Graph Theory and DNA Nanostructures

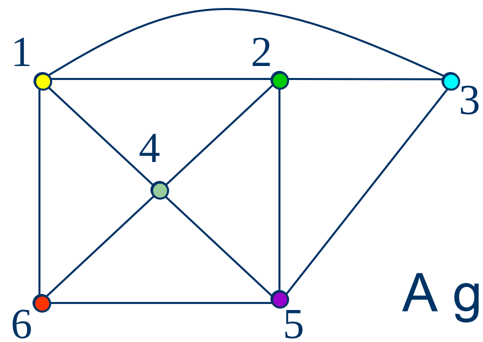
Laura Beaudin, Jo Ellis-Monaghan*, Natasha Jonoska, David Miller, and Greta Pangborn

VGN NATIONAL INSTITUTES OF HEALTH INBRE
NIH

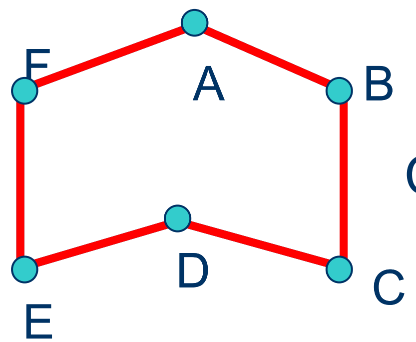
NATIONAL SECURITY AGENCY

SAINT MICHAEL'S COLLEGE

A graph is a set of vertices (dots) with edges (lines) connecting them.

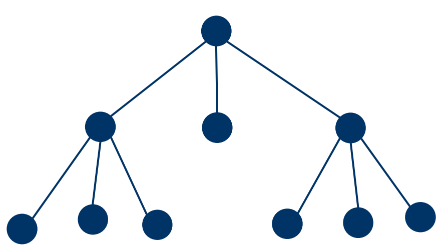


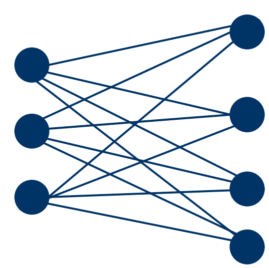
A graph



C_6 (cycle)

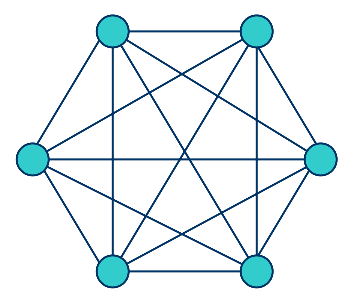
(tree)





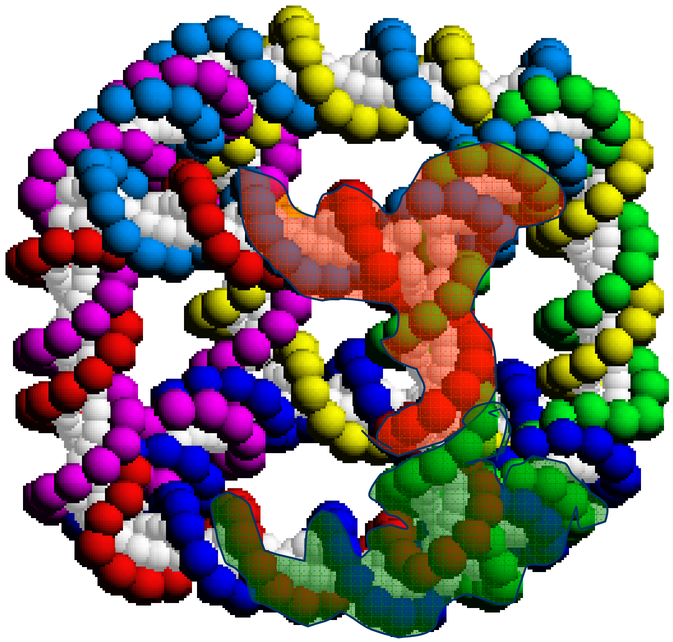
$K_{3,4}$ (bipartite)

K_6
(complete)

