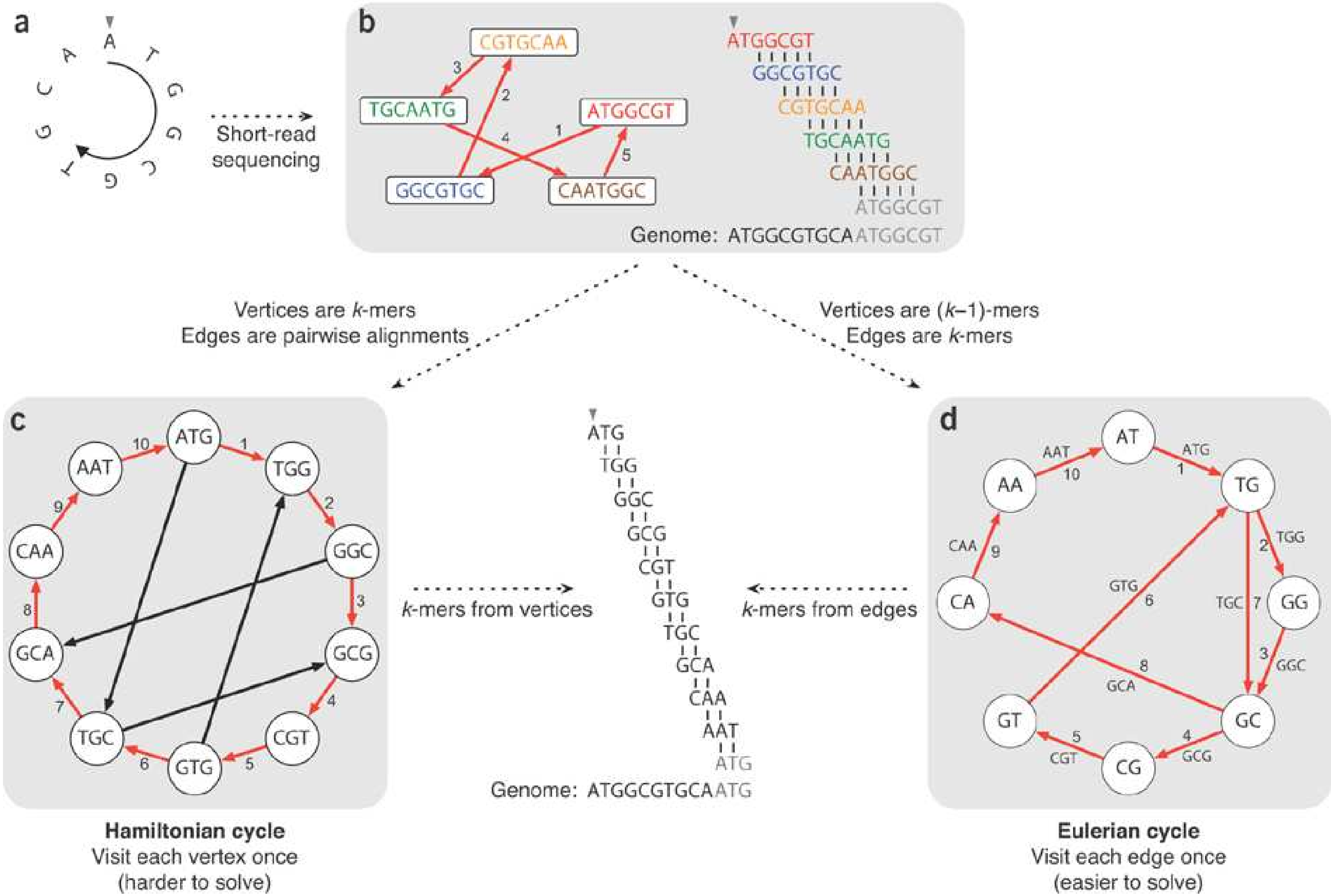
weight $W(T) = a^{|V_1|} b^{|V_2|} c^{|V_3|}$

