Graph cycles, ergodicity and energy landscapes

Konstantin Klemm and Peter F. Stadler

Department of Bioinformatics

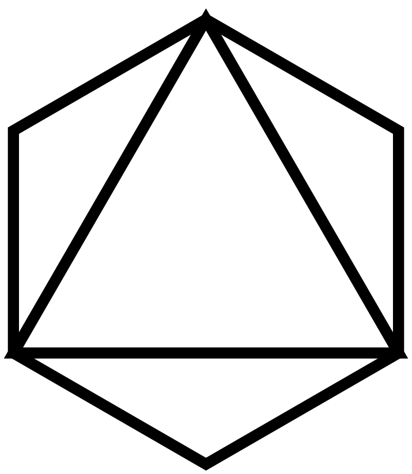
University of Leipzig

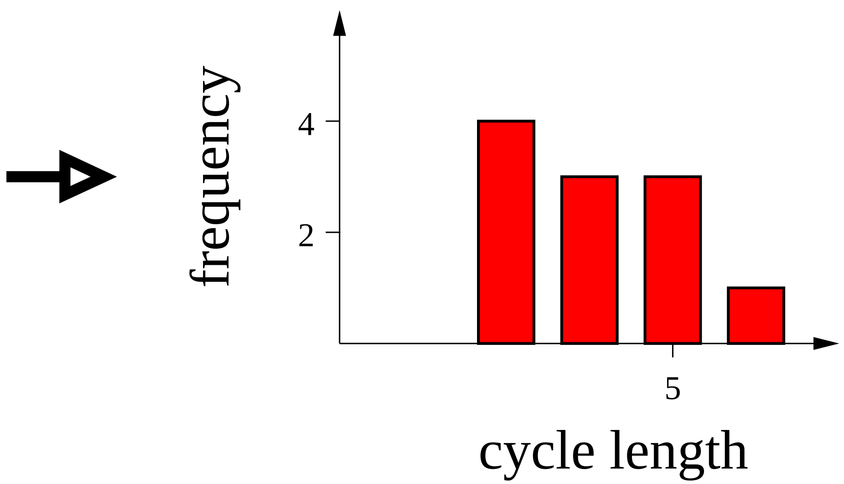
Problem setting

- Given: Simple graph $G(V, E)$

- Wanted: Cycle length distribution

How many cycles of length h does the graph contain?





Outline

- Motivation: growth exponents, model validation
- Method: Markov chain Monte Carlo in cycle space
- Problem: find ergodic move set
- Energy landscapes

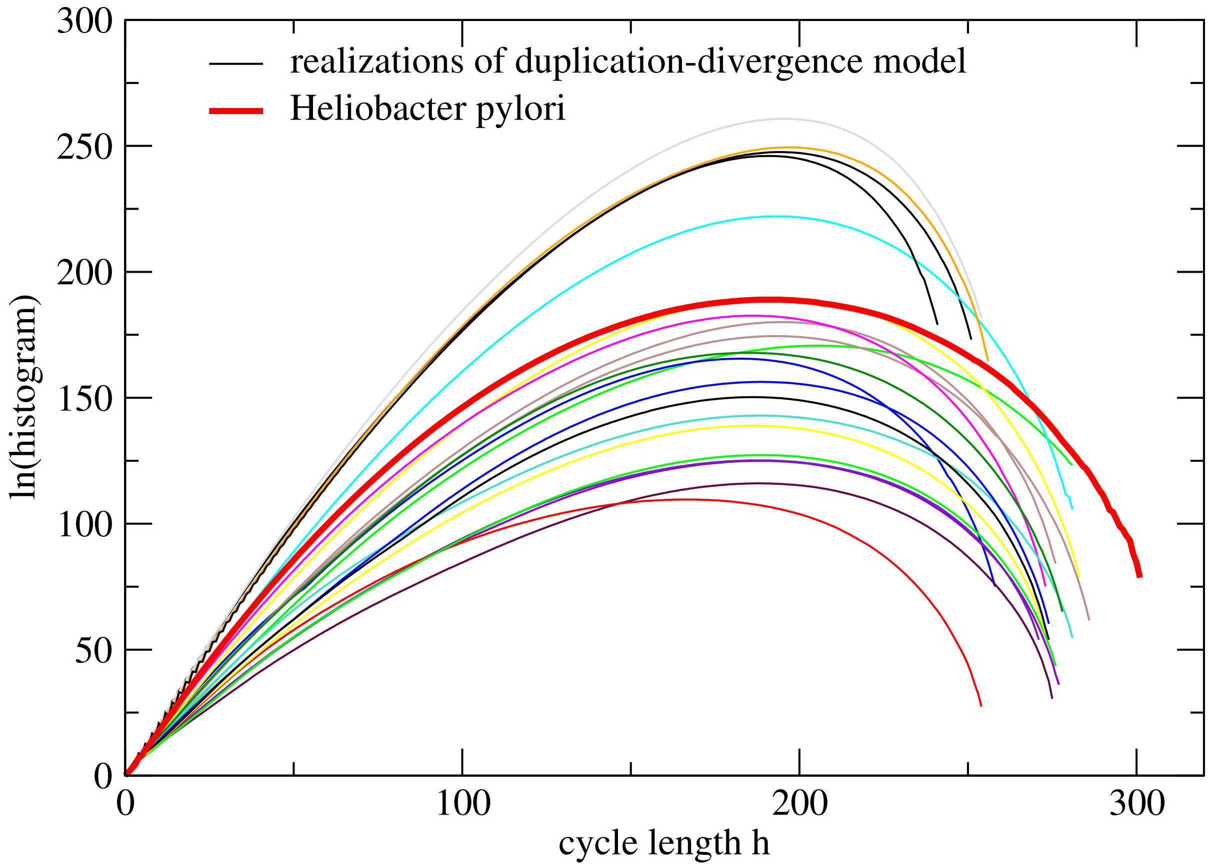
Motivation (1): model verification

Verification of graph models

- Internet, WWW
- social and economic (trade) networks
- metabolic and gene regulatory networks
- protein-protein interaction (PPI) networks
- ...

Cycles in real network and model

PPI network of bacterium H. Pylori compared with instances from duplication-divergence model (Vázquez et al., 2003).



Motivation (2): growth exponents

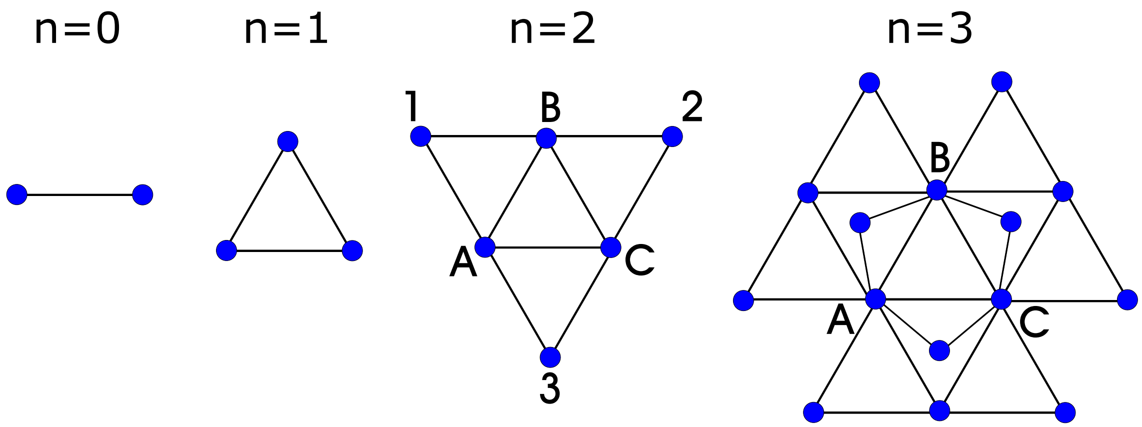
Consider graphs growing by iterative addition of nodes and edges. Average cycle length increases as