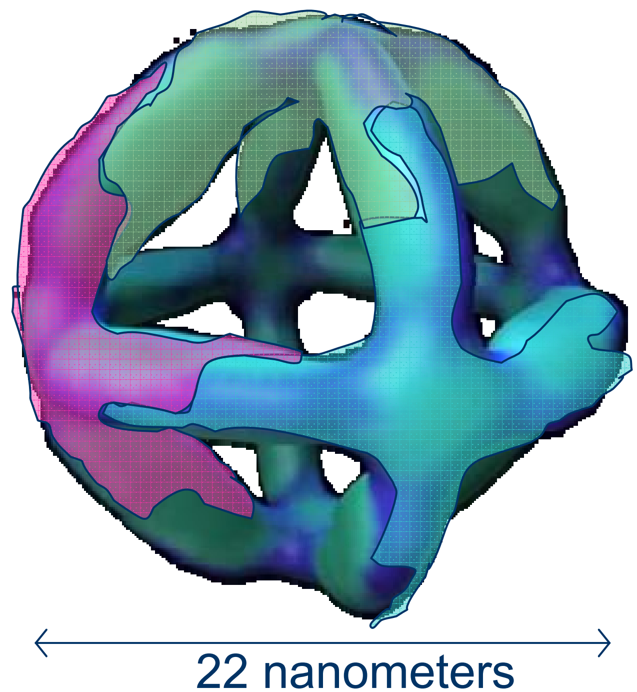


What are self-assembled DNA nanostructues?

