Graph polynomials motivated by Gene Rearrangements in Ciliates

Hendrik Jan Hoogeboom
Leiden NL

joint with

Robert Brijder
Hasselt B
natural computation

bio inspired computing

comput molecular biol

DNA computing
DNA en grafen
pizza canibale
inhoudsopgave

- unscrambling genes in ciliates
- de Bruijn graphs & gene assembly
- sorting by reversal
inhoudsopgave

- unscrambling genes in ciliates
- de Bruijn graphs & gene assembly
- sorting by reversal
genome: turning cabbage into turnip

‘van kool naar raap’
genome: turning cabbage into turnip
<table>
<thead>
<tr>
<th>12</th>
<th>31</th>
<th>34</th>
<th>28</th>
<th>26</th>
<th>17</th>
<th>29</th>
<th>4</th>
<th>9</th>
<th>36</th>
<th>18</th>
<th>35</th>
<th>19</th>
<th>16</th>
<th>14</th>
<th>32</th>
<th>33</th>
</tr>
</thead>
<tbody>
<tr>
<td>22</td>
<td>15</td>
<td>11</td>
<td>27</td>
<td>5</td>
<td>20</td>
<td>13</td>
<td>30</td>
<td>23</td>
<td>10</td>
<td>6</td>
<td>3</td>
<td>24</td>
<td>21</td>
<td>8</td>
<td>25</td>
<td>27</td>
</tr>
</tbody>
</table>

 mitochondrial DNA: worm Ascaris Suum into human
model: reality and desire
Bad Components

Non-Hurdles
- E
- B

Hurdles

Simple Hurdles
- A
- D
- C

Super Hurdles
- F

hurdle & fortress

\[ d(\pi) \geq b(\pi) - c(\pi) + h(\pi) + f(\pi) \]
inhoudsopgave

- unscrambling genes in ciliates
- de Bruijn graphs & gene assembly
- sorting by reversal
Nicolaas Govert "Dick" de Bruijn

(9 July 1918 – 17 February 2012)

- De Bruijn sequence (1946)
- BEST Theorem (1951)
- Automath (1967)

http://www.win.tue.nl/automath/images/photos/deBruijn.jpg
de Bruijn, van Aardenne-Ehrenfest, Smith & Tutte

http://spikedmath.com/327.html

\[ ec(G) = t_w(G) \prod_{v \in V} (\deg(v) - 1)! \]

BEST. THEOREM. EVER!

Directed trees

Euler cycles

speciaal geval \( \deg(v) = 2 \)

http://myweb.lsbu.ac.uk/~whittyr/MathSci/TheoremOfTheDay/CombinatorialTheory/BEST/TotDBEST.pdf
genetic / physical map
chain-termination sequencing
(Sanger)
300 - 1000 mer

http://en.wikipedia.org/wiki/DNA_sequencing
How to apply de Bruijn graphs to genome assembly
(Phillip E C Compeau, Pavel A Pevzner & Glenn Tesler)
inhoudsopgave

- unscrambling genes in ciliates
- de Bruijn graphs & gene assembly
- sorting by reversal
Two different sorts of nuclei:
• **micronucleus** (small, diploid, reproduction)
• **macronucleus** (large, polyploid, cell regulation)

gene assembly: **micronucleus** → **macronucleus**

http://en.wikipedia.org/wiki/Ciliate
“This indicates that, in principle, these unicellular organisms may have the capacity to perform at least any computation carried out by an electronic computer”
[L. Landweber, L. Kari: Universal Molecular Computation in Ciliates, 2002]

“The way that ciliates transform genes from their micronuclear (storage) form into their macronuclear (expression) form constitutes a very interesting case of ‘DNA computing in vivo’ ”
[Ehrenfeucht, Prescott, Rozenberg: Computational Aspects of Gene (Un)Scrambling in Ciliates, 2002]
1. Loop recombination

2. Hairpin recombination

3. Double-loop recombination

pointers between two copies of 4
look at the “gymnastics”