$$\langle h \rangle \sim N^{\alpha}$$

as a function of graph size N, with exponent

$$\alpha \in [0, 1]$$

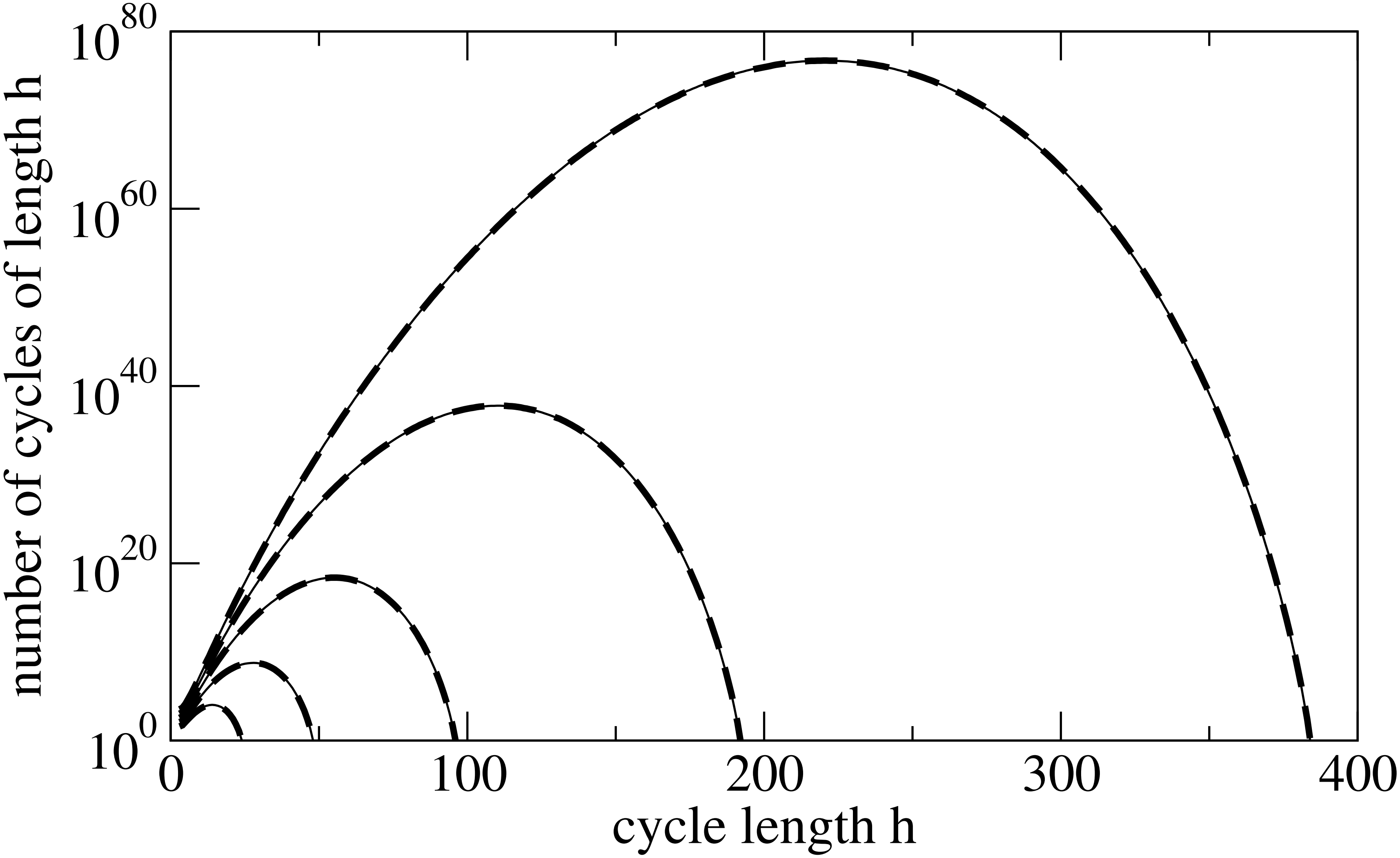
characteristic of the growth rule.



