Large-deviation properties of random graphs

Alexander K. Hartmann

Institut für Physik
Universität Oldenburg

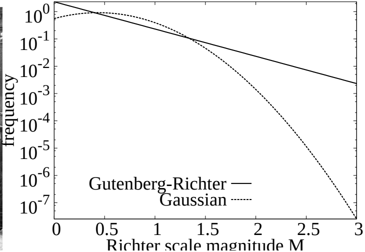
MAPCON 12, 15. May 2012

Rare events

- Storms
- Stock-market crashs
- Floodings $\longrightarrow$
- Earth quakes $\downarrow$
- ...

Oldenburg August 2010

San Fracisco 1906

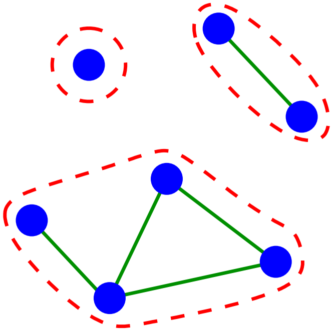


Outline

- Algorithm
- Large-deviation graph properties
  (ER/2d lattice random graphs)
  largest component
  number of components
- Sampling of graphs

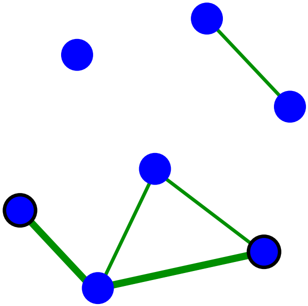
Graphs

- Graph $G = (V, E)$
- connected components: transitive closure of "connectivity relation"



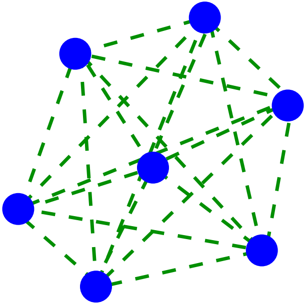
Graphs

- Graph $G = (V, E)$
- connected components: transitive closure of "connectivity relation"
- diameter: longest among all pairwise shortest paths (within components)



Graphs

- Graph $G = (V, E)$
- connected components: transitive closure of "connectivity relation"
- diameter: longest among all pairwise shortest paths (within components)



- Random graphs:

  N vertices, each edge tentative (ij) with prob. p.

  - Erdös-Rényi: $(ij) \in N^{(2)}$, $p = c/N \rightarrow$ finite connect. c
  - two-dim. percolation: $(ij) \in$ square lattice, $p =$ const

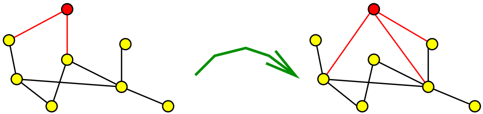
  N vertices, given (sampled) degree sequence $k_1 \ldots, k_i$, e.g., scale-free $P(k) \sim k^{-\gamma}$
  each graph with same probability ("configuration model")

Physics Approach

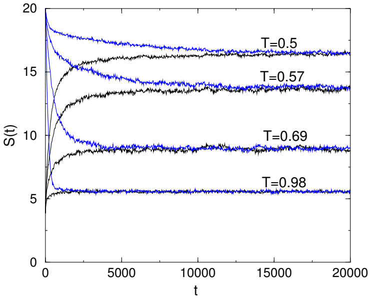
- Idea:

model	$\leftrightarrow$	physical system
quenched realisation	$\leftrightarrow$	degrees of freedom $\vec{x}$ (state)
quantity “score” S	$\leftrightarrow$	energy $E(\vec{x})$

  (ground state: often known)
  simulate at finite T
  Monte Carlo moves:
  change realisat. a bit



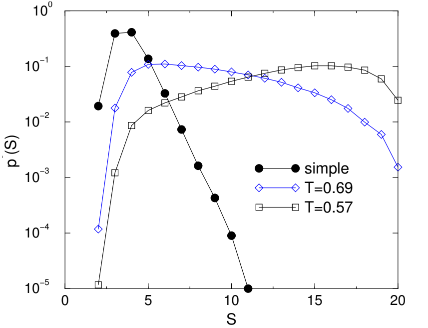
- Simulation at different T
  (using $(\mathrm{MC})^3$/PT)
  Example
  (sequence alignment)
  equilibration:
  start with ground state/
  with random state



