# Graph Models for DNA Structures

#### Jo Ellis-Monaghan

St. Michaels College, Colchester, VT 05439 e-mail: jellis-monaghan@smcvt.edu website: http://academics.smcvt.edu/jellis-monaghan



Work supported by the Vermont Genetics Network through NIH Grant Number 1 P20 RR16462 from the INBRE program of the National Center for Research Resources



On the care and maintence of your Mathematician...

YES !!



**N000**...

### Also an info-mercial for...

Mary Cox— Knots and the Stuck Unknot in Hydrocarbon Chains





Jess Scheld— DNA Sequencing and Eulerian Circuits







Marisa Debowsky— Biomolecular Computing and Topological Graph Theory Problems motivated by applications in biology

New Mathematical theory and tools

Existing Mathematical theory and tools

## COMMUNICATION IS KEY...

- **1. Explain the biological problem to the mathematician (problem formulation).**
- 2. Develop the necessary and sufficient formalism to model the problem.
- 3. Apply/develop mathematical theory and tools.
- 4. Communicate the mathematics to the biologist *in a* way that actually informs the problem.

#### **GRAPHS AND NETWORKS**

A **Graph or Network** is a set of *vertices* (dots) with *edges* (lines) connecting them.



### **BIOMOLECULAR COMPUTING**



L. M. Adleman, Molecular Computation of Solutions to Combinatorial Problems. *Science*, **266** (5187) Nov. 11 (1994) 1021-1024.

- 1. Encode a question in a biological structure
- 2. Apply a biological process to the structure
- 3. Be able to isolate a solution to the question from the result of the applied process



### A CHARACTERIZATION FOR DNA NANO-STRUCTURES

http://seemanlab4.chem.NYU.edu

A theorem of C. Thomassen specifies precisely when a graph may be constructed from a single strand of DNA, and theorems of Hongbing and Zhu to characterize graphs that require at least *m* strands of DNA in their construction. (1990 and 1998).

Theorem: A graph G may be constructed from a single strand of DNA if and only if G is connected, has no vertex of degree 1, and has a spanning tree T such that every connected component of G - E(T) has an even number of edges or a vertex v with degree greater than 3.

## GARBAGE TRUCKS???



### Oriented Walk Double Covering and Bidirectional Double Tracing





"The authors of this paper came across the problem of bidirectional double tracing by considering the so called "garbage collecting" problem, where a garbage collecting truck needs to traverse each side of every street exactly once, making as few U-turns (retractions) as possible."

(See M. Debowsky poster for more info and progress!)

# (subtitle: why there aren't many pure math posters....)

**Theorem 3.3.** (The recipe theorem.)

Let  $\Psi$  be a map from a minor closed subset  $\mathcal{M}$  of ribbon graphs with  $D_{ijk} \in \mathcal{M}$  to a commutative ring  $\mathcal{R}$  with unity such that

(i)

$$\Psi(G) = \begin{cases} \Psi(G/e) + \Psi(G \setminus e) & \text{if } e \text{ is ordinary} \\ x\Psi(G/e) & \text{if } e \text{ is } a \text{ bridge} \end{cases}$$

- (ii)  $\Psi(D_{ijk}) = \alpha s^j q^{i-2j-k} r^k$ , where  $s = \Psi(D_{210})$ ,  $q = \Psi(D_{100})$  and  $r = \Psi(D_{101})$ .
- (iii)(iiia) Ψ(G∪H) = Ψ(G)Ψ(H) if G, H are one vertex graphs,
  - (iiib) Ψ(E) = α if E is the edgeless graph on one vertex or the empty graph.

(iv) There are elements u and v in  $\mathcal{R}$  such that

$$\begin{array}{l} (iva) \ (q-1)^2 u^2 = s-2q+1, \\ (ivb) \ (q-1)uv = r-1, \\ (ivc) \ v = v^2. \end{array}$$

Then

$$Ψ(G) = α^{k(G)}R(G; x, q - 1, u, v),$$

on  $\mathcal{R}$ , where k(G) is the number of componentes of G.