- A self-assembled DNA cube and Octahedron

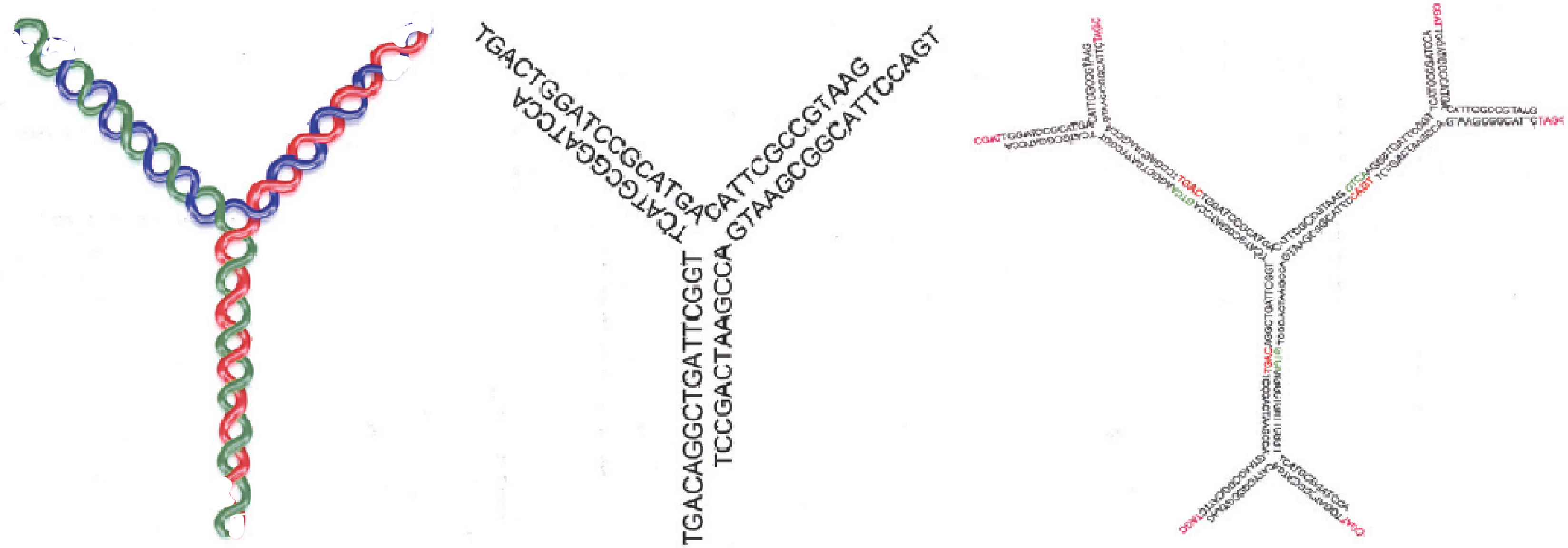


http://seemanlab4.chem.nyu.edu/nanotech.html



The molecular building blocks

- *K*-armed branched junction molecules

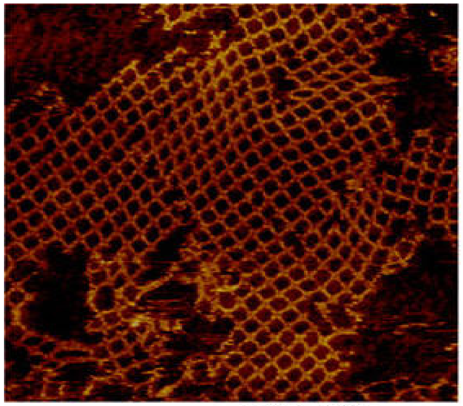


Y-shaped DNA. Schematic diagrams of the structure (left) and sequence (middle) of Y-DNA, and dendrimer-like DNA (right).

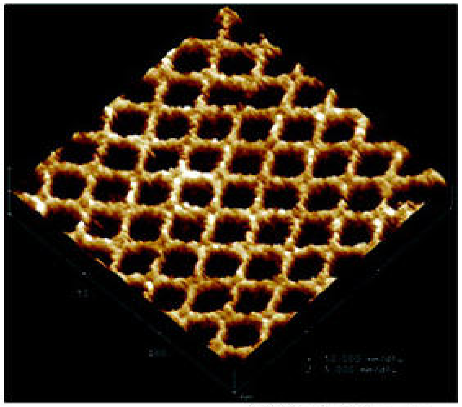
D. Luo, “The road from biology to materials,” Materials Today, **6** (2003), 38-43

Why self-assembling nanostructures?

- Biomolecular computing (Hamilton Cycle/3-Sat)
- Nanoelectronics
- Fine screen filters (lattices) at the nano-size scale
- Biosensors and drug delivery mechanisms



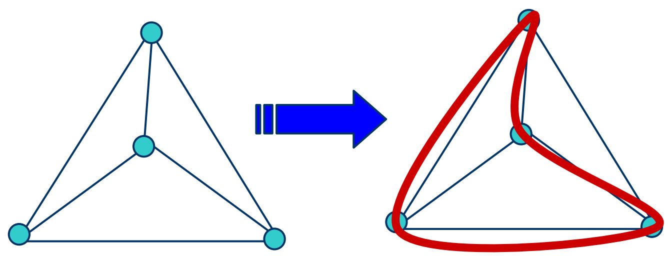
500x500 nm



150x150 nm

http://www.nanopicoftheday.org/2004Pics/April2004/DNAmesh.htm

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

The application↔theory cycle

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.*

The fundamental questions

Given a target graph,

1. what is the minimum number of k-armed branched junction molecules that must be designed to create the graph?
2. What is the minimum number of bond types needed?
3. What is the combinatorial structure of the molecules in a minimal set?

Three different laboratory constraints

1. The incidental construction of a graph smaller than G is acceptable
2. The incidental construction of a graph smaller than G is not acceptable but a graph with the same size as G (same number of edges and vertices) is acceptable
3. Any graph incidntally constructed must be larger than G.

In all cases, we assume flexible armed molecules (abstract, not embedded, graphs).

a $\hat{a}$

ATTCG GGTAACATTCG

TAAGCCCATTG TAAGC

Definitions

Sticky end types a, b, c, $\hat{c}$, $\hat{a}$, etc. label unpaired arms sticking off of molecules.

Types a and $\hat{a}$, are complementary sticky ends.

A bond-edge is an edge formed by joining two complementary sticky ends.

A tile represents a branched junction molecule with a specific set of sticky ends.

