PROPERTIES OF THE INTERLACE POLYNOMIAL

Jo Ellis-Monaghan*
St. Michaels College,
Colchester, VT 05439
Irasema Sarmiento
CINVESTAV, Mexico

e-mail: jellis-monaghan@smcvt.edu,
website: http://academics.smcvt.edu/jellis-monaghan

The motivating problem: DNA sequencing

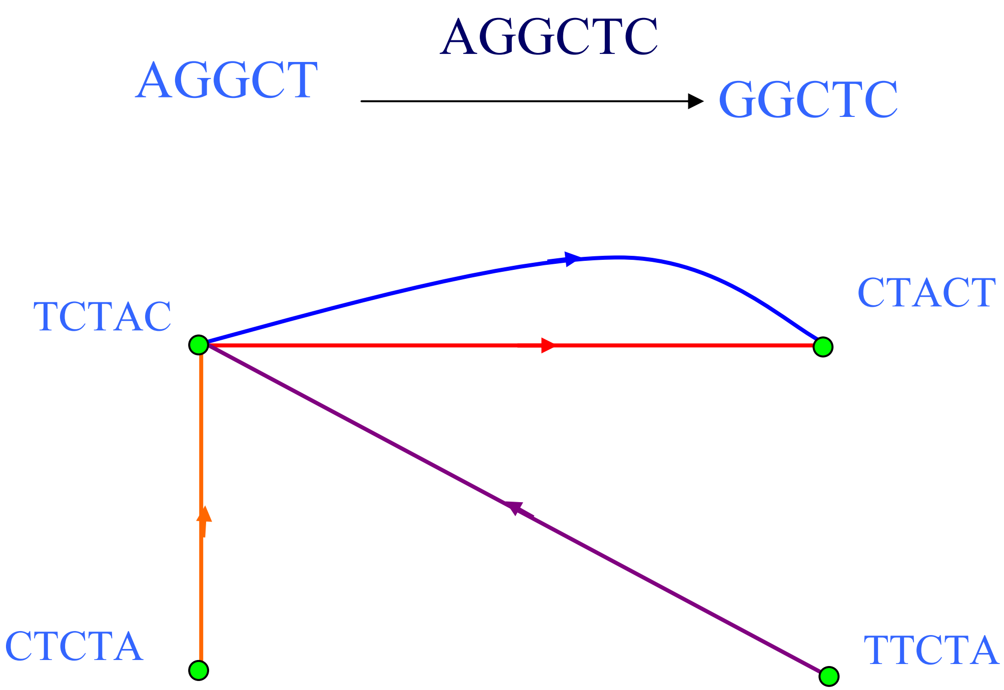
It is very hard in general to "read off" the sequence of a long strand of DNA. Instead, researchers probe for "snippets" of a fixed length, and read those.

The problem then becomes reconstructing the original long strand of DNA from the set of snippets.

AGGCTC, TCTACT, CTCTAC, TTCTAC,
TCTACT, GCTCTC TCTCTA, CTCTCT,
GTTCTA,...

Creating the DeBruijn Graph

AGGCTC, TCTACT, CTCTAC, TTCTAC, TCTACT, GCTCTC TCTCTA, CTCTCT, GTTCTA,...



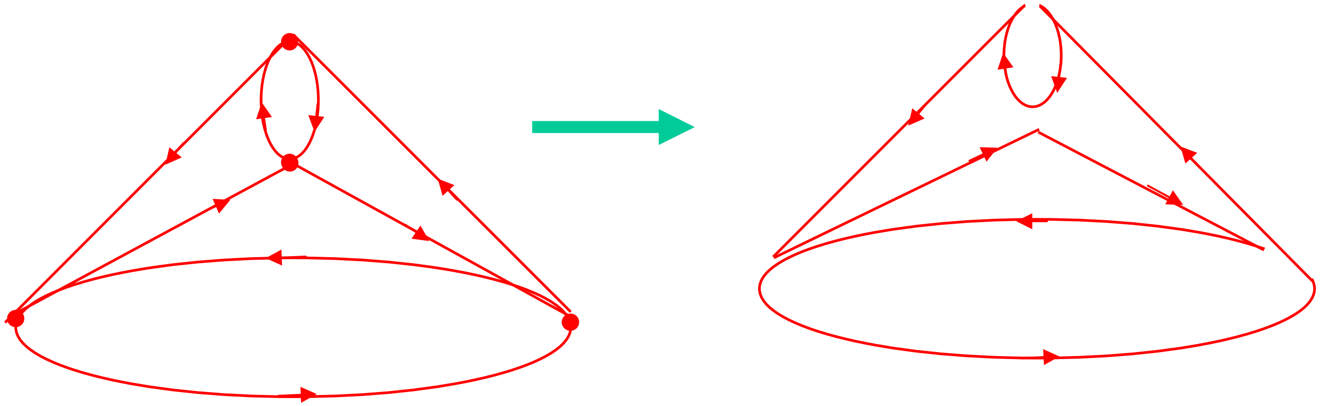
Create a directed edge for each snippet by placing a 'tail' vertex for first *n-1* letters, and a 'head' vertex for last *n-1* letters.

