From Las Vegas to Monte Carlo and back: Sampling cycles in graphs

Konstantin Klemm

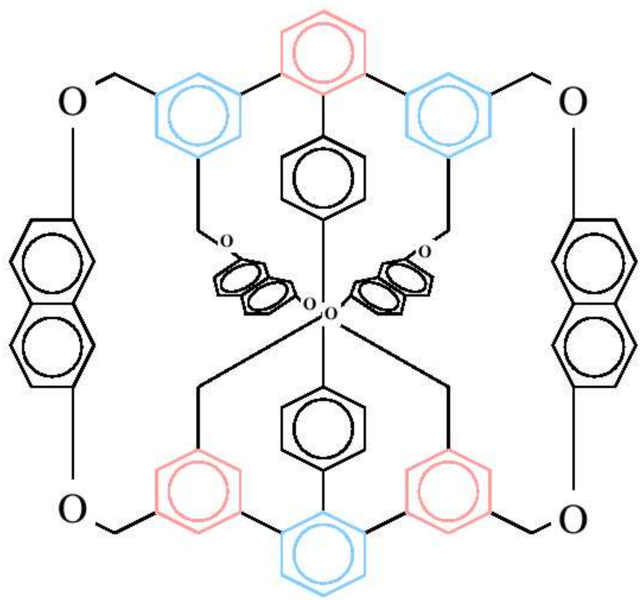
SEHRLÄSSIGER
LEHRSÄSS--EL for Bioinformatics

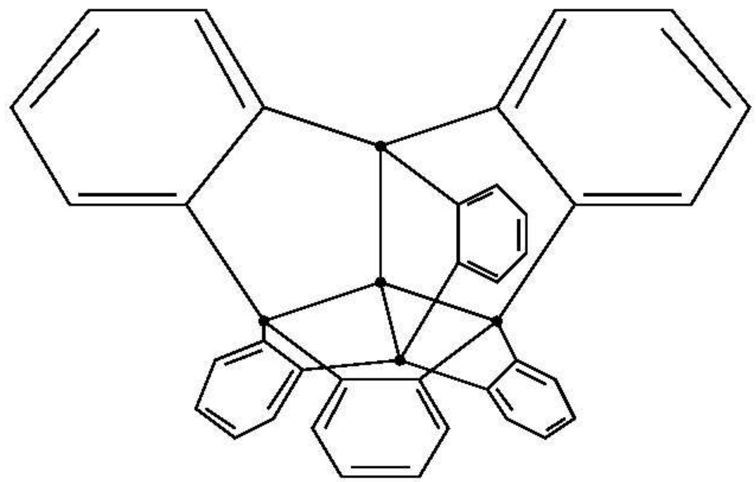
University of Leipzig

Agenda

- Cycles in graphs: Why? What? Sampling?
- Method I: Las Vegas
- Method II: Monte Carlo
- Robust cycle bases

Why care about cycles? (1)