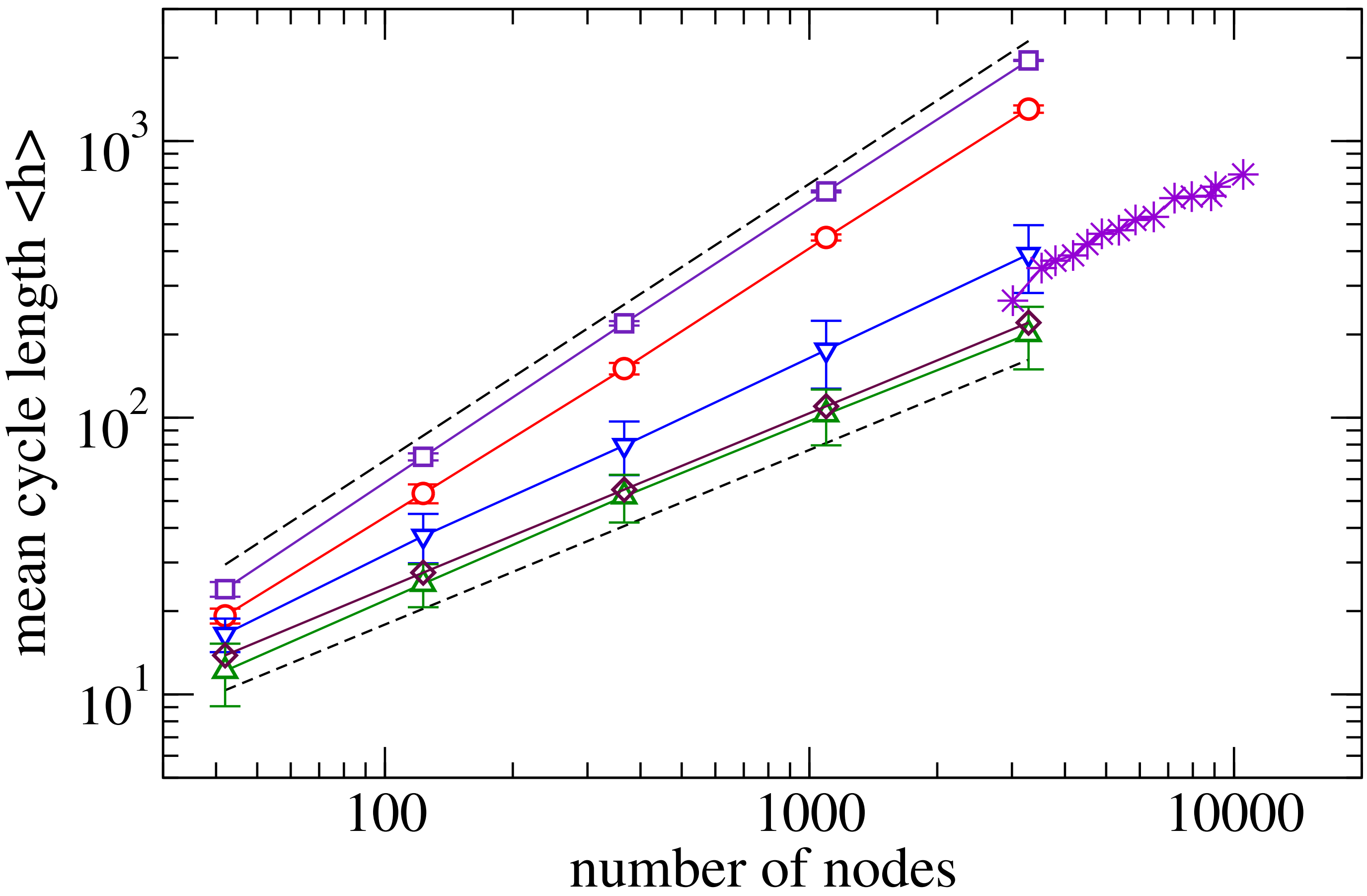
Example: growing 2-tree

$$\alpha = \frac{\ln 2}{\ln 3}$$

Cycle lengths in growing 2-tree



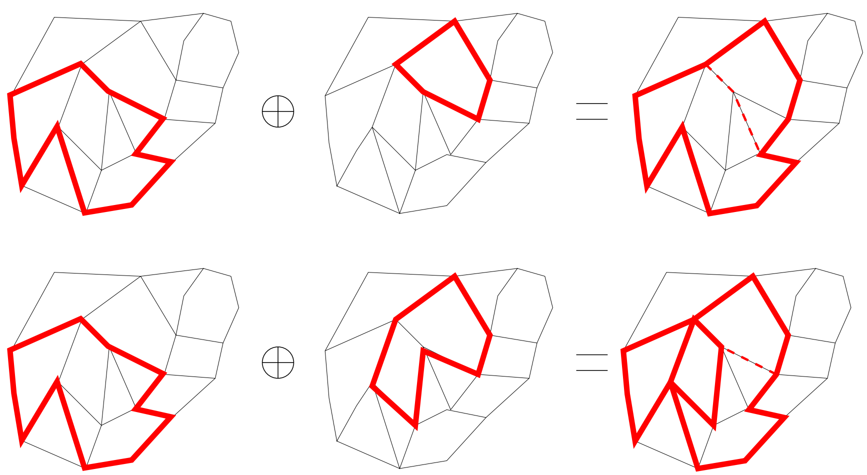
Growth under different rules



Markov chain Monte Carlo in cycle space

- microstate = cycle

- Markov step = addition of “detour” to current cycle.



- But: Sum of two cycles is not necessarily a cycle.

The algorithm ($\beta = 0$)

1. Choose generating system M of cycle space

2. Set initial cycle $C_0 := 0$ (empty cycle) and $t = 0$

3. (Propose) Draw random element $P \in M$

4. (Accept) If $C + P$ is a simple cycle or empty, set $C_{t+1} := C_t + P$
   (Reject) Otherwise, set $C_{t+1} = C_t$

5. Increment t and resume at 3 (or stop if desired chain length reached)

Ergodicity and the generating system M

Ergodicity of the Markov process is ensured if each cycle is a sum of cycles in M such that each partial sum is also a cycle:

Generating system M of cycle space of G is *ergodic* if
For all cycles C in G there are $t \in \mathbb{N}$ and $(C_1, C_2, \ldots, C_t) \in M^t$ such that for all $i \in \{1, \ldots t\}$

$$S_i := \sum_{j=1}^{i} C_j$$

is a cycle and

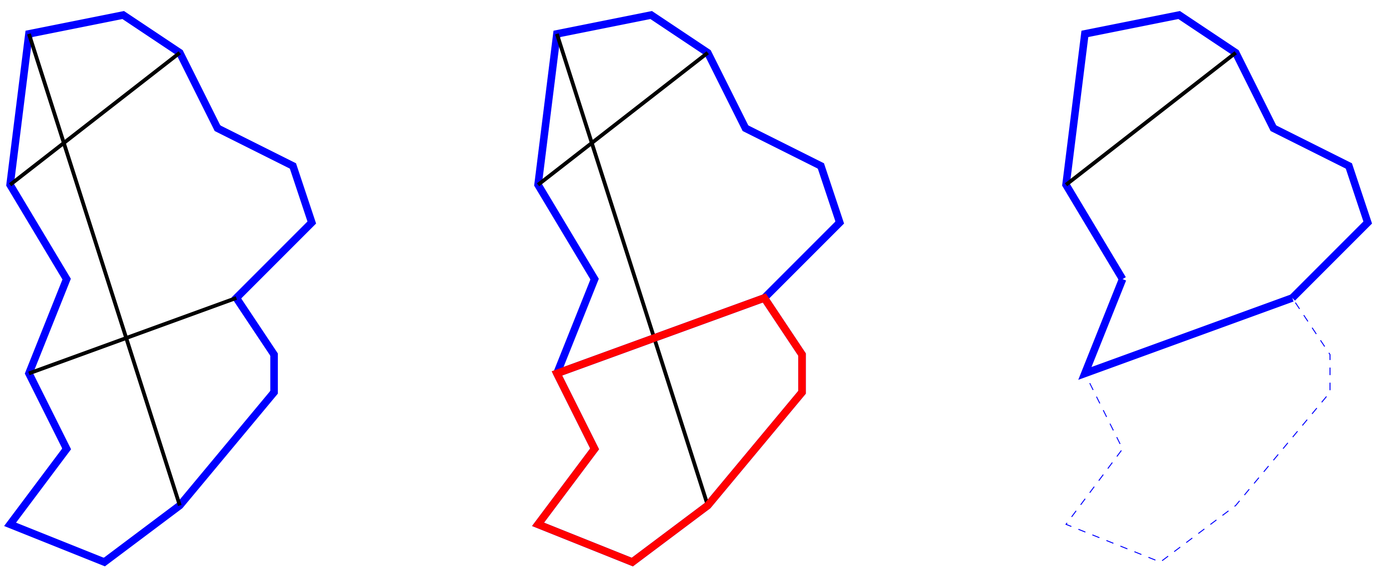
$$C = S_t \ .$$

<u>Problem:</u> Find possibly small ergodic M.

M = set of chordless cycles (1)

- A cycle C is *chordless* if there is $W \subseteq V$ such that C is the subgraph induced by W.

- The set of all chordless cycles is an ergodic generating system of cycle space.

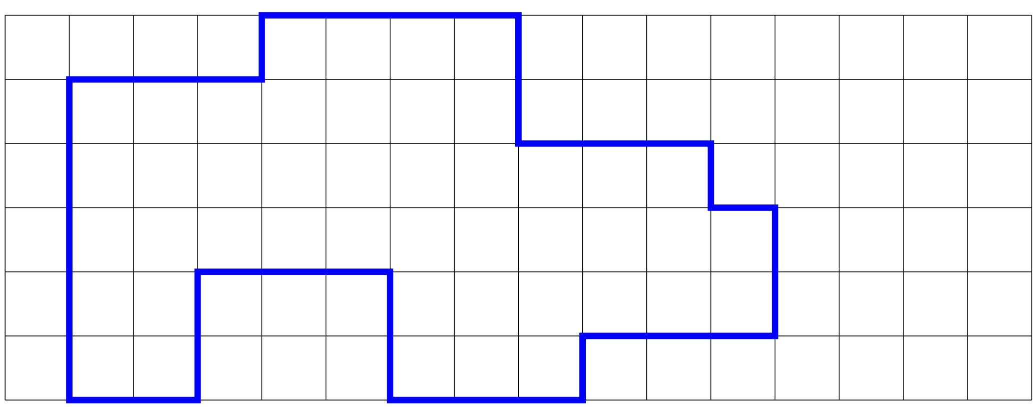


M = set of chordless cycles (2)

- decomposition into shorter and shorter cycles
  $\Rightarrow$ energy (=length) landscape without local minima