$$M(G, W; y) = \sum_{T \in \mathcal{T}(\vec{G})} W(T) y^{k(T) - c(\vec{G})}$$

polynomial	a	b	c
Martin	1	1	0
(3-way)	1	1	1
assembly	0	p	1
Penrose	0	1	-1

$$= 3^3 - 3^2 - 3^2 - 3^2 + 3 + 3 + 3 - 3$$

F. Jaeger: On transition polynomials of 4-regular graphs (1990)



How to apply de Bruijn graphs to genome assembly, Compeau, Pevzner & Tesler

de Bruijn Graphs for DNA Sequencing
originally recursive definition

simple graph G (with loops)

interlace polynomial

(single-variable, vertex-nullity)

$$q(G;y) = \sum_{X \subseteq V(G)} (y-1)^{n(\,A(G)[X]\,)}$$

as Tutte, but: vertices vs. edges,
algebraic vs. combinatorial

Arratia, Bollobás, Sorkin: The interlace polynomial: a new graph polynomial (2000)

Aigner, van der Holst: Interlace polynomials (2004)

Bouchet: TutteMartin polynomials and orienting vectors of isotropic systems (1991)

Def. $q(G;y) = 1$ if $n = 0$

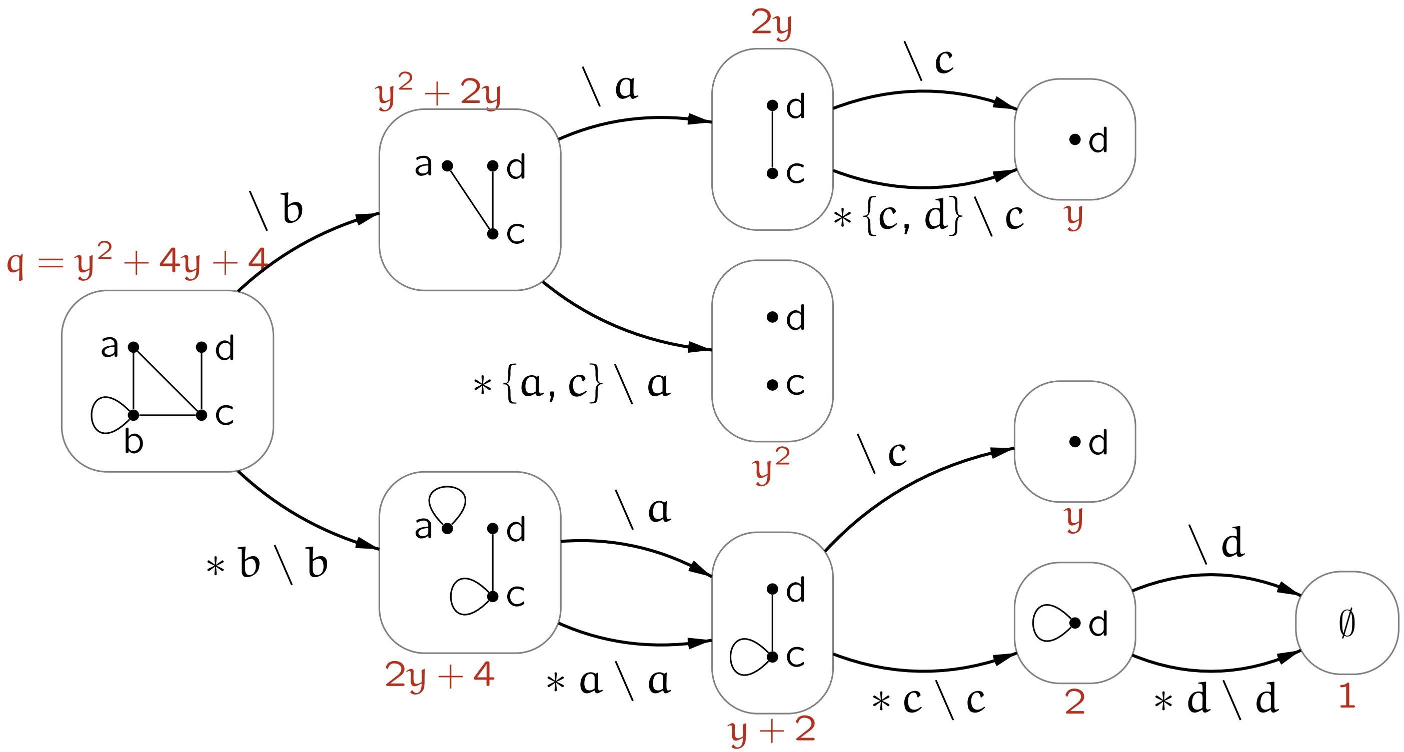
$q(G;y) = y\,q(G \setminus v; y)$ $\quad v$ isolated (unlooped)