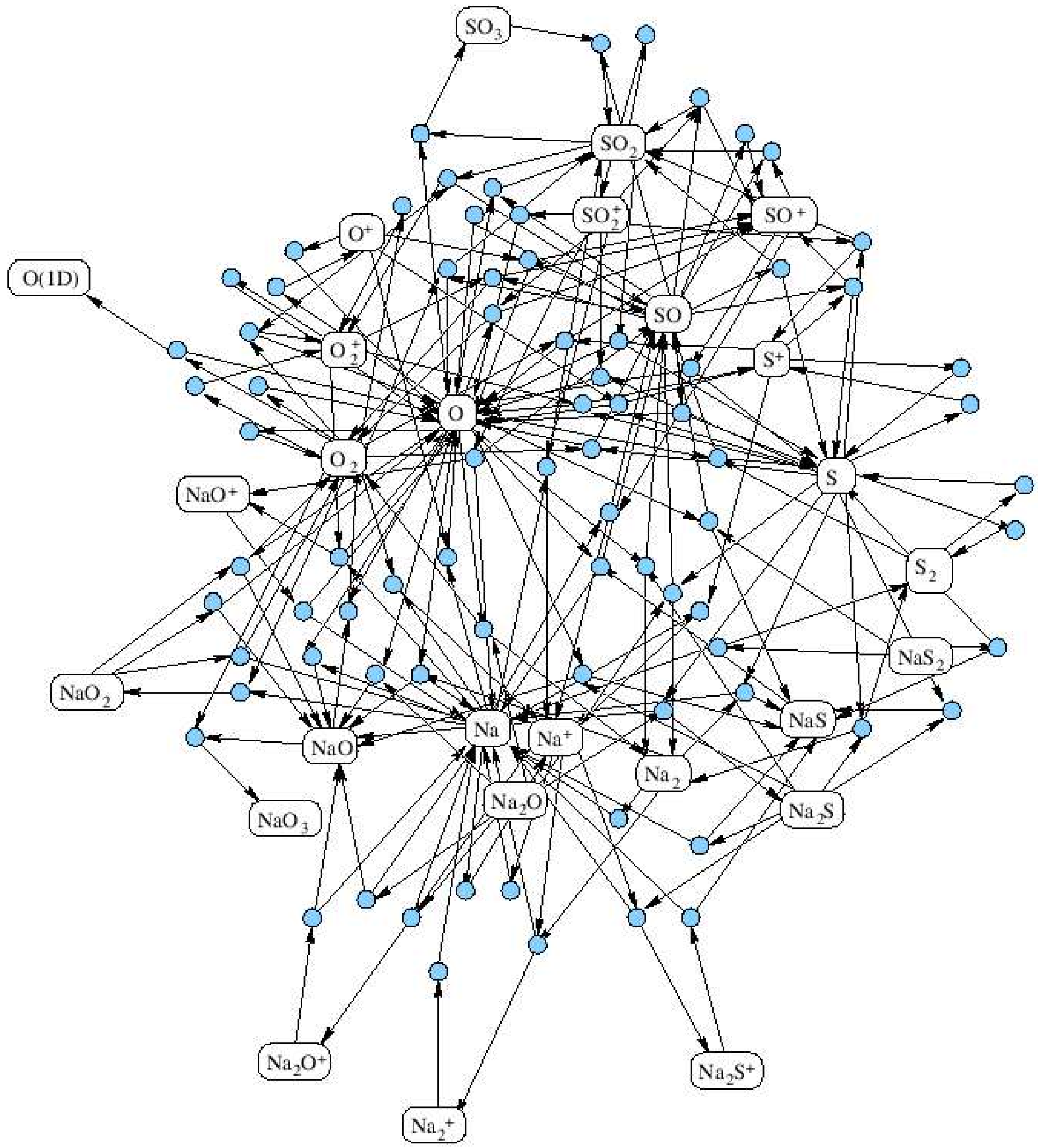




Chemical Ring Perception

Why care about cycles? (2)

Analysis of chemical reaction networks (Io's athmosphere)



Why care about cycles? (3)

- Protein interaction networks
- Internet graph
- Social networks
- ...

Aim: Detailed comparison between network models and empirical networks with respect to presence / absence of cycles

Sampling

- Interested in average value of a cycle property $f(C)$

$$\langle f \rangle = \frac{1}{|\{\text{cycles}\}|} \sum_{C \in \{\text{cycles}\}} f(C)$$

with $f(C) = |C|$ or $f(C) = \delta_{|C|,h}$ or $\ldots$

- exhaustive enumeration of {cycles} not feasible

- approximate $\langle f \rangle$ by summing over representative, randomly selected subset $S \subset \{\text{cycles}\}$

- How do we generate S then?

Las Vegas

- Sampling method based on self-avoiding random walk [Rozenfeld et al. (2004), `cond-mat/0403536`]

- probably motivated by the movie "Lost in Las Vegas" (though the authors do not say explicitly)