A pot P is a set of tiles such that for any sticky end type a on a tile in P, there is a sticky end of type $\hat{a}$ on some tile in P

A complex is an arrangement of tiles from a pot type P with as many adjoined complementary sticky ends as possible with the given tiles

A complete complex is a complex which has no unadjoined sticky ends

a $\hat{a}$

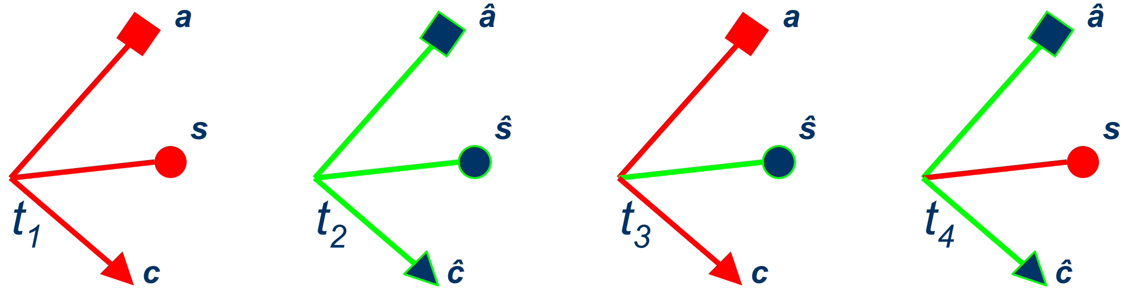
ATTCG GGTAACATTCG

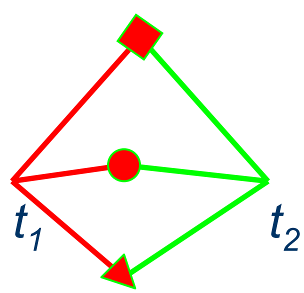
TAAGCCCATTG TAAGC

Example

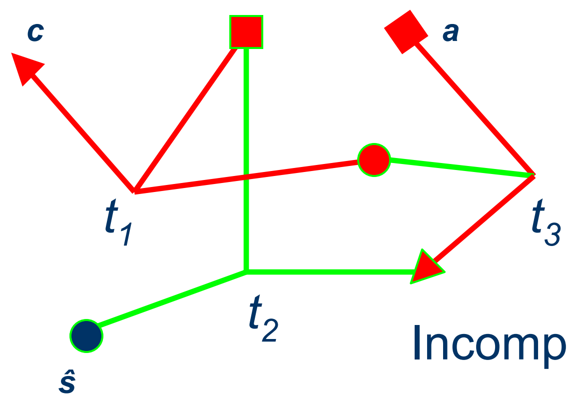
Both complete complexes and incomplete complexes can be constructed from the this pot P with 4 tiles:

P:





Complete complex



Incomplete complex

Simple constraints

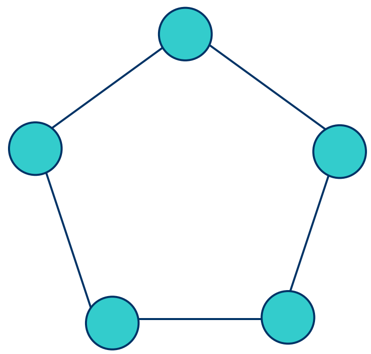
1. A graph G may be constructed as a complete complex from pot P if and only if the number of hatted sticky ends of each type used in the construction of G equals the number of unhatted sticky ends of the same type that appear in the construction.

2. The total number of hatted sticky end types must equal the total number of unhatted sticky end types in a complete complex.

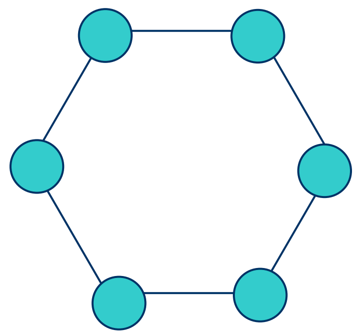
These constraints drive parity arguments.

You try one….

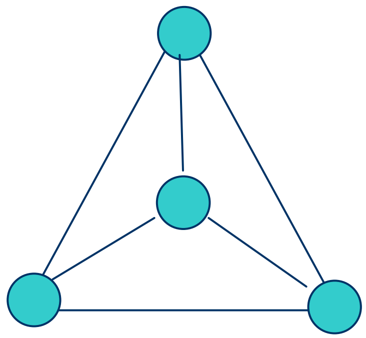
How many tiles to get it, but nothing smaller?