- But: set of chordless cycles can be too large to be generated beforehand

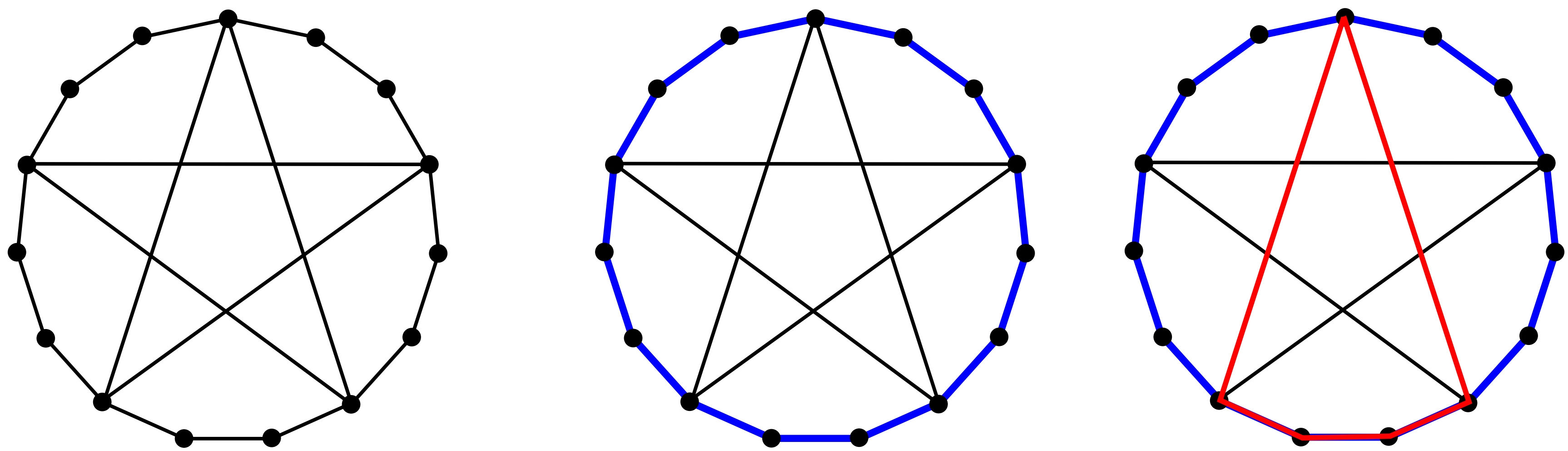
- example: square lattice



Short (geodesic) cycles?

- A cycle S is short (geodesic) if it is not the sum of two shorter cycles.

- Set of short cycles ergodic? No, counterexample:

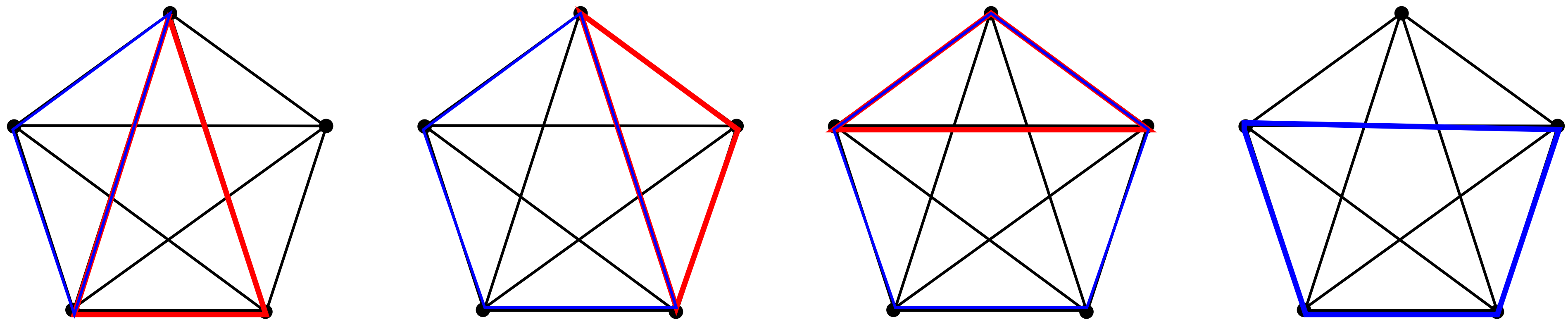


Complete graphs

- Complete graph K_N on N nodes has $O(N^3)$ chordless cycles (triangles).

- Choose a vertex x. The set M of all triangles containing x is an ergodic generating system of the cycles of K_N.

- Resulting energy landscape <u>with local minima</u>



Summary

- Introduced Monte Carlo method for obtaining statistics of cycles in graphs.

- Remaining task: general ergodicity criterion for move sets.

- Move set involving all chordless cycles is ergodic and generates completely smooth energy landscape (no local minima).

- Smaller move sets known for some cases: complete graphs, planar graphs. Smoothness of energy landscape is lost.

…

Robust cycle bases

Kainen (2000): A cycle basis $\mathcal{B}$ is robust if for every [simple] cycle Z there is a linear ordering of the subset $\mathcal{C}(G, \mathcal{B}, Z)$ such that, as each element in the resulting sequence is added to form the sum Z, it intersects the *sum* of those preceding it in a nontrivial path. In this case, the partial sums must be cycles. A cycle basis is called *cyclically robust* when the sum of the new cycle and those that went before remains a cycle.

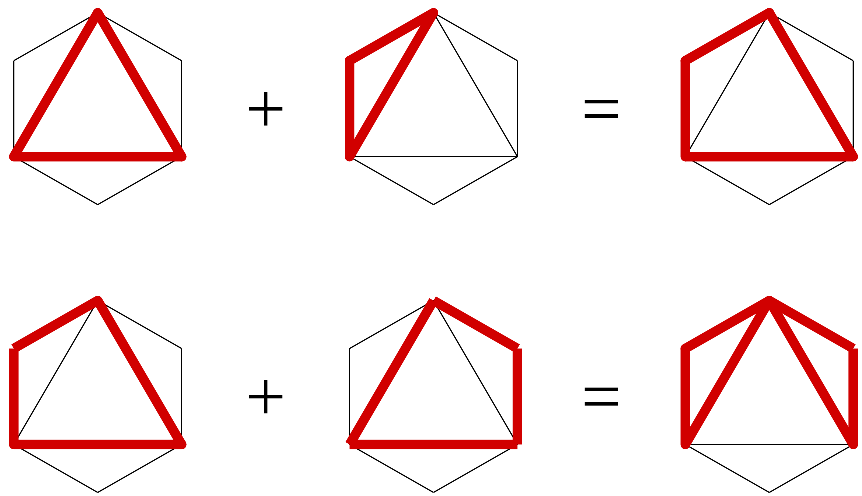
Relevance here:
basis (cyclically) robust $\Rightarrow$ ergodic Monte Carlo

Robust cycle bases — known results

- planar graphs: planar basis, basis cycles are outlines of faces in a planar embedding
- complete graphs (Kainen): pick arbitrary vertex x, basis cycles are all triangles containing x
- slightly more general: graphs spanned by a star (argument analogous to complete graphs)
- No general criterion for existence of (cyclically) robust bases