1. Choose starting vertex s.

2. Hop to randomly chosen neighbour, avoiding previously visited vertices except s.

3. Repeat 2. unless reaching s again or getting stuck

Las Vegas – trouble

- Number of cycles of length h in complete graph K_N

$$W(h) = (2h)^{-1}\frac{N!}{(N-h)!}$$

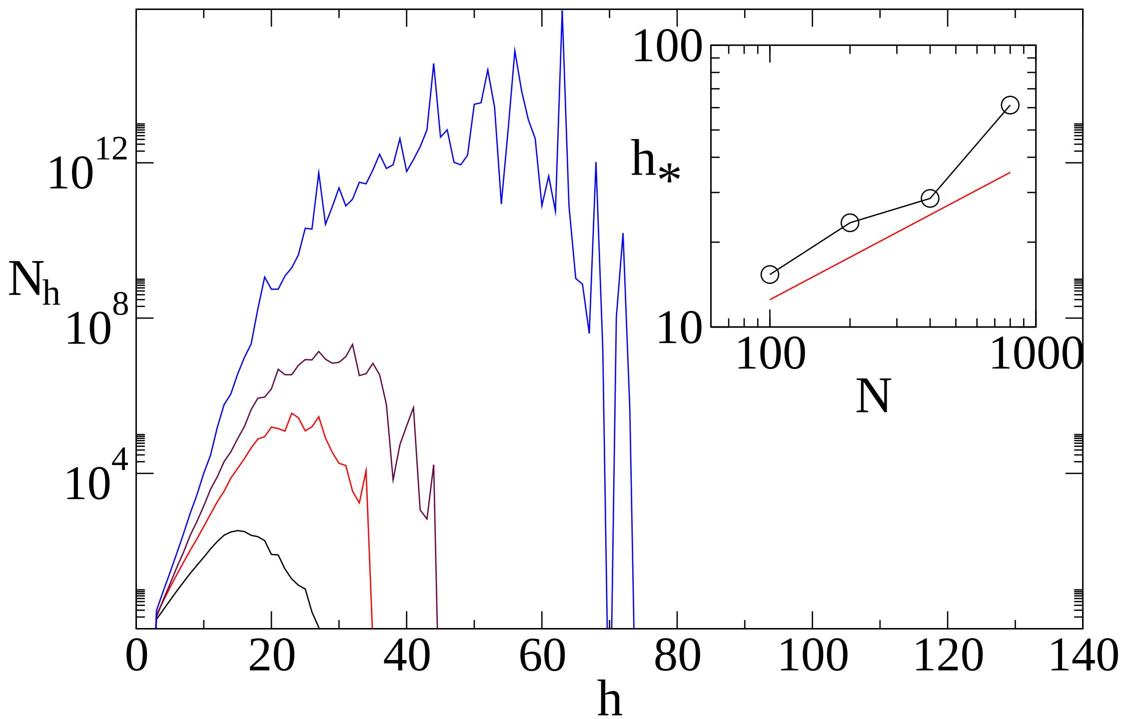
- For $N = 100$, $W(100)/W(3) \approx 10^{150}$

- Flat cycle length distribution in K_N from Rozenfeld method

$$p(h) = \frac{1}{(N-2)}$$

- Undersampling of long cycles

Las Vegas — results

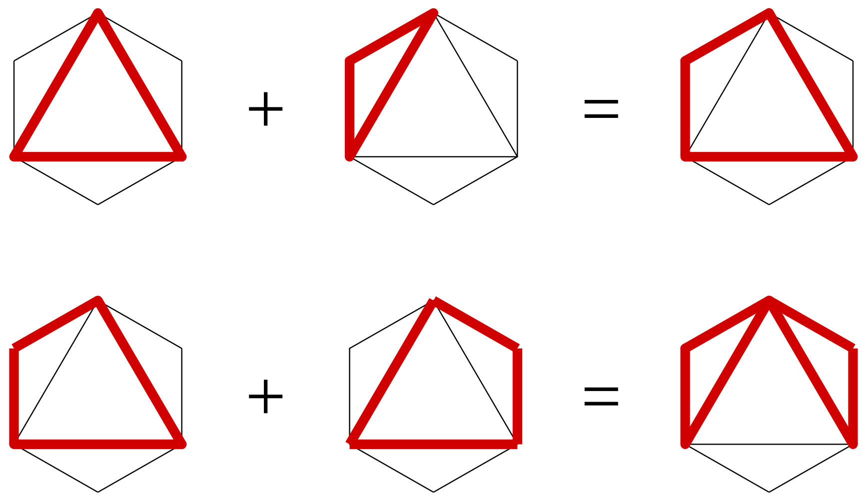


Generalized random graphs ("static model") with $N = 100, 200, 400, 800$, $\langle k \rangle = 2$, $\beta = 0.5$

Leaving Las Vegas ...