C_5



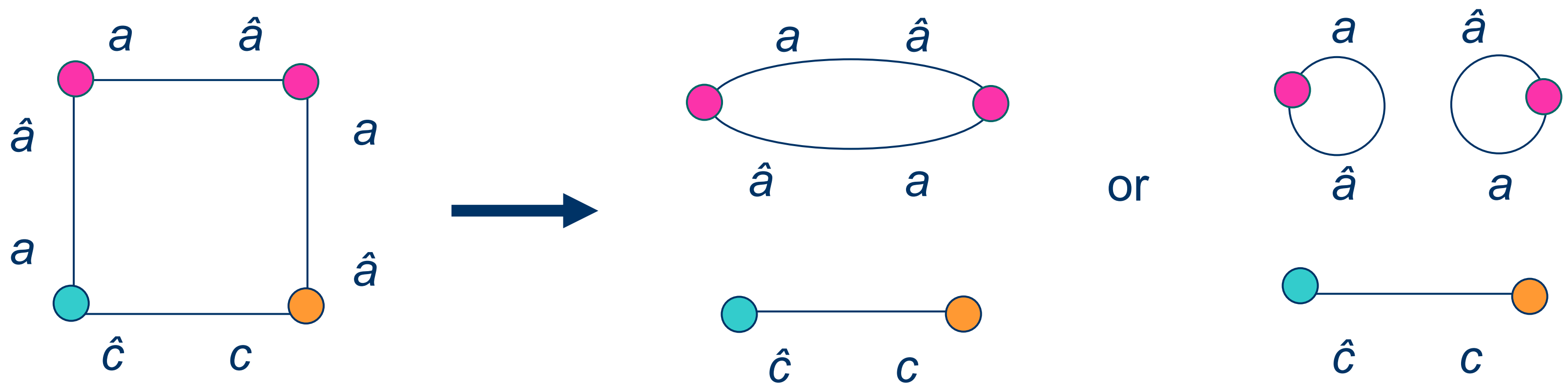
C_6



K_4

Some things to consider

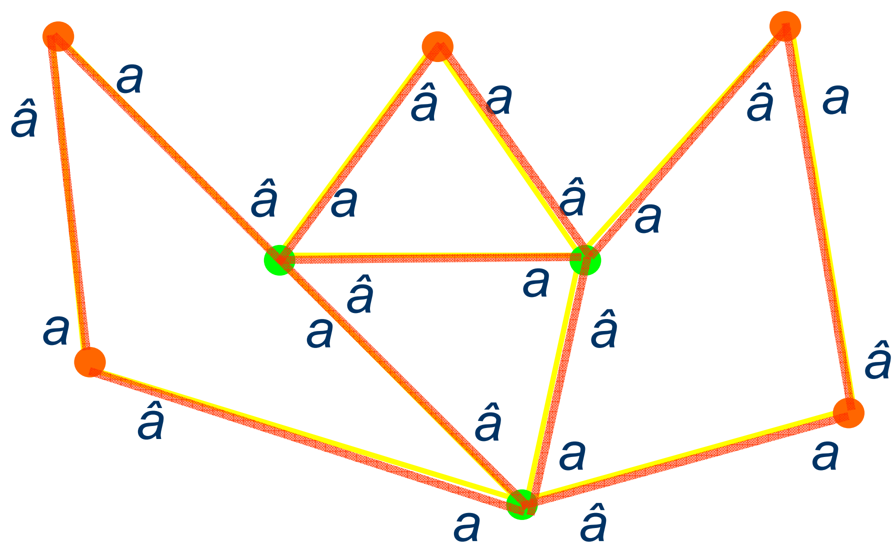
- Since every edge in a graph *G* represents the connection of two complementary sticky ends, a complete complex will be required to construct *G*.
- Since a tile can not represent two vertices of different degree can represent the same tile type, at least the number of different vertex degrees in *G* are needed.
- Under the restrictions of scenario 3, no two adjacent vertices can represent the same tile type because multi-edges and loops could be formed by swapping sticky ends.



Scenario 1 example

The vertex sequence of a graph *G* is the list of vertex degrees in *G*.