Monte Carlo — summing cycles

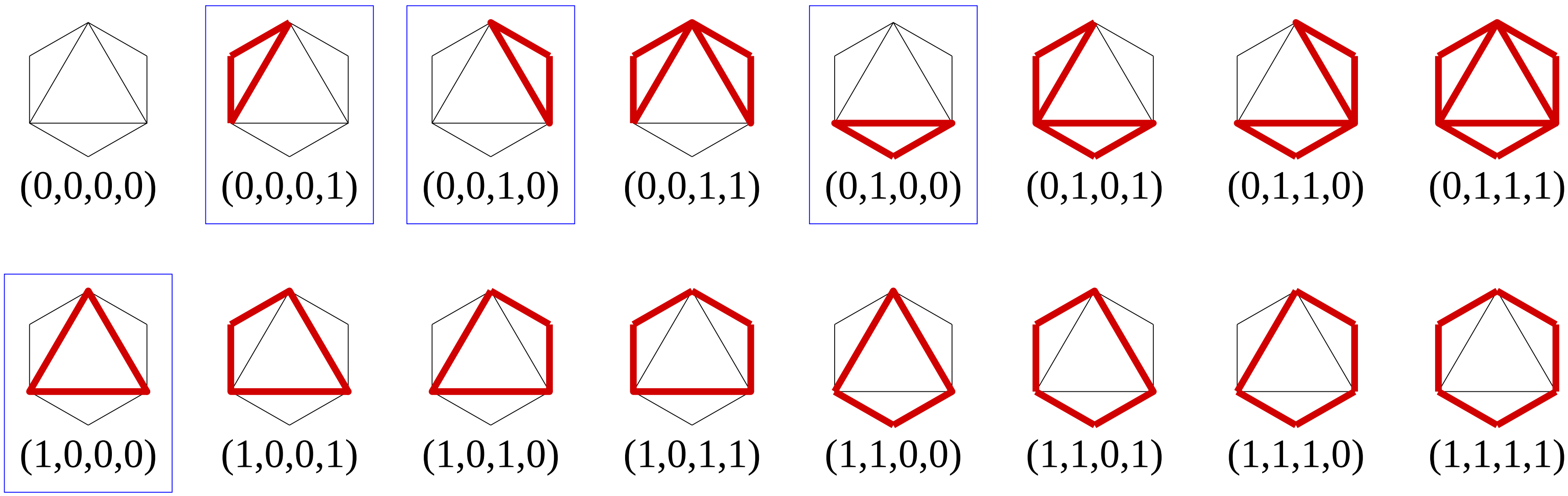
- Sum of two cycles yields new cycle:



- (generalized) cycle: subgraph, all degrees even

- simple cycle: connected subgraph, all degrees = 2.

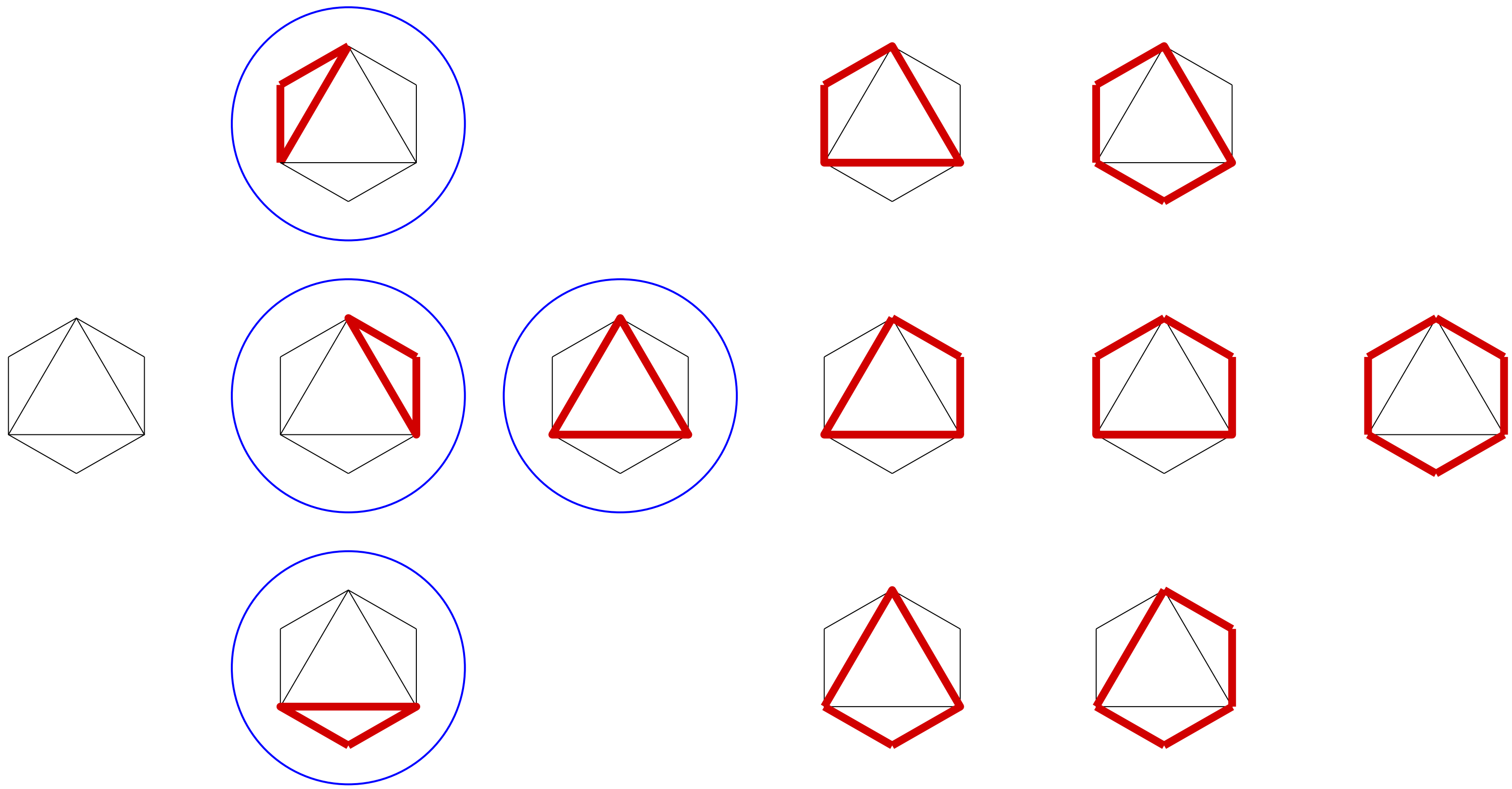
Monte Carlo — cycle space



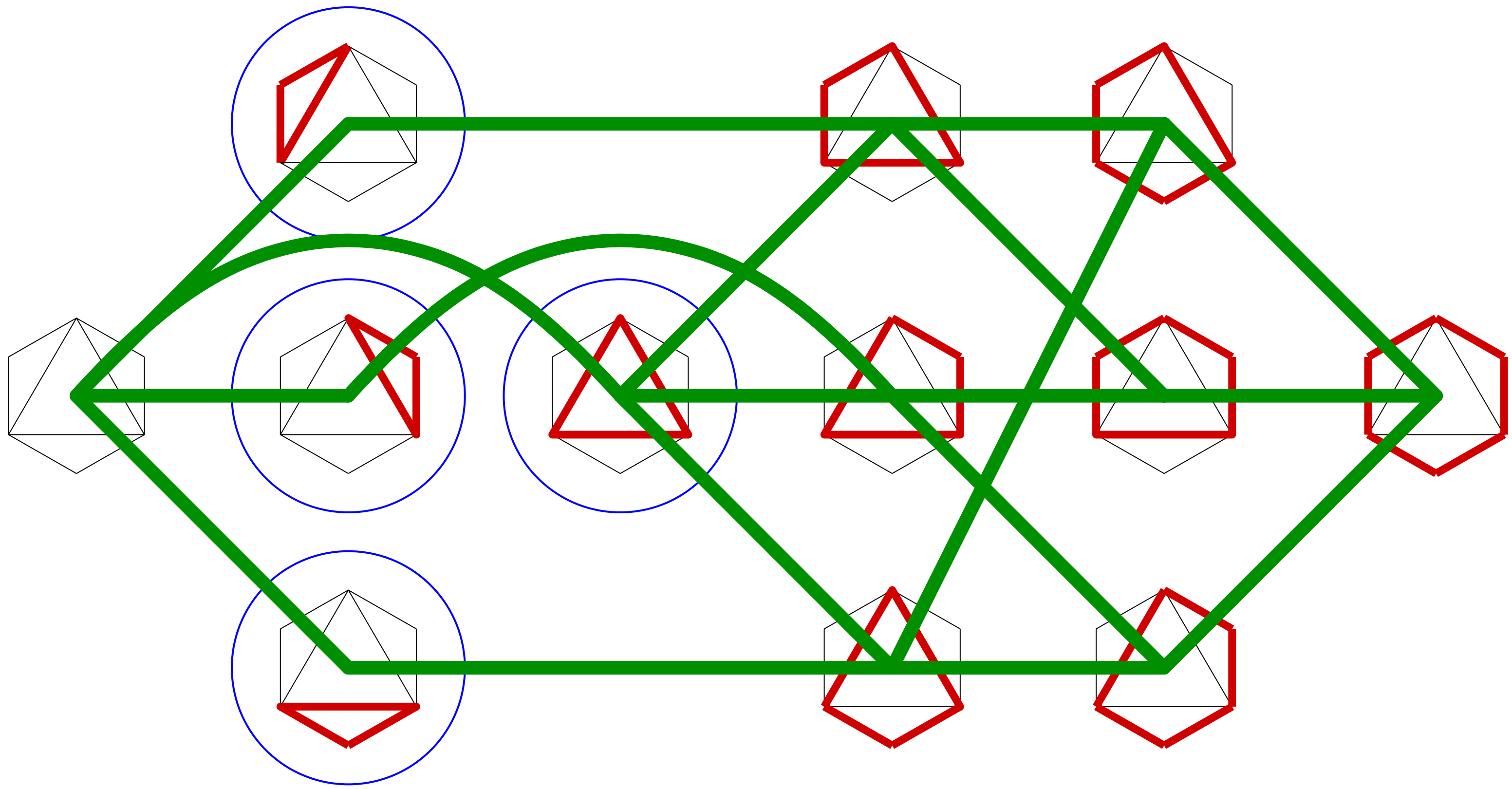
- cycle space: contains all (generalized) cycles
- finite-dimensional vector space, has cycle basis