Monte Carlo — summing cycles

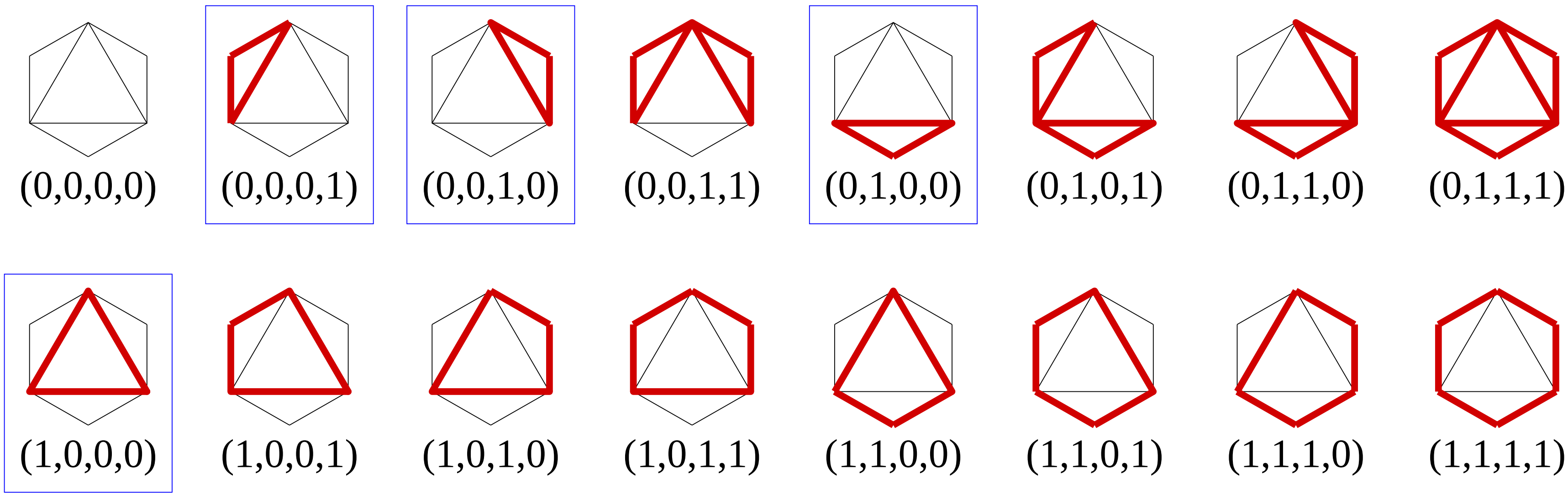
- Sum of two cycles yields new cycle:



- (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!!!

Monte Carlo — algorithm

1. Generate cycle basis of simple cycles $B_1, B_2, \ldots, B_\nu$

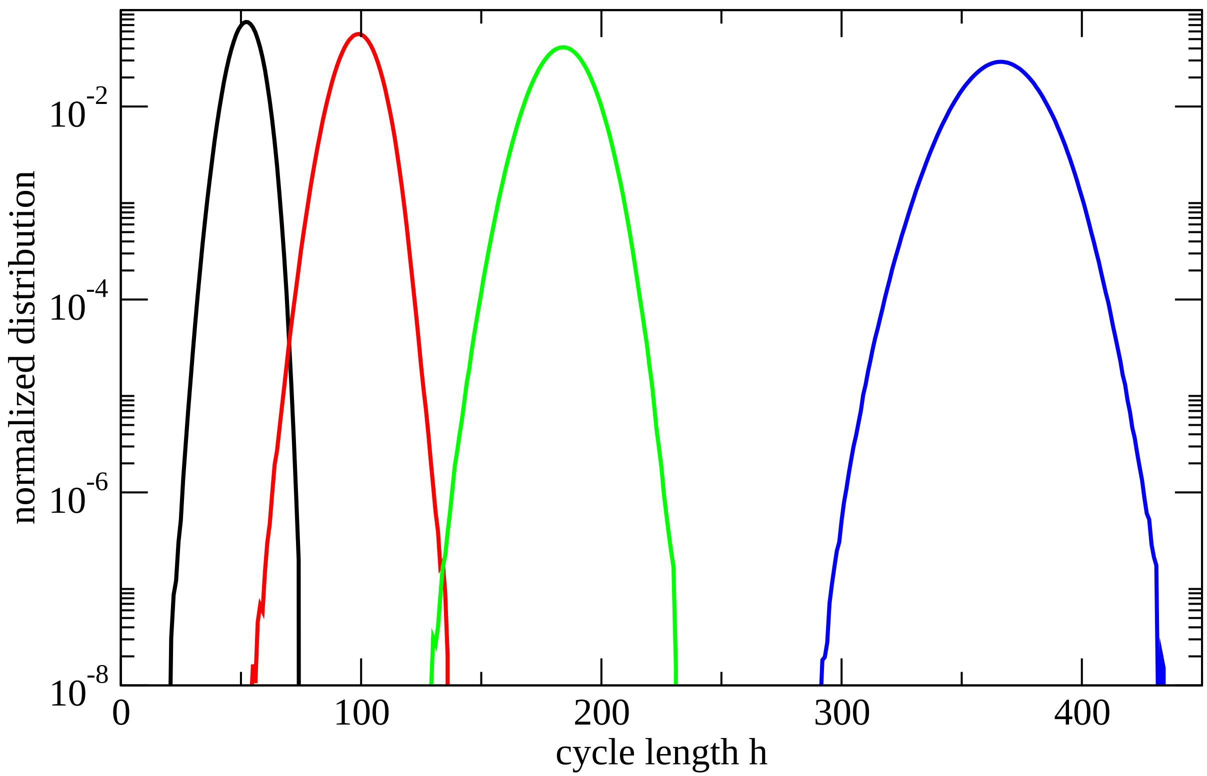
2. Set current cycle $C := 0$ (empty cycle)