Don't repeat vertices.

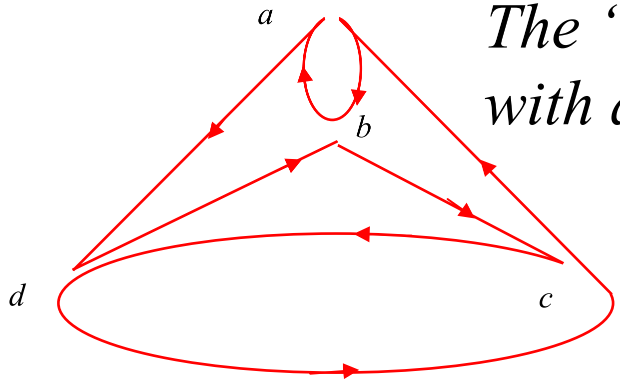
Enumerating the reconstructions

This leads to a directed graph with the same number of in-arrows as out-arrows at each vertex.

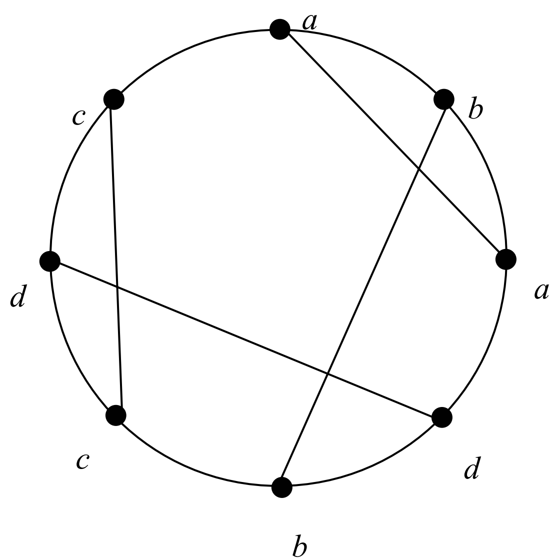
The number of reconstructions is then equal to the number of paths through the graph that traverse all the edges in the direction of their arrows.



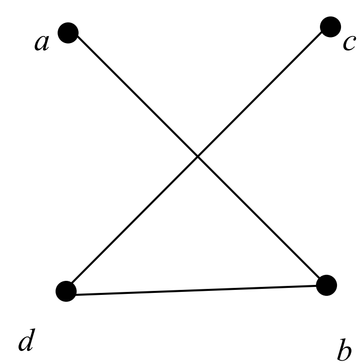
We will often focus on, not on the “snippet” graph, but on an associated circle graph.



The “snippet” graph, with an Euler circuit



A chord diagram

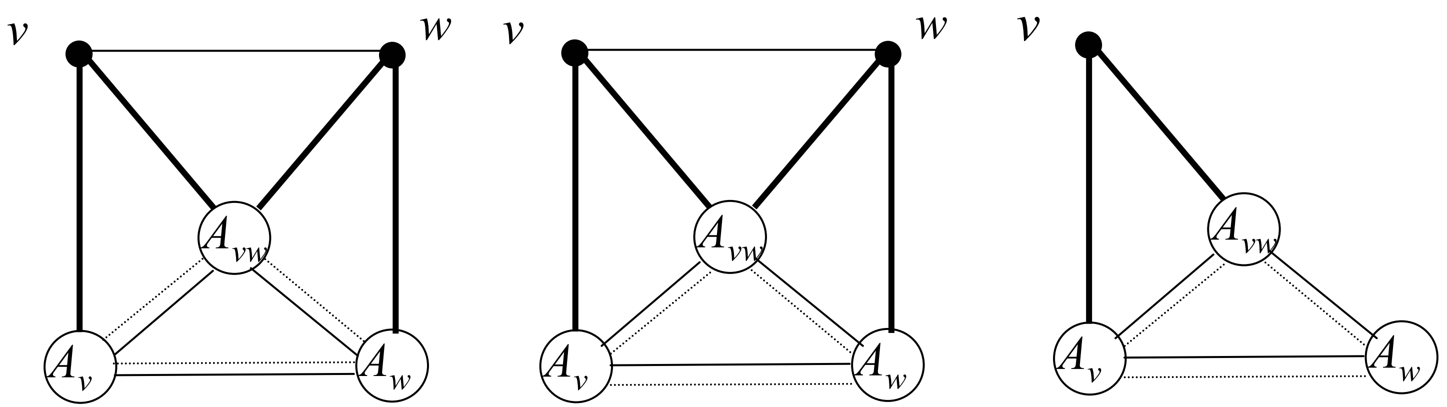


The associated circle graph

The Interlace Polynomial

Arratia, Bollobás, Sorkin, ’00.

