Approximate counting of large subgraphs in random graphs with statistical mechanics methods

Guilhem Semerjian

LPT-ENS
Paris

13.03.08 / Eindhoven

in collaboration with

Rémi Monasson, Enzo Marinari and Valery Van Kerrebroeck

[EPL (06), J. Stat. Mech. (06), PRE (07)]

Outline

Two perspectives on...

... the “Statistical Mechanics (models) on Random Structures”

Emphasis can be put :

- either on the model

  example : spin-glasses on Erdos-Renyi random graphs [Viana, Bray]

- or on the random structure

  model chosen for the “geometric” information
  it provides about the graph

  examples :

  - $q \to 1$ Potts model for percolation [Fortuin, Kasteleyn]
  - $O(n \to 0)$ model for self-avoiding walks [de Gennes]

Graph theoretic problems

Decision problems: existence of a subgraph in a graph

- small (bounded size) subgraph $\Rightarrow$ easy (example: triangles)
- large subgraphs :
  - can be easy (Eulerian circuit, perfect matching)
  - or difficult, Hamiltonian circuit is NP-complete

Enumeration problems: count the number of such subgraphs

generally difficult (#P-complete) for large subgraphs...
...even when the existence is easy to decide

- Eulerian circuits in undirected graphs
- perfect matchings in non-planar graphs

Construction problems: exhibit such a subgraph

Random graphs context

$\mathcal{N}_H(G)$: number of distinct copies of H in G

G drawn from a random ensemble $\Rightarrow \mathcal{N}_H$ is a random variable

Natural questions :

- conditions on the random ensemble to ensure

$$\lim_{N\to\infty} \mathrm{Prob}[\mathcal{N}_H > 0] > 0 \ \ ?$$

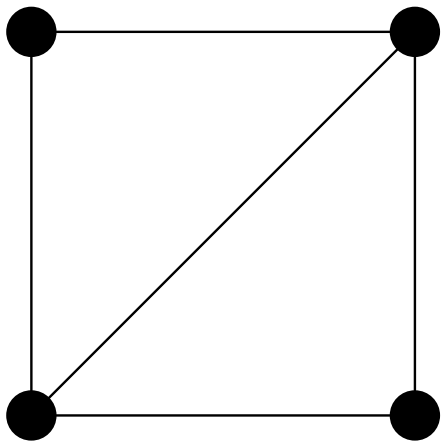
- detailed description of the distribution of $\mathcal{N}_H$?

One example: circuits

Graphs G with N vertices and M edges $\quad M = \theta(N)$

Circuit (loop, cycle) of length L : closed, non-intersecting path which visits L vertices (and L edges)

$\mathcal{N}_L(G)$: number of distinct circuits of length L in G



$$N = 4$$

$$M = 5$$

$$\mathcal{N}_3 = 2$$

$$\mathcal{N}_4 = 1$$

Most interesting regime : $N, L \to \infty$, with $\quad \ell = \lim L/N > 0$

Long circuits in random graphs

Non-exhaustive list of rigrous results :

- random c-regular graphs (for $c \geq 3$)
  - are Hamiltonian with high probability [Robinson and Wormald, Janson]
  - distribution of $\mathcal{N}_{\ell N}$ characterized for all ℓ [Garmo]
- fixed degree distribution, supported on $[3, k_{\max}]$

  conjectured to be Hamiltonian [Wormald]
- Erdos-Renyi random graphs $G(N, p = c/N)$
  - c close to 1 (percolation regime) [Flajolet and Knuth and Pittel]
  - c large (but independent on N) [Frieze]
  - $c = \theta(\log N)$ [Pósa]

Long circuits in random graphs

Difficulty : $\mathcal{N}_L \sim e^{N\sigma}$ $(L/N \to \ell)$

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[\mathcal{N}_L] = \sigma_{\mathrm{a}}(\ell)$$

$$\lim_{N\to\infty} \frac{1}{N} \log \mathbb{E}[\mathcal{N}_L^2] = 2\sigma_{\mathrm{a}}(\ell) + \delta(\ell)$$

$\delta(\ell) > 0$: second moment method inapplicable
(except for random regular graphs, where $\delta(\ell) = 0$)

quenched entropy : $\sigma_{\mathrm{q}}(\ell) = \lim_{N\to\infty} \frac{1}{N} \mathbb{E}[\log \mathcal{N}_L] \quad \leq \sigma_{\mathrm{a}}(\ell)$

$\sigma_{\mathrm{q}}(\ell)$ gives the typical behaviour of $\mathcal{N}_L$,
while $\sigma_{\mathrm{a}}(\ell)$ is dominated by rare events