3. (Propose) Draw random index $i \in \{1, 2, \ldots, \nu\}$

4. (Accept) If $C + B_i$ is simple or null cycle, set $C := C + B_i$
   (Reject) Otherwise, leave C unchanged

5. Resume at 3. (or stop after desired number of iterations)

Monte Carlo — preliminary results

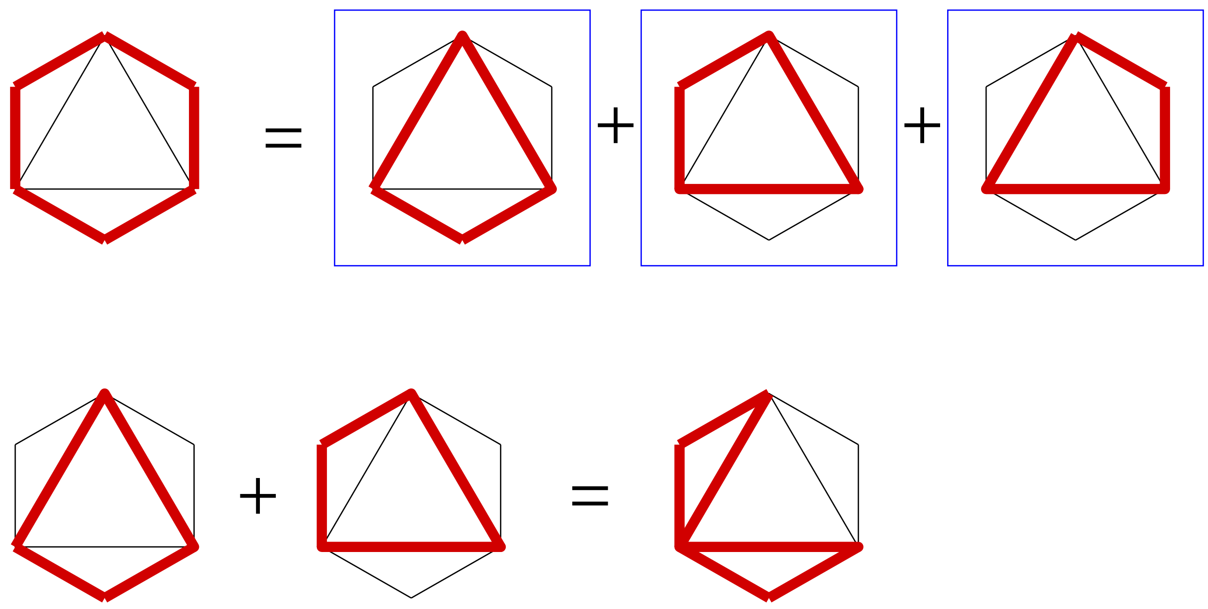


Generalized random graphs (”static model”) with $N = 100, 200, 400, 800$, $\langle k \rangle = 2$, $\beta = 0.5$

Monte Carlo — properties

- detailed balance — ok
- adjacent simple cycles equiprobable
- extensions easy, e.g. Metropolis with Energy := cycle length
- But: ergodicity ?!?

Monte Carlo — trouble



- reachability of cycles depends on choice of cycle basis
- long cycles particularly difficult to reach — less space for "maneuvers"

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

Summary / Outlook

- Naive approaches tend to give bad statistics due to under-sampling of long cycles
- Cycle space method is powerful if ergodicity can be ensured.
- Still lost and hopping through Las Vegas?
- Escape by
  (1) finding robust cycle bases, and/or
  (2) considering move sets beyond cycle bases

Co-starring: Peter F. Stadler