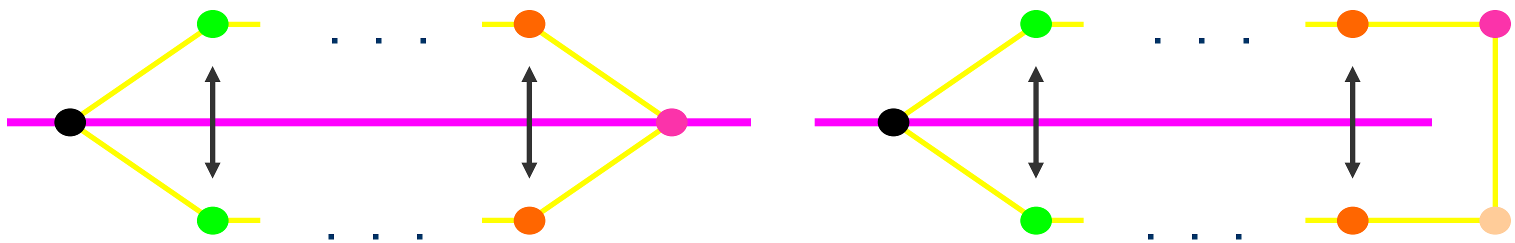
For Eulerian graphs the minimum number of tile types is just the number of different digits that appear in the vertex sequence. This can be shown by labeling sticky end types as we follow a graphs Euler circuit (labeling sticky end type *a* for outgoing sticky ends and *â* for incoming sticky ends).



Only 1 bond-edge type is required for Eulerian graphs, and only as many tile types as valencies!

Scenario 2 example

The minimum number of tile types required to construct a cycle such that no smaller graphs can be constructed out of the tiles is $\left\lceil \frac{n}{2} \right\rceil + 1$ where n is the number of vertices in the cycle C_n.

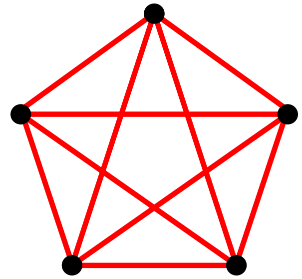


n Even *n* Odd

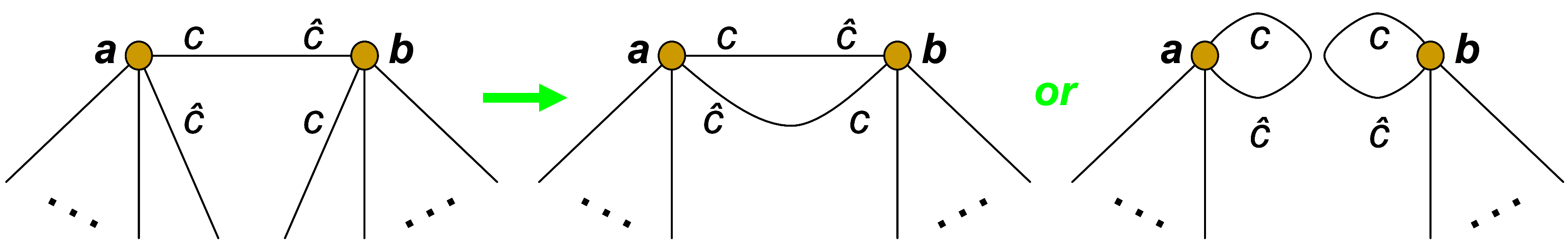
The bisecting line reflects identical tile types

The minimum number of bond-edge types in this case is $\left\lceil \frac{n}{2} \right\rceil$.

Scenario 3 example



Complete graphs K_n can only be constructed using n tile types and n-1 bond-edge types. Since every vertex in a complete graph is adjacent to every other vertex, no two vertices can represent the same tile type under the constraints of scenario 3. The image below shows the result of two tiles (***a*** and ***b***) of the same type appearing in K_n.



A complex other than K_n is formed!

Proof techniques

Get upper bounds by finding a set of tiles that suffice to build the graph.

Lower bounds/unwanted graphs are hard. A combination of number theory and linear algebra, on equations determined by equivalence of hatted and unhatted sticky ends of a given type in complete complex.