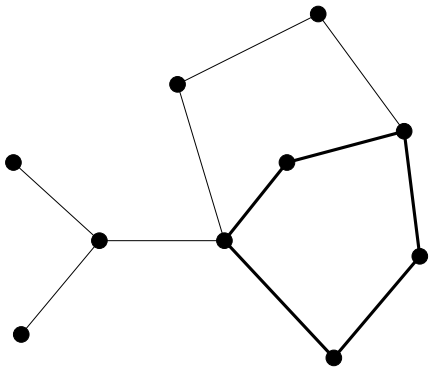
Statistical mechanics model

degrees of freedom : $S_l \in \{0,1\}$ on the M edges of the graph

$\mathcal{C} = \{S_1, \ldots, S_M\}$ = subgraph of G (retain edges with $S_l = 1$)

$S_l = 0$

$S_l = 1$



$$w(\mathcal{C}) = \begin{cases} 0 & \text{if } \mathcal{C} \text{ is not a circuit} \\ u^L & \text{if } \mathcal{C} \text{ is a circuit of length } L \end{cases}$$

$$\mathrm{Prob}[\mathcal{C}] = \frac{w(\mathcal{C}; u, G)}{Z(u, G)} \quad , \qquad \text{with} \quad Z(u, G) = \sum_L \mathcal{N}_L(G)\, u^L$$

Canonical computation (generating function of $\mathcal{N}_L$)

$$Z \sim \int d\ell \; e^{N[\sigma(\ell) + \ell \log u]} \sim e^{N[\sigma(\ell(u)) + \ell(u) \log u]}$$

Statistical mechanics model

does $w(\mathcal{C})$ have a simple (local) description ? yes (almost...)

$$w(\mathcal{C}) = \left(\prod_{l=1}^{M} u^{S_l}\right)\left(\prod_{i=1}^{N} w_i(\{S_l\}_{l\in A(i)})\right)$$

$$w_i = \begin{cases} 1 & \text{if there are 0 or } k=2 \text{ edges with } S_l = 1 \text{ around vertex } i \\ 0 & \text{otherwise} \end{cases}$$

[There can be several vertex disjoint circuits in $\mathcal{C}$, should not be relevant in the thermodynamic limit]

Related statistical mechanics studies :

- $k = 1$ (matchings) [Zhou and Ou-Yang, Zdeborová and Mézard]
- $k \geq 3$ (k-regular subgraphs) [Pretti and Weigt]

The results of the cavity method

Random graphs converge locally to Galton Watson random trees with offspring distribution $\tilde{q}_k$:

- c-regular random graphs, $\tilde{q}_k = \delta_{k,c-1}$
- Erdos-Renyi random graphs, $\tilde{q}_k = e^{-c} c^k / k!$

For each value of u, solve a recursive distributional equation :

$$y \stackrel{\mathrm{d}}{=} \frac{u \sum_{1 \le i \le k} y_i}{1 + u^2 \sum_{1 \le i < j \le k} y_i y_j} \quad , \quad k \sim \tilde{q}_k \quad , \quad y_1, \ldots, y_k \text{ iid copies of } y$$

Then compute $\ell(u) = \frac{c}{2} \mathbb{E}\left[\frac{u y_1 y_2}{1 + u y_1 y_2}\right]$, $\sigma_{\mathrm{q}}(\ell(u)) = \mathbb{E}[f(y_1, \ldots, y_k)]$