*Proof.* We first suppose G is a (signed) chord diagram. We note that if G consists of a single chord or is empty the result follows immediately since  $R(D_{100}; X, Y, Z, W) = 1 + Y$  and  $R(D_{101}; X, Z, W) = 1 + YZW$ . Assume by induction that  $\Psi(G) = \alpha R(G; x, q - 1, u, v)$  if G is a signed chord diagram with fewer than m chords. Then the function  $\Psi'(G) = \Psi(G) - \alpha R(G; x, q - 1, u, v)$  vanishes on signed chord diagrams with fewer than m chords and satisfies (i) for all ribbon graphs. Thus, as in the proof of Theorem 2 of [BR02], if  $D_1$  and  $D_2$  are related diagrams with m chords,

$$\Psi'(D_1) - a\Psi'(D'_1) = \Psi'(D_2) - a\Psi'(D'_2),$$

A topological Tutte polynomial was developed in 2000-2, independent of any particular application.

Turns out that it encodes information about graphs constructed from DNA

Understanding this polynomial will help develop tools to analyze the structural properties of DNA nano-constructs.

The theory is of intrinsic mathematical interest....

### DNA SEQUENCING

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



A very fancy polynomial, the interlace polynomial, of Arratia, Bollobás, and Sorkin ,2000, encodes the number of ways to reassemble the original strand of DNA, as does a simpler one, the circuit partition polynomial (work begun in late 70's by a variety of people, but not recognized for the biology application until 2000 or so).

(See J. Scheld poster for more info and progress!)

#### NEW RESULTS

Lemma 4.4. Let F be an element of  $\{x, y\}^V$ . Then

 $dim(L \cap \widehat{F}) = dim(U) = |F_x| - r(M),$ 

where M is the adjacency matrix of  $G \mid F_x$ .

*Proof.* Let  $\psi : \mathcal{P}(F_x) \to GF(2)^{F_x}$  be the homomorphism whose matrix M is the adjacency matrix of  $G \mid F_x$ . That is, the rows and columns of M are indexed by the elements of  $G \mid F_x$  and  $M_{ij} = 1$  iff i and j are joined by an edge in  $G \mid F_x$ . Also  $\psi(P) = M \cdot \chi_P$ , where  $\chi_P$  is the characteristic polynomial of P written as a column.

We have that  $P \in Ker(\psi)$  if and only if  $v(P) = M \cdot \chi_P = 0$  (the all zero column). That is, for every  $i \in F_x$ ,  $0 = \sum_{j \in F_x} M_{ij} \chi_{P_j} = \sum_{j \in P} M_{ij}$ . In other words, for every  $i \in F_x$ , the cardinality of  $star(v) \cap P$  is even.

Thus  $L \cap \widehat{F}$  is the kernel of  $\psi$  and  $dim(Im(\psi)) = r(M)$ , where r is the GF(2) rank function. On the other hand  $dim(\mathcal{P}(F_x)) = |F_x|$ .

Therefore

$$\dim(L \cap \widehat{F}) = \dim(U) = |F_x| - r(M).$$

Theorem 4.5. Let G be a simple graph. Let q(G; x) be the interlace polynomial of G. Then

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

where M(W) is the adjacency matrix of  $G \mid W$ .

*Proof.* By Theorem refBou2.8, given a simple graph G on the vertex set V, supplementary vectors A and B of  $K^V$ , and  $L = \{A(N(P)) + B(P) : P \in \mathcal{P}(V)\}$ , then the pair S = (L, V) is an isotropic system. Moreover, the mapping  $X \to A(N(P)) + B(P)$  is a linear bijection from  $\mathcal{P}(V)$  onto L. Let  $A = \bar{x}, B = \bar{y}$  and  $C = A + B = \bar{z}$  as before. It follows that there is an isotropic system S such that (G, A, B) is its graphic presentation.

Bouchet proved ([Bou]) also that

 $q(G;\xi) = m(S,C;\xi),$ 

The alternating sum of possible reconstructions of a DNA strand is exactly 2 if and only the circle graph associated to any Eulerian circuit of the 'snippet' graph is a pendant-duplicate graph (identifiable in polynomial time).

### To grow a happy, healthy, and vigorous collegue: *Talk to your mathematician!*\*

There is a lot of math out there.

Applied mathematicians are interested in making the connection between the math and the application and then moving the theory in directions motivated by the application.

If the math is "out there" in a purely metaphysical rather than practical sense, the demands of the application may open up an entirely new branch of mathematics.

#### THIS IS ALL VERY EXCITING...

(See M. Cox poster for some cool new directions!)

\* Singing songs and playing gentle music is not generally required, although donuts may stimulate growth.