- Wang-Landau approach

Distribution of Scores

- Raw result $\longrightarrow$ (simple $\leftrightarrow T = \infty$) at low T: high scores prefered
- MC moves: $\vec{x} \to \vec{x}'$ change on "element" probability $= f_a$



$$\Pr(\text{acceptance}) = \min\{1, \frac{\exp(S(\vec{x}')/T)}{\exp(S(\vec{x})/T)}\} = \min\{1, e^{\Delta S/T}\}$$

- $\Rightarrow$ equilibrium distribution $Q_T(\vec{x}) = P(\vec{x}) e^{S(\vec{x})/T} / Z(T)$
  with $P(\vec{x}) = \prod_i f_{x_i}$, $Z(T) = \sum_{\vec{x}} P(\vec{x}) e^{S(\vec{x})/T}$

$$\Rightarrow \; p_T(S) = \sum_{\vec{x}, S(\vec{x})=S} Q_T(\vec{x}) = \frac{\exp(S/T)}{Z(T)} \sum_{\vec{x}, S(\vec{x})=S} P(\vec{x})$$

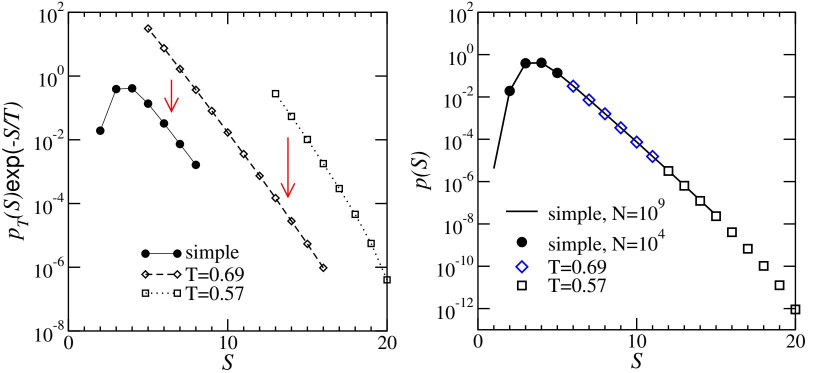
$$\Rightarrow \quad p(S) = p_T(S) Z(T) e^{-S/T}$$

[AKH, PRE 2001]

Match Distriutions

$\left[\, p(S) = p_T(S)Z(T)\exp(-S/T)\,\right]$

rescaling with $\exp(-S/T)$ — $Z(T)$ by "matching"