$$q(G,x)=\begin{cases} x^{n} & \text{if } G=E_n \text{, the edgeless graph on } n \text{ vertices} \\ q(G-v,x)+q\left(G^{vw}-w,x\right) & \text{if } vw\in E(G) \end{cases}$$



G

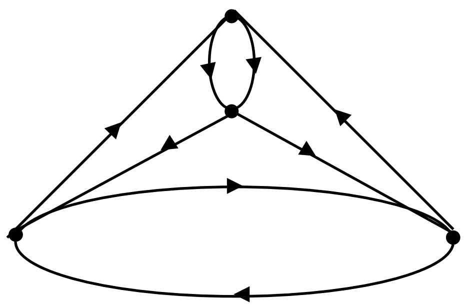
G^{vw}
(note interchange of edges and non-edges among A_v, A_w and A_{vw})

G^{vw}-*w*

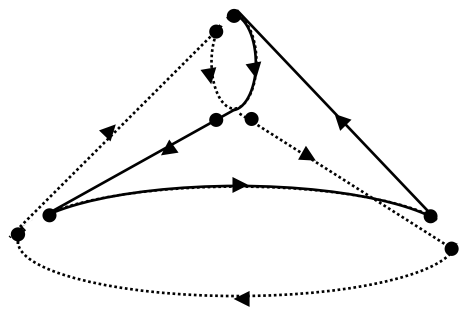
Yes, *q* is well-defined (yucky).

What does $q(G)$ count?

If H is a circle graph, then $q(H)$ counts families of Eulerian circuits in an associated Eulerian graph:



A 4-regular Eulerian digraph



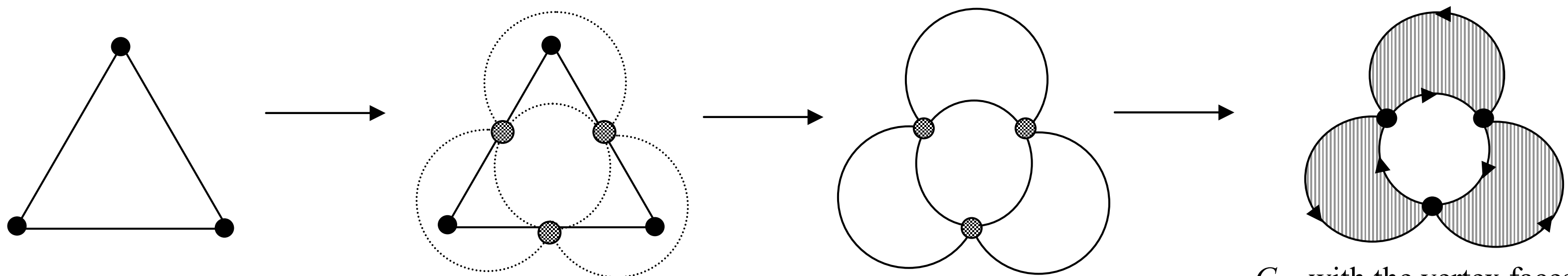
A graph state with 2 components

$$\sum f_k\left(\vec{G}\right)x^k = xq(H; x+1)$$

Where $f_k\left(\vec{G}\right)$ is the number of graph states with k components, and H is the circle graph associated to any Eulerian circuit of $\vec{G}$. [ABS ’04]

From Interlace to Tutte

Theorem: If G is a planar graph, and H is the circle graph of some Eulerian circuit of $\vec{G}_m$, then $q(H; x) = t(G; x, x)$.



A planar graph G

The medial graph G_m

G_m with the vertex faces colored black, oriented so that black faces are to the left of each edge.

Proof: Exploits the relation between $\sum f_k\left(\vec{G}\right)x^k$ (essentially the Martin polynomial) and the Tutte polynomial—See Martin and Las Vergnas.

Cor: The interlace polynomial is computationally intractable.

Question: Tractable classes? YES—stay tuned.

What does $q(G)$ count in general, and is there a closed form?

Theorem:

$$q(G) = \sum_{W \subseteq V(G)} (x-1)^{|W| - r(M(W))}$$

where $M(W)$ is the adjacency matrix of G restricted to W, and r is the GF_2 rank function.

Proof: Bouchet's theory of isotropic systems and the fact that $q(G)$ can be viewed as a special case of a Tutte-Martin polynomial of an isotropic system.

(Independently by Aigner& van der Holst, '04, using different techniques)

The Gamma Invariant

Let γ (G) be the coefficient of x in $q(G)$, analogously to the β-invariant, which is the coefficient of x in the Tutte polynomial.