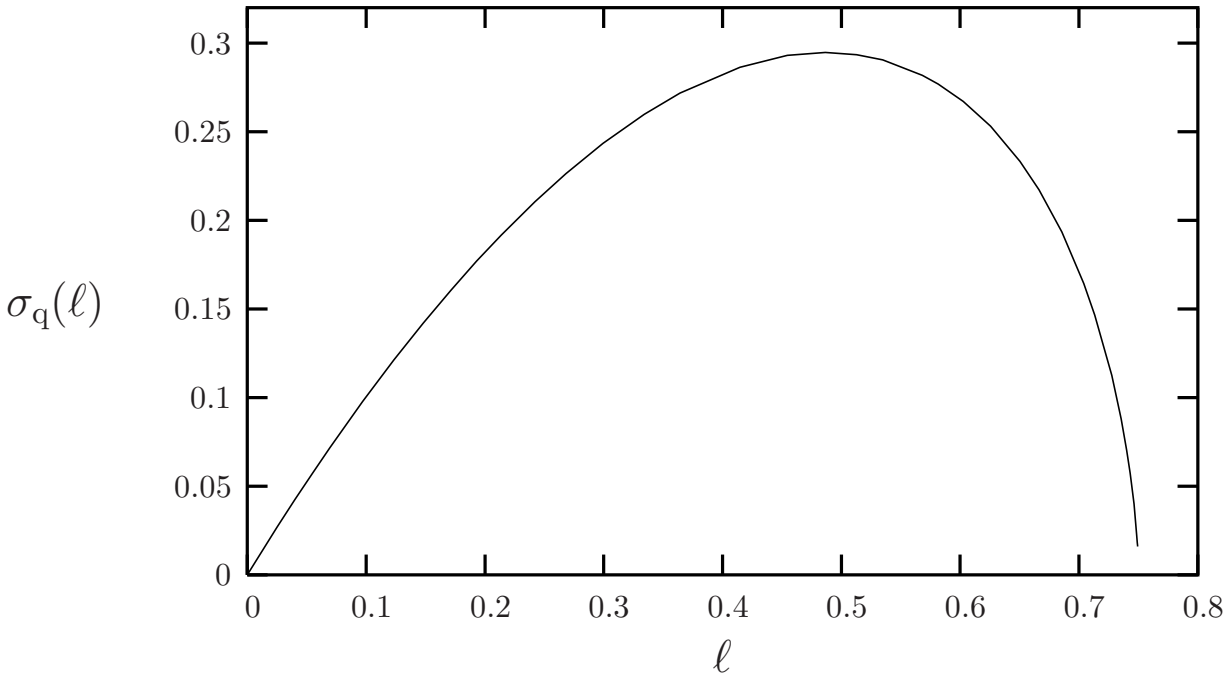
$\Rightarrow$ parametric expression of the quenched entropy $\sigma_{\mathrm{q}}(\ell)$

The results of the cavity method

For random regular graphs the equation can be solved explicitly,
gives back the rigorously known results $(\sigma_{\mathrm{q}}(\ell) = \sigma_{\mathrm{a}}(\ell))$

For generic $\tilde{q}_k$, distributional equation can be solved numerically
(population dynamics algorithm)

Example :
Erdos-Renyi
random graphs
with $c = 3$



The results of the cavity method

Longest circuits ($u \to \infty$) :

- if the minimal degree is larger than 3, $\ell(u \to \infty) \to 1$,
  as previously conjectured [Wormald]

  quantitative refinement of the conjecture:

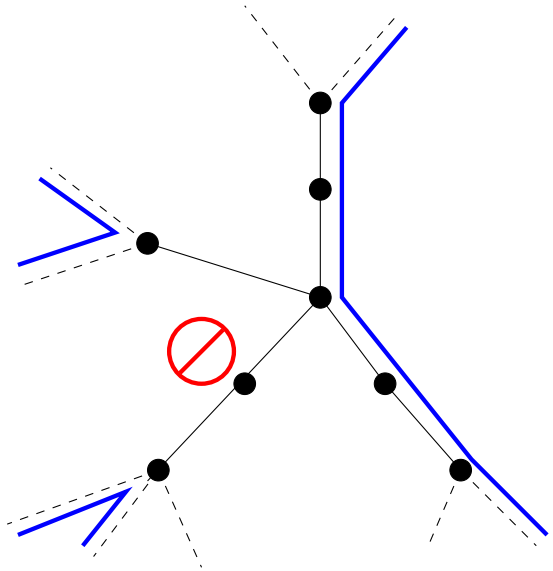
  prediction of the typical number of Hamiltonian cycles

  for instance for graphs with half of the vertices of degree 3,
  half of degree 4

The results of the cavity method

Longest circuits ($u \to \infty$) :

- if the minimal degree is 2, $\ell_{\max} < 1$



Conjecture in the limit of a small fraction of degree 2 sites:

$$\ell_{\max} = 1 - \sum_{k=3}^{\infty} q_k \binom{k}{3} \tilde{q}_1^3 + O(\tilde{q}_1^4)$$

hence an explicit formula for the length of the largest circuits in Erdos-Renyi random graphs, in the large c limit

in agreement with the rigorously known bounds [Frieze]

Conclusions

- Outcomes
  - New conjectures, in particular for the longest circuits
  - Algorithm (Belief Propagation) for counting on a single graph
  - Algorithm for constructing the circuits
    [numerical confirmation of the conjecture on the Hamiltonianicity of graphs with degree 3 and 4]

- Perspectives
  - Replica symmetry breaking
    [longest cycles with a large fraction of degree 2 vertices, correlated extreme value problem]
  - Large deviations from the typical cases
    [probability that a c-regular random graph is not Hamiltonian ?]
  - Towards a rigorous proof
    independent sets [Bandyopadhyay and Gamarnik]
    matchings [Bayati and Nair]