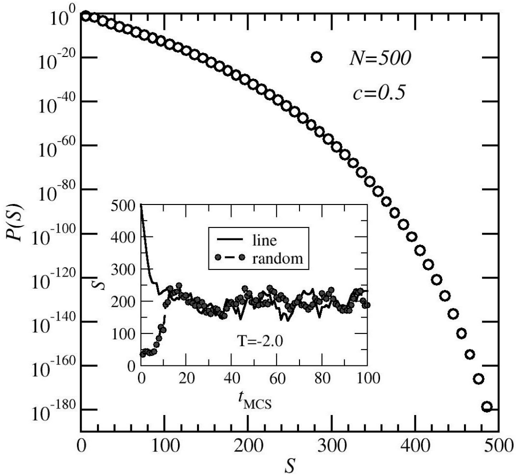


agrees with large statistics simple sampling

agrees with (for this example) known exact result

Results: Erdős-Rényi

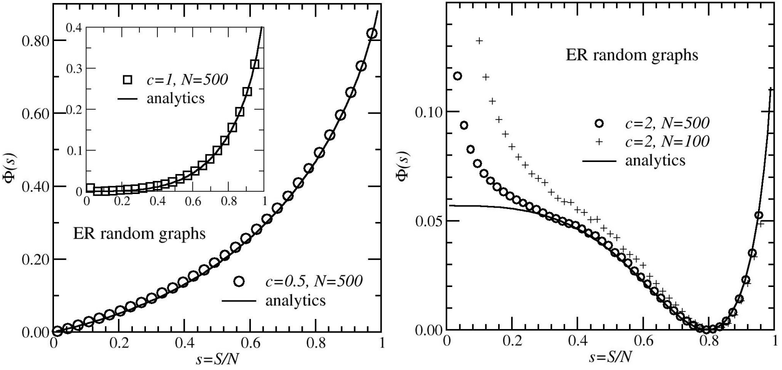
Size S of largest component (connectivity c)



[AKH, Eur. Phys. J. B (2011)]

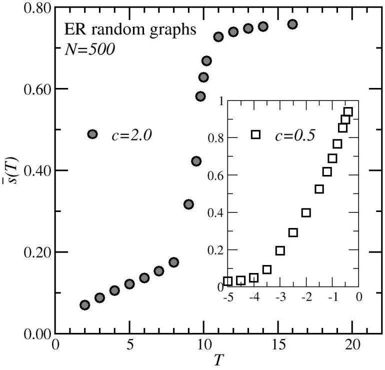
- Rate function $\Phi(s) \equiv -\frac{1}{N} \log P(s)$, $s = S/N$
- Comparison with exact asymptotic result

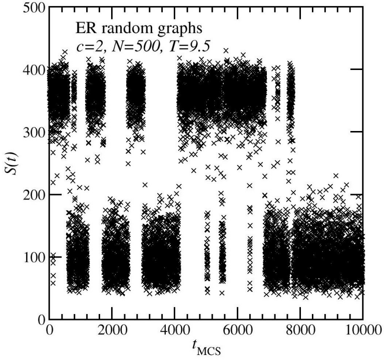
  [M. Biskup, L. Chayes, S.A. Smith, Rand. Struct. Alg. 2007]
- $\rightarrow$ evaluate algorithm $\rightarrow$ works very well
- $\rightarrow$ finite-size corrections visible



Phase transition

- Cluster size as function of (artificial) temperature
- 1st order transition in percolating phase

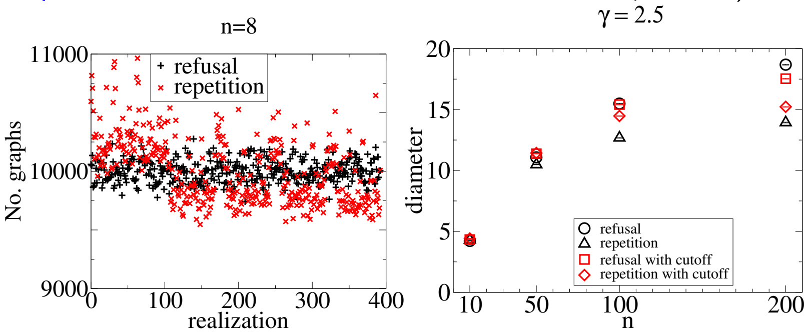




- → large system sizes not fully accessible (→ use Wang-Landau algorithm here)

Bias in Configuration model

- Configuration model: k “stubs” for each node of degree k. Randomly draw pairs of stubs. If multiple/self edge:
  refusal: start graph from scratch
  repetition: redraw pair
- Repetition is biased: relevant for measurements ($N \to \infty$)?