Linear Algebra kicks butt!!!

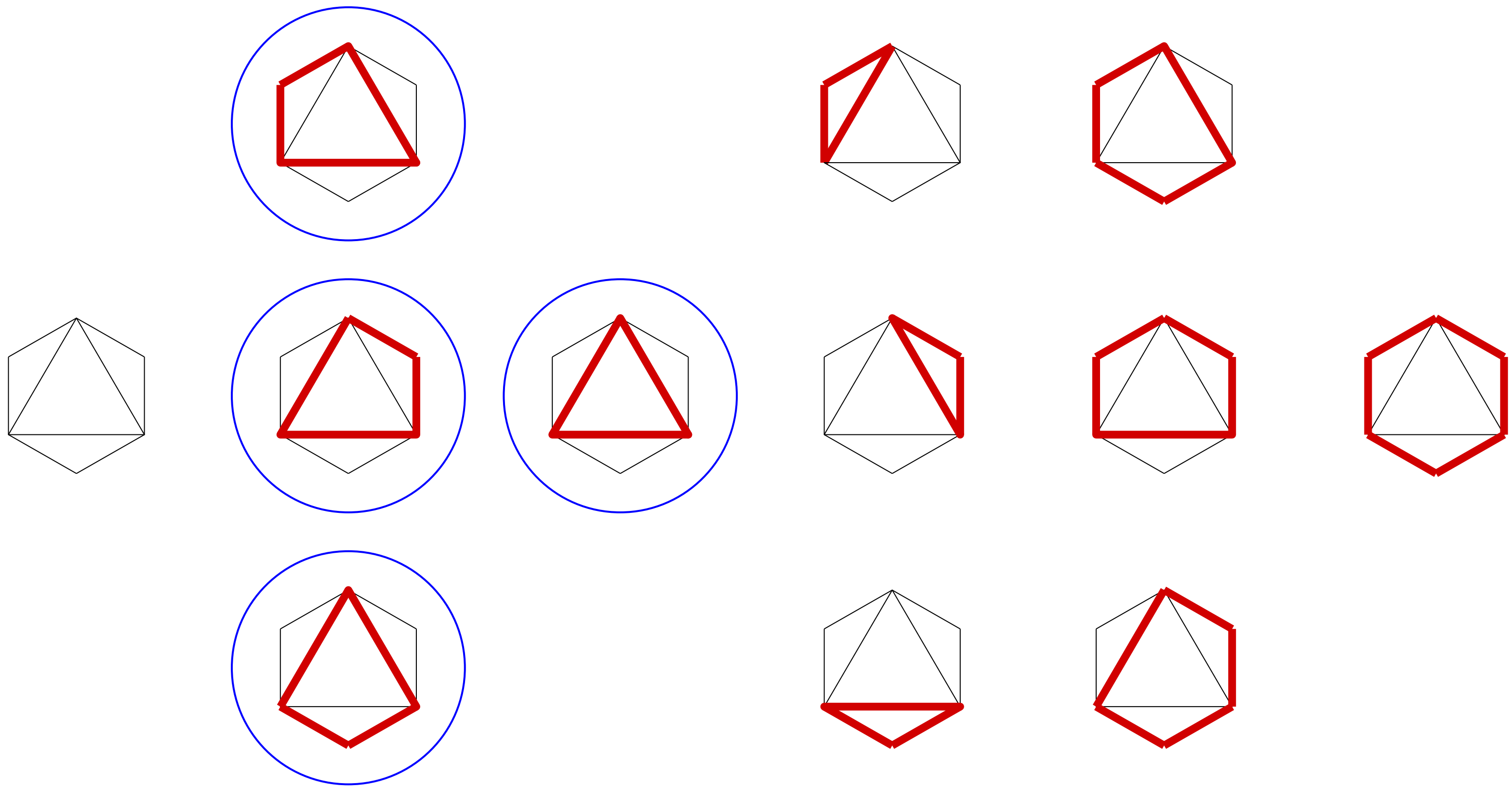
Connected transition graph



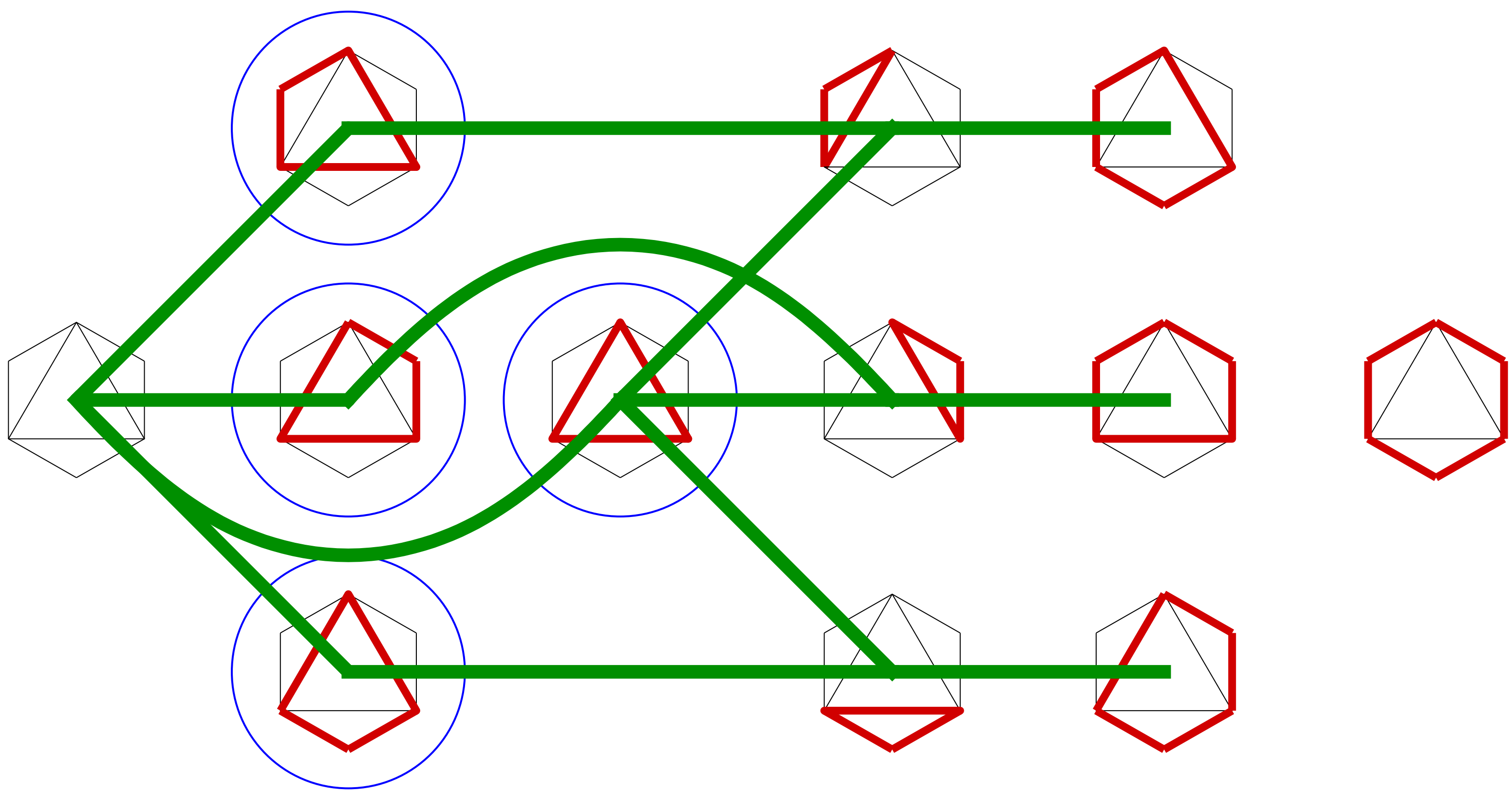
Connected transition graph