E.g. tiles $t_1 = \{a^{n-1}\}$ $t_2 = \left\{ \hat{a}^{\frac{n}{2}}, a^{\frac{n}{2}-1} \right\}$

suffice for K_n for n even, in Scenario 2. To show that no smaller graph on m vertices results from x tiles of type 1 and y tiles of type 2, we show this has a unique solution:

$$\left\{ \begin{array}{l} x + y = m \\ x(n-1) + y\left(\frac{n}{2} - 1\right) = y\left(\frac{n}{2}\right) \end{array} \right\} \qquad x = \frac{m}{n} \qquad y = \frac{m(n-1)}{n}$$

However, x and y must be integers, so this is a contradiction.

Table A: Minimum Tile Types

Scenario 1	$T_1(G)$ = minimum number of tile types required if complexes of smaller size than the target graph are allowed.
General graph *G*	The number of different vertex degrees ≤ $T_1(G)$ ≤ the number of different even vertex degrees + 2*(the number of different odd vertex degrees).
Trees	The number of different vertex degrees ≤ $T_1(G)$ ≤ the number of different vertex degrees + 1.
C_n	$T_1(C_n) = 1$.
K_n	$T_1(K_n) = 1$ if n is even, and $T_1(K_n) = 2$ if n is odd.
$K_{n,m}$	$T_1(K_{n,m}) = 1$ if $n=m$ and even, and $T_1(K_{n,m}) = 2$ otherwise.
K-regular graphs	$T_1(G) = 1$ if n is even, and $T_1(G) = 2$ if n is odd.
Scenario 2	$T_2(G)$= minimum number of tile types required if complexes of the same size as the target graph, but not smaller, are allowed.
Trees	$T_2(T)$ = the number of different lesser size subtree sequences.
C_n	$T_2(C_n) = \lceil n/2 \rceil + 1$.
K_n	$T_2(K_n) = 2$ if n is even, and $T_2(K_n) = 3$ if n is odd.
$K_{n,m}$ *with $n \neq m$*	$T_2(K_{n,m}) = 2$ if $gcd(m,n)=1$, and $T_2(K_{n,m}) = 3$ if $gcd(m,n)>1$.
$K_{n,n}$	$2 \leq T_2(K_{n,n}) \leq 3$.
Scenario 3	$T_3(G)$ = minimum number of tile types required if complexes of the same size as (or smaller than) the target graph are not allowed.
Trees	$T_3(T)$ = the number of induced subtree isomorphisms.
C_n	$T_3(C_n)= \lceil n/2 \rceil + 1$.
K_n	$T_3(K_n) = n$.
$K_{n,m}$	$T_3(K_{n,m}) = min(n,m)+1$.

Table B: Minimum Bond-Edge Types	
Scenario 1	$B_1(G)$= minimum number of bond-edge types required if complexes of smaller size than the target graph are allowed.
General graph *G*	$B_1(G)$= 1 for all graphs.
Scenario 2	$B_2(G)$ = minimum number of bond edge types required if complexes of the same size as the target graph, but not smaller, are allowed.
Trees	$B_2(T)$ = the number of different sizes of lesser size subtrees.
C_n	$B_2(C_n) = \lceil n/2 \rceil$.
K_n	$B_2(K_n)$ = 1 if n is even, and $B_2(K_n)$ = 2 if n is odd.
$K_{n,m}$	$B_2(K_{n,m})$ = 1 if $gcd(m,n)$=1 , and $B_2(K_{n,m})$ = 2 if $gcd(m,n)$>1.
Scenario 3	$B_3(G)$ = Minimum number of bond edge types required if complexes of the same size as (or smaller than) the target graph are not allowed.
C_n	$B_3(C_n) = \lceil n/2 \rceil$.
K_n	$B_3(K_n) = n - 1$.

Thus far, the same pots have achieved both minimum tile types and minimum bond-edge types, but we don't know if this is always possible.

Pending…

- Various lattices, both 2 and 3 dimensional (as incomplete complexes?)
- Tubes (C_m x P_n) (ditto)
- C_m x C_n
- Various Platonic and Archimedean solids

And a whole other kettle of fish…

- Same set up and questions, but now assume rigid armed molecules—i.e. a fixed rotation (or location) of the sticky end types about a tile vertex.
- Edge-length constraints—because the helixes have to twist, if we call a twist a unit, each edge is of integer length.
- Rigid edges.