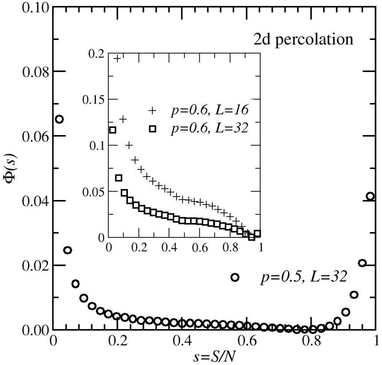


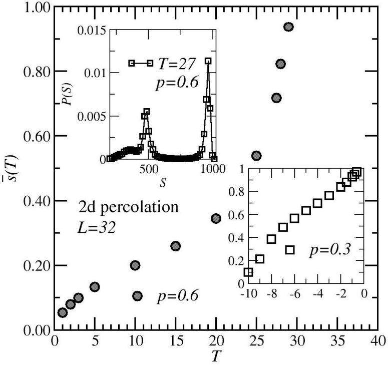
[H. Klein-Hennig, AKH, Phys. Rev. E 2012]

- $\to$ Markov chains/ hidden variables/ throw-away edges/ ...

Two-dimensional percolation

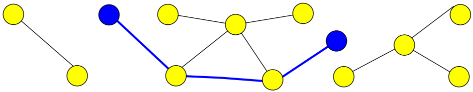
- $N = L \times L$, edge density p
- No exact result known (to me)
- Results comparable to Erdős-Rényi random graphs but stronger finite-size effects





Graph Diameter

- Diameter $d^\star$:= Longest of all shortest $i \to j$ paths



- Random graphs: ($c < 1$): Gumbel distribution

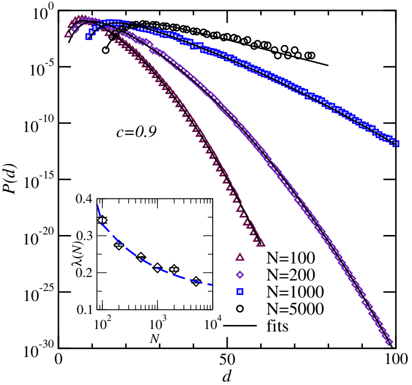
$$Pr_G(d^\star = d) = \lambda e^{-\lambda(d-d_0)} e^{-e^{-\lambda(d-d_0)}}$$

- (sloppy) explanation:
  graph = forest
  $d = \max_{\text{trees } T} d(T)$
  $\to$ Gumbel distribution
- Fit to

$$P(d) = P_G(d) e^{-a(d-d_0)^2}$$

"gaussianized" Gumbel

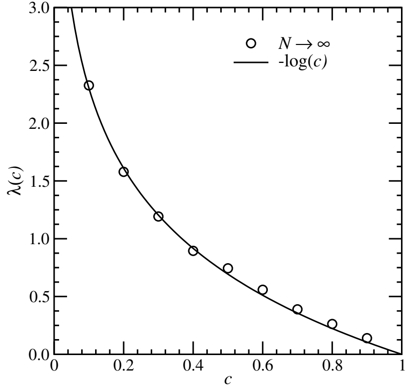
[AKH, M. Mézard, in preparation]



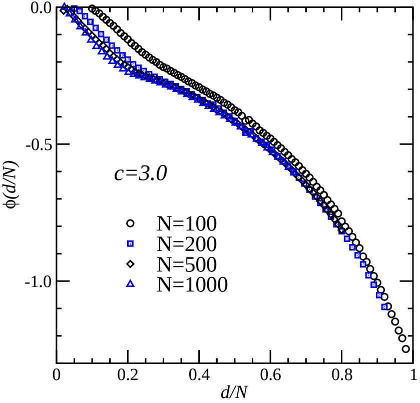
Close to $c = 1$, asymptotically

$$\lambda(c) = -\log c$$

[T. Luczak, Rand.Struct.Alg., 1998]



Percolating region:
more complex distributions



Summary

- Large-deviation properties
- Physics approach:
  study system at artificial finite temperature
  (or, in principle, Wang-Landau algorithm $+$ modifications)
- Full distribution of size of largest component
- Erdős-Rényi random graphs: matches well analytics
  1st order transition in percolating phase
  (also: number of components, 2d percolation, diameter)
- Simple sampling of configuration model is biased

Work more efficiently: read/write/edit scientific paper summaries
`www.papercore.org` (open access)

Summer school: Efficient Algorithms in Computational Physics
Bad Honnef (Germany), 10-14. September 2012