Theorem: There exists a class of graphs, called *Pendant-Duplicate graphs,* which are characterized by $\gamma = 2$, and for which q is polynomial time to compute.

P-D Graphs:

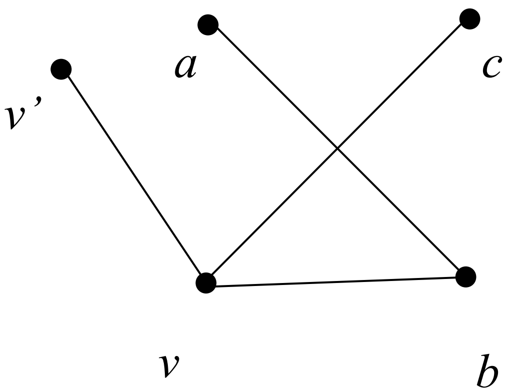
Start with a single vertex and either:

Add a pendant vertex to a vertex v, i.e. add a new vertex v' and an edge vv', or

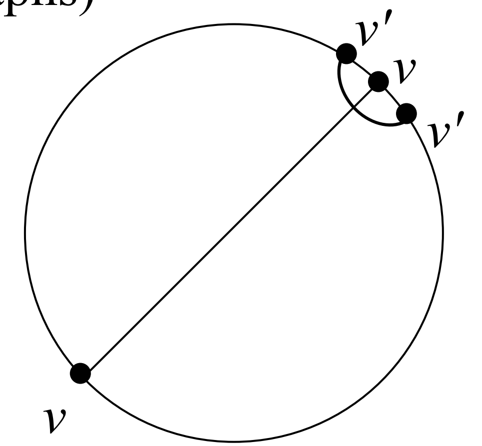
Add a duplicate vertex of a vertex v, i.e. add a new vertex v' and edge uv', if and only if uv is an edge.

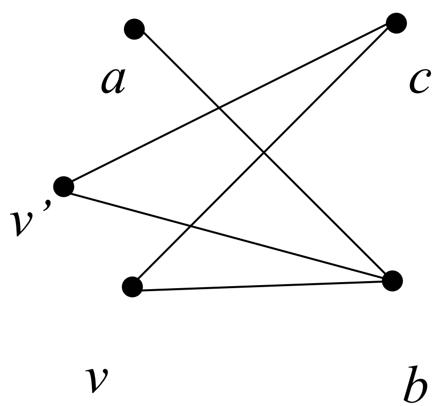
Pendant-Duplicate Graphs

Effect of adding a pendant vertex or duplicating a vertex
(P-D graphs are circle graphs)

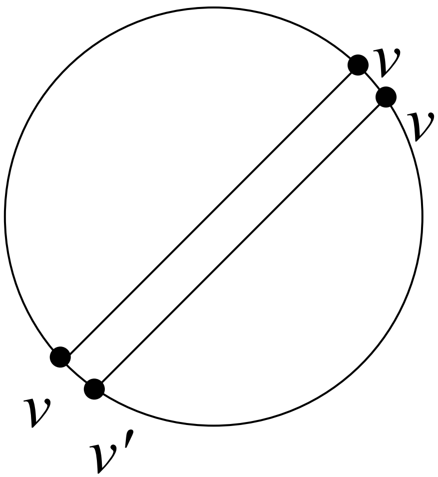


Adding a pendant vertex to v.





Duplicating the vertex v.



Relation between P-D graphs and series-parallel graphs

Theorem: H is a connected P-D graph with at least two vertices if and only if it is the circle graph of an Euler circuit in $\vec{G}_m$, where G is a series parallel graph.

Proof: By induction.

Proof that P-D graphs are characterized by gamma invariant 2

The proof is analogous to Brylawski's proof that series parallel graphs are characterized by beta invariant of 1, but instead of deletion/contraction, the 'dual' operations for q are:

Pendant: If G' is the graph that results from adding a pendant vertex w to a vertex v of G, then $q(G'; x) = q(G; x) + xq(G \setminus v; x)$.

Duplicate: If G' is the graph that results from duplicating a non isolated vertex v, then $q(G'; x) = q(G; x) + xq(G^{uv} \setminus u; x)$ where uv is an edge of G.

Proof then follows from using this to show that gamma is invariant under adding duplicate or pendant vertices.

Consequence for original application

A set of subsequences of DNA has the alternating sum of possible reconstructions = 2

iff

the circle graph H associated to any Eulerian circuit of the 'snippet' graph is a pendant-duplicate graph.

Proof: Recall that $\sum f_k\left(\vec{G}\right)x^k = xq(H;x+1)$

Thus $\sum f_k\left(\vec{G}\right)(x-1)^k = (x-1)q(H;x)$

Reconstructions with k components

Equate coefficients of x to get: $\sum f_k\left(\vec{G}\right)(-1)^k = 2$