Disconnected transition graph



Disconnected transition graph



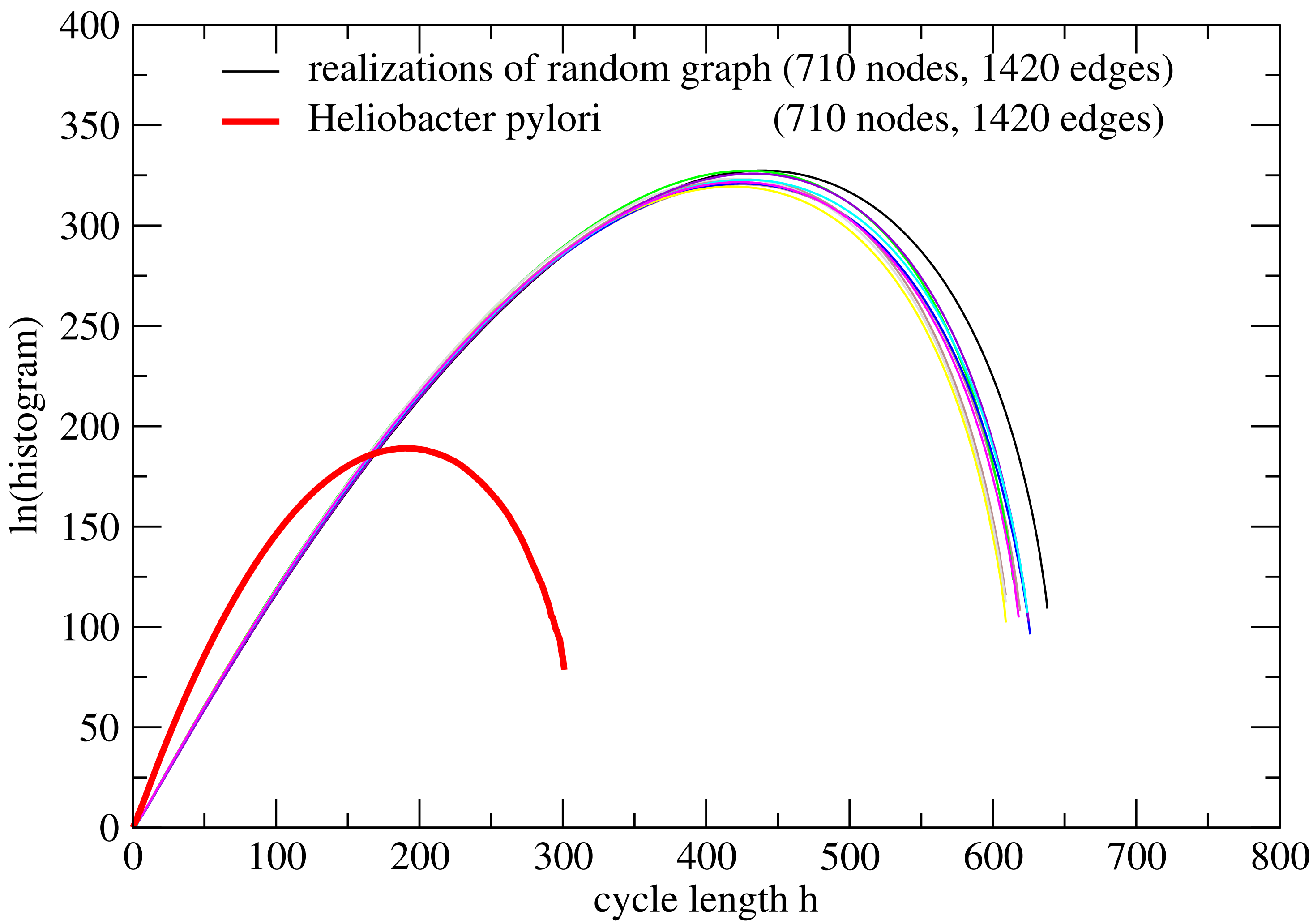
Application: protein-protein interactions