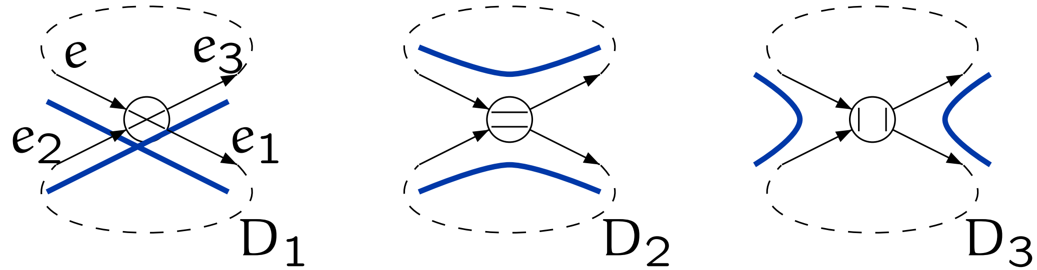
$q(G;y) = q(G \setminus v; y) + q((G * v) \setminus v; y)$

local & $\quad v$ looped

edge complement $\quad q(G;y) = q(G \setminus v; y) + q((G * e) \setminus v; y)$

$e = \{v, w\}$ unlooped edge





4-regular graph G with Eulerian system C

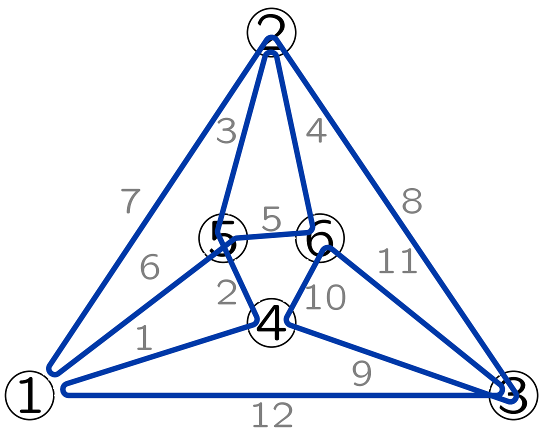
k circuit partition of $E(G)$, partition vertices:

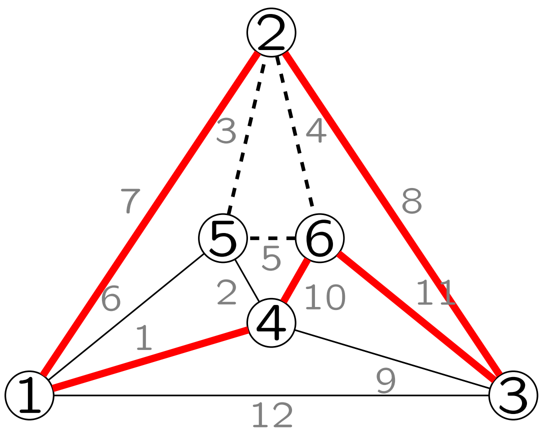
D_1 follows C

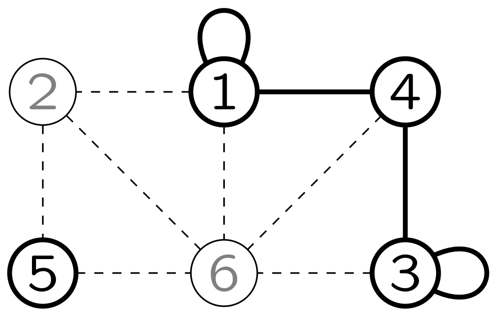
D_2 orientation consistent

D_3 orientation inconsistent

Thm. Then $k - c(G) = n(\,(I(C) + D_3) \setminus D_1\,)$







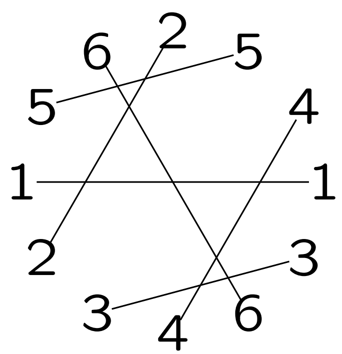
$$\begin{array}{c|cccc} & 1 & 3 & 4 & 5 \\ \hline 1 & 1 & 0 & 1 & 0 \\ 3 & 0 & 1 & 1 & 0 \\ 4 & 1 & 1 & 0 & 0 \\ 5 & 0 & 0 & 0 & 0 \end{array}$$

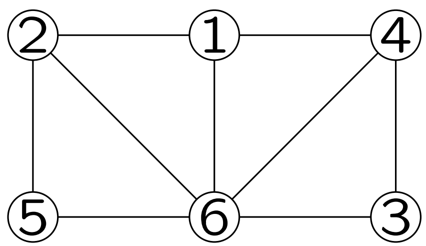
$$q(I(C);y) = \sum_{X \subseteq V(G)} (y-1)^{n(\,A(I(C))[X]\,)}$$

$$k - c(G) = n(\,(I(C) + \underbrace{D_3}_{=\varnothing}) \setminus D_1\,)$$

$$m(\vec{G};y) = \sum_{T \in \mathcal{T}(\vec{G})} (y-1)^{k(T)-c(\vec{G})}$$

Thm. $m(\vec{G};y) = q(I(C);y)$





$$w = 1\,4\,\underline{\mathbf{5}\,2\,6\,\mathbf{5}}\,1\,2\,3\,4\,6\,3$$

Thm. $q(G;y) = q(G*v;y)$ $\quad$ v looped

$q(G;y) = q(G*e;y)$ $\quad$ e unlooped edge

$q(G;y) = q(G*X;y)$ $\quad$ $G[X]$ nonsingular

Thm.

$m(\vec{G};-1) = (-1)^n(-2)^{a(\vec{G})-1}$	$q(G;-1) = (-1)^n(-2)^{n(A(G)+I)}$
$m(\vec{G};0) = 0$, when $n > 0$	$q(G;0) = 0$ if $n > 0$, no loops
$m(\vec{G};1)$ #Eulerian systems	$q(G;1)$ #induced subgraphs with odd number of perfect matchings
$m(\vec{G};2) = 2^n$	$q(G;2) = 2^n$
$m(\vec{G};3) = k\|m(\vec{G};-1)\|$ odd k	$q(G;3) = k\|q(G;-1)\|$ odd k

defined and studied these polynomials for Δ-matroids

(some things become less complicated that way)

nullity corresponds to minimal ‘size’

two directions $\Leftrightarrow$ deletion and contraction

need another ‘minor’ for third direction

thank you ciliates!

when studying new polynomials

look back at old ones

connections to recursive formulations, and

special evaluations

THANKS