GENOME GYMNASTICS: UNIQUE MODES OF DNA EVOLUTION AND PROCESSING IN CILIATES

David M. Prescott

In some ciliates, the DNA sequences of the germline genomes have been profoundly modified during evolution, providing unprecedented examples of germline DNA malleability. Although the significance of the modifications and malleability is unclear, they may reflect the evolution of mechanisms that facilitate evolution. Because of the modifications, these ciliates must perform remarkable feats of cutting, splicing, rearrangement and elimination of DNA sequences to convert the chromosomal DNA in the germline genome (micronuclear genome) into gene-sized DNA molecules in the somatic genome (macronuclear genome). How these manipulations of DNA are guided and carried out is largely unknown. However, the organization and manipulation of ciliate DNA sequences are new phenomena that expand a general appreciation for the flexibility of DNA in evolution and development.
Prescott: Genome Gymnastics, Nature Reviews, Genetics
4-regular graph

Prescott: Genome Gymnastics, Nature Reviews, Genetics
Prescott: Genome Gymnastics, Nature Reviews, Genetics
How to study intramolecular gene assembly?
Eulerian circuits in 4-regular graphs

string = Euler circuit
126134563245

4-regular graph + Euler circuit
Eulerian circuits in 4-regular graphs

String = Euler circuit
126134563245

1-2-6 & 3-2-4

4-regular graph + Euler circuit
Eulerian circuits in 4-regular graphs

string = Euler circuit
126134563245

1–2–6 & 3–2–4

we can split split or invert

1245, 61345632

4–regular graph + Euler circuit

1–2–3 & 6–2–4

123654316245

invert
reordering at vertices

Kotzig 1966

interlaced splits: swap

p ... q ... p ... q ...
“gene assembly polynomial”

\[ t(t - 1)(t - 2)(t^7 - 12t^6 + 67t^5 - 230t^4 + 529t^3 - 814t^2 + 775t - 352) \]
The Tutte polynomial is defined as:

\[ T_G(x, y) = \sum_{A \subseteq E} (x - 1)^{c(G[A])} c(G)(y - 1)^{c(G[A]) + |A| - |V|} \]

number components of induced subgraph $G[A]$
Graph polynomials

Tutte rather generic

- combinatorial
- algebraic
- recursive (deletion contraction)

definitions / interpretations

Interesting special values

$T_G(1,1)$ # spanning forests

Related
- knots “Jones polynomial”
- Statistical physics

Extended to matroids
⇒ motivation: counting intermediate products
assembly polynomial

w = 1 1 2 3 2 3

follows C
consistent with C}

inconsistent with C

split
invert

2 consistent; 2 components

\( p^3 t + 2p^2 t + p^2 + pt^2 + 2p + t \)
F. Jaeger: On transition polynomials of 4-regular graphs (1990)

Roger Penrose, Applications of Negative Dimensional Tensors
Theorem (Traldi–Cohn–Lempel)
- C Euler cycle in 4–regular graph
- P circuit partition, transitions wrt C
  \[ D_1 \text{ follows } C \]
  \[ D_2 \text{ consistent with } C \]
  \[ D_3 \text{ inconsistent with } C \]
then

\[ |P| = \nu( (I(C) + \Delta(D_3)) \setminus D_1 ) + 1 \]

nullity

overlap/interlace graph deletion

loops / diagonal matrix

M.Cohn, A.Lempel: Cycle decomposition by disjoint transpositions (1972)
L.Traldi: Binary nullity, Euler circuits and interlace polynomials (2011)
4-regular graph from doc-string

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

$D_1 = \{ 2, 6 \}$
$D_2 = \{ 4, 5 \}$
$D_3 = \{ 1, 3 \}$

$I(C_w)$

$\left| P \right| = \nu( (I(C) + \Delta(D_3)) \setminus D_1 ) + 1$

$I(C_w)$

rank=2 nullity = 2
what have we learned (for free)

1. as transition polynomial
   - evaluations at special points
   - recursive formulations
     "contraction–deletion"

2. algebraic formulation
   using $I(C_w)$ interlace graph
   - related to bracket polynomial
   - if $I(C_{w1}) = I(C_{w2})$ then equal pols
dankuwel ...

klaar