A different assembly method

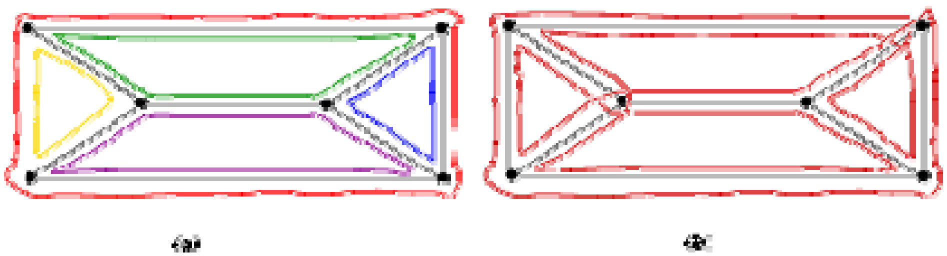


Fig. 1. One graph made by DNA strands in two ways.

'zipping together' single strands of DNA

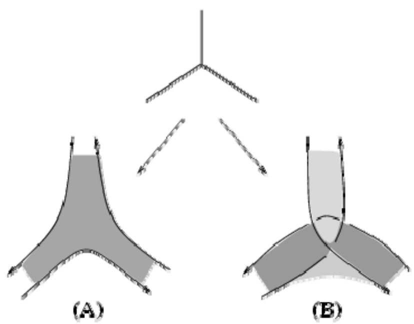
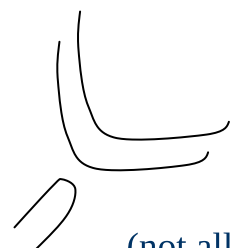


Fig. 2. Neighborhood of a vertex



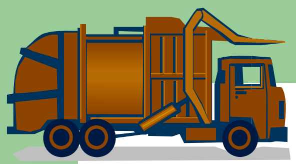
(not allowed)

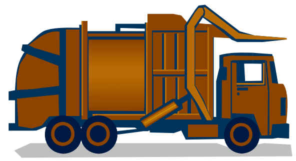
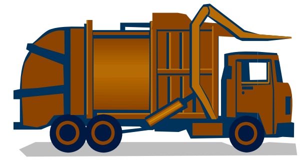
N. Jonoska, N. Saito, '02

A characterization

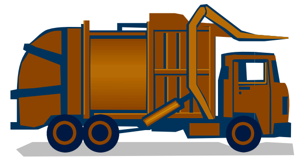
- A theorem of C. Thomassen specifies precisely when a graph may be constructed from a single strand of DNA, and theorems of Hongbing and Zhu to characterize graphs that require at least m strands of DNA in their construction.

- 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.*



You never know….



Oriented Walk Double Covering and Bidirectional Double Tracing

Fan Hongbing, Xuding Zhu, 1998

“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.”

Bibliographical References

F. Hongbing, X Zhu, "Oriented Walk Double Covering and Bidirectional Double Tracing," J. Graph Theory **29** (1998) 89-102.

N. Jonoska, G. L. McColm, A. Staninska. "The Spectrum of a Pot with DNA molecules," University of South Florida Department of Mathematics.

N. Jonoska, G. L. McColm, A. Staninska. "The Graph of a Pot with DNA molecules," University of South Florida Department of Mathematics. 2006.

D. Luo, "The road from biology to materials," Materials Today, **6** (2003), 38-43.

C. Thomassen, "Bidirectional Retraction-Free Double Tracings and Upper Embeddability of Graphs," J. Combin. Theory, Ser. B **50** (1990) 198 -207.

http://www.nanopicoftheday.org/2004Pics/April2004/DNAmesh.htm

http://seemanlab4.chem.nyu.edu/nanotech.html

Acknowledgement:

The project described was supported in by the Vermont Genetics Network through NIH Grant Number 1 P20 RR16462 from the INBRE program of the National Center for Research Resources, and by a National Security Agency Standard Grant.