- H. pylori network from DIP, version April 24, 2005

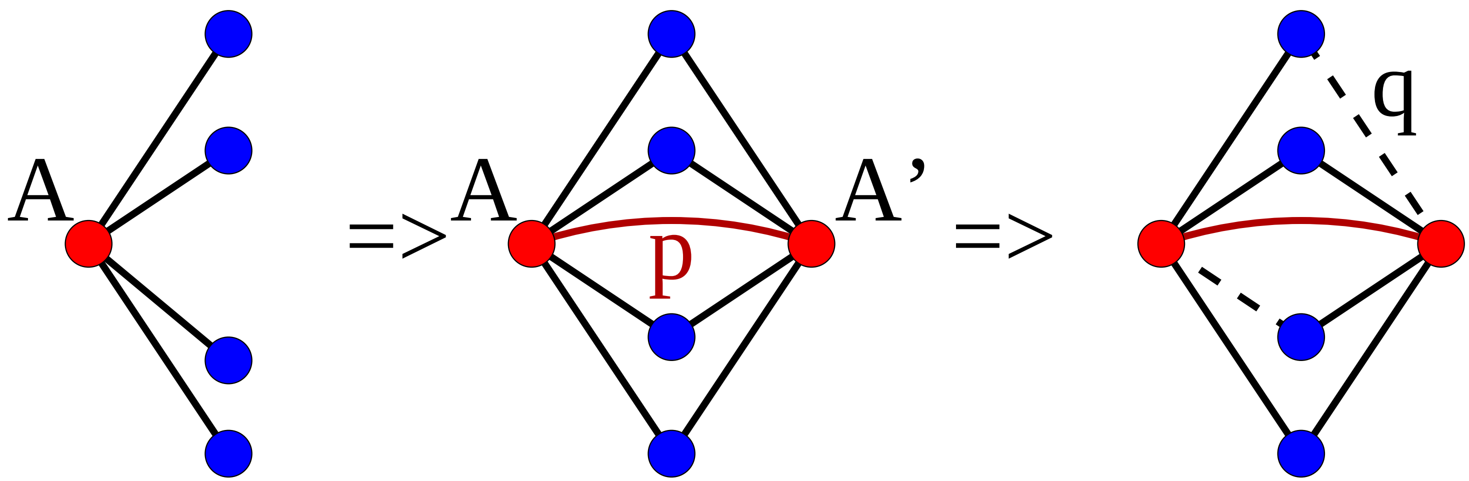
- 710 proteins

- 1420 interactions

PPI vs. random graphs



Duplication-divergence model

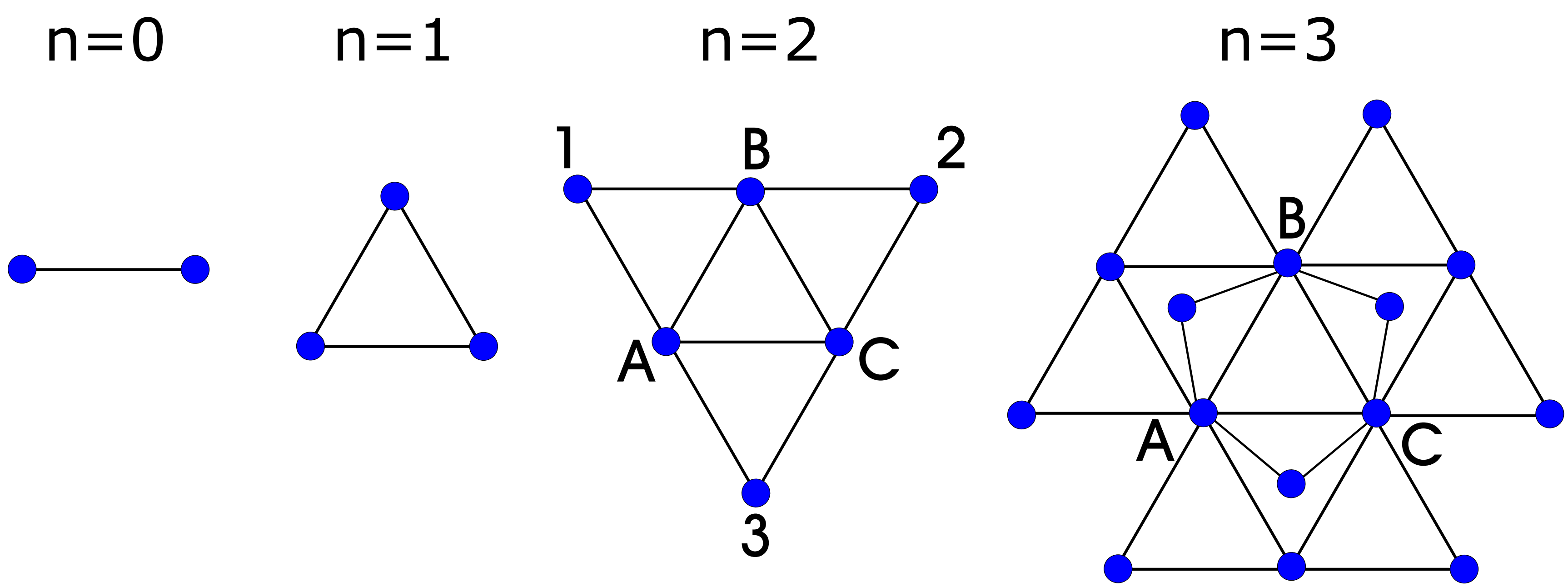


1. duplication: generate new node A' with same neighbors as randomly chosen node A. Add edge AA' with probability p.

2. divergence: For each pair of edges $(AX, A'X)$, with probability q remove one of the edges (chosen with prob. 1/2); resume at 1.

Vázquez et al. (2003)

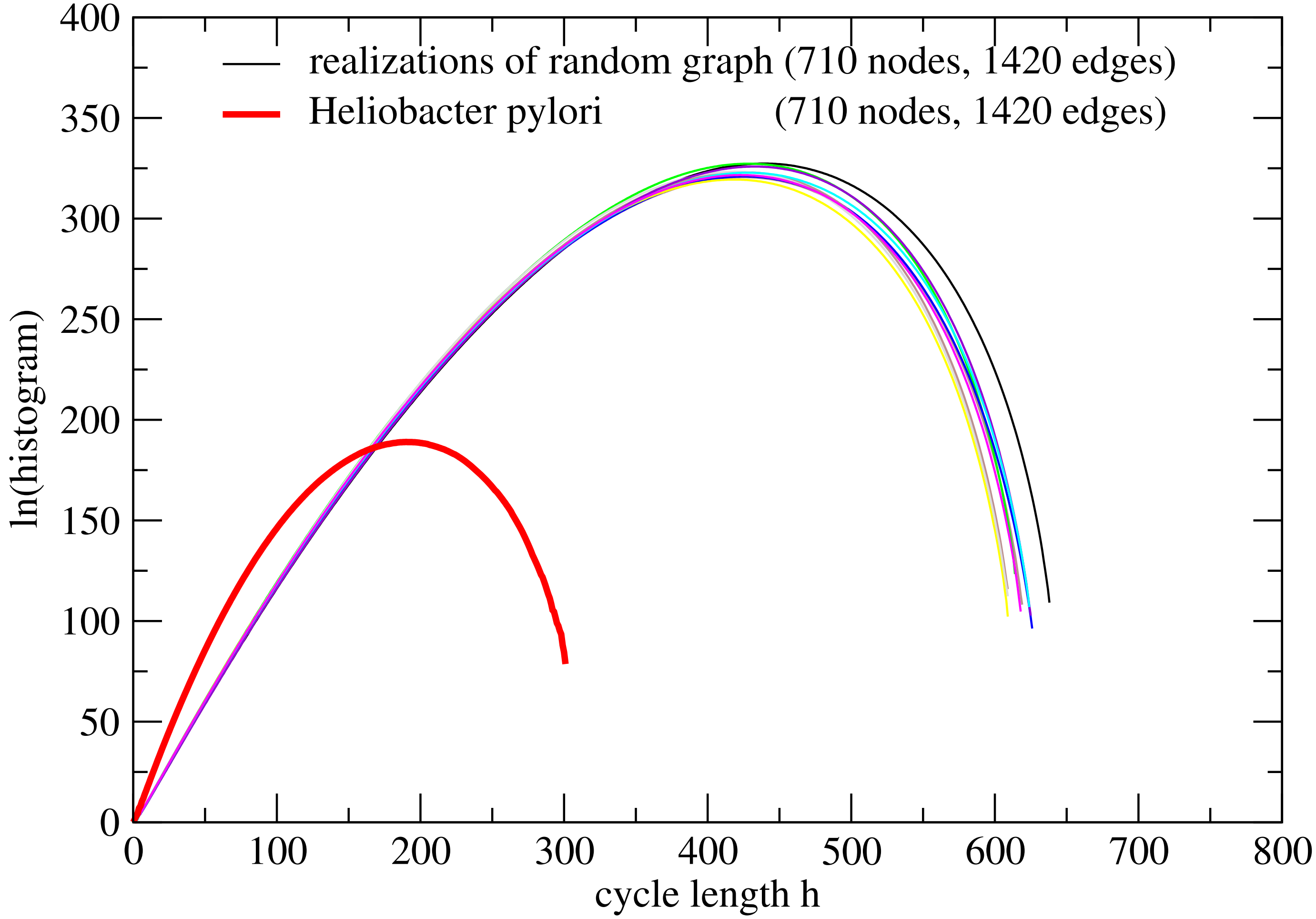
Application: growing graphs



$$c^{(n+1)}(h) = \sum_{l=3}^{h} \binom{h}{h-l} c^{(n)}(l) \text{ for } l \geq 4, \text{ and}$$

$$c^{(n+1)}(3) = c^{(n)}(3) + 3^n$$

Protein Interaction vs. random graphs



H. pylori network from DIP, version April 24, 2005, 710 proteins, 1420 interactions