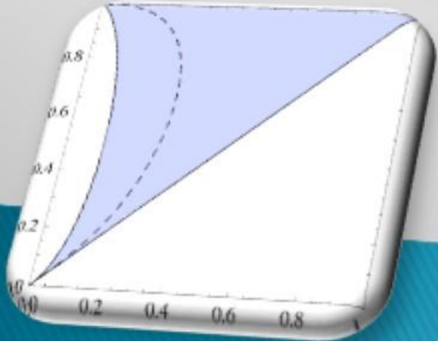
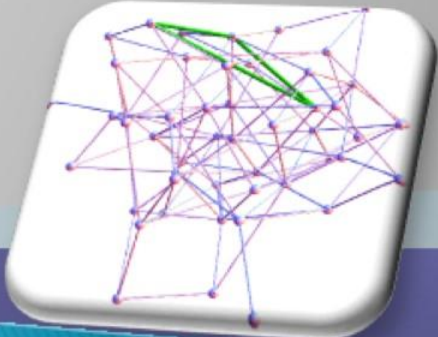
Erdős Lecture Series, Memphis, Mar 2014

Large deviations in random graphs: revisiting the infamous upper tail



Eyal Lubetzky
Microsoft Research



Based on joint works with Yufei Zhao

Large deviations in random graphs

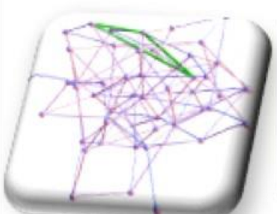
- QUESTION [Chatterjee and Varadhan (2011)]:
  - Fix $0 < p < r < 1$.
  - Take $G \sim \mathcal{G}(n,p)$ conditioned on having at least as many triangles as a typical $\mathcal{G}(n,r)$.
  - Is $G \approx \mathcal{G}(n,r)$, namely, are they close in cut-distance?

- Possibilities: extra triangles due to

replica symmetry

1. *(yes)* overwhelming # edges, uniformly distributed.
2. *(no)* fewer edges, arranged in a special structure.

symmetry breaking

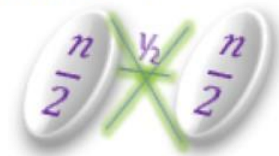


cut distance between G_n and $\mathcal{G}(n,r)$:

$$\delta_{\square}(G_n, r) = \max_{A,B \subset V} \frac{1}{n^2} \big| e(A,B) - r|A||B| \big|$$

Related example: triangle-free $\mathcal{G}(n, p)$

- Let $G \sim \mathcal{G}(n, 1/2)$ conditioned on **triangle-free**:
  [Erdős, Kleitman, Rothschild '76]: w.h.p. G bipartite.



- Let $G \sim \mathcal{G}(n, m)$ conditioned on triangle-free:
  - [Osthus, Prömel, Taraz '03]: bipartite w.h.p. if $\sqrt{\frac{3}{16}\frac{\log n}{n}} \le \frac{m}{\binom{n}{2}} < 1$.
- Sparse $G \sim \mathcal{G}(n, p)$ $(p \ll 1)$:
  - $p \asymp 1/n$: Poisson approximation [Bollobás '81].
  - $p \ge \frac{\log n}{n}$: how can we make G triangle free?
    i. force a given bipartition: $(1-p)^{n^2/4} \approx e^{-n^2 p/4}$
    ii. every triplet must miss an edge: $\left(1-p^3\right)^{n^3/6} \approx e^{-\frac{1}{6}n^3 p^3}$
    Is a typical triangle-free bipartite for $p \gtrsim 1/\sqrt{n}$?

Chatterjee–Varadhan breakthrough

- Powerful new framework for large deviation principle in $\mathcal{G}(n,p)$ using Szemerédi's regularity & graph limits.
- In particular, for upper tails of triangles:
  - Q: Let $G \sim \mathcal{G}(n,p)$ conditioned on $\geq \frac{1}{6}n^3r^3$ triangles for $0 < p < r < 1$. Is $G \approx \mathcal{G}(n,r)$, namely, is $\delta_\square(G,r)$ small?
  - A: depends on (p,r)…
  - Bounds for symmetry replica region.
  - $p \geq \frac{2}{2+e^{3/2}} \approx 0.31$: always symmetric.
  - Multiple phase transitions for small p.
  - e.g., $p = 1/4$ and $r = 1/2$?

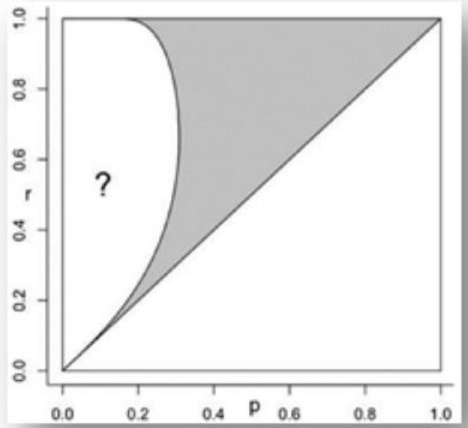


figure from [Chatterjee–Dey '10]

The rate function

- Intimately related (and preliminary) question:
  - Fix $0 < p < r < 1$, let $G \sim \mathcal{G}(n,p)$ and $X = \#$ triangles in G. What is $\mathbb{P}\left(X \geq \frac{1}{6}n^3r^3\right)$?
  - *e.g.*, for $p = \frac{1}{4}$ and $r = \frac{1}{2}$ is it $\exp\left[-\left(\frac{1}{4}\log 2 + \frac{1}{4}\log\frac{2}{3} + o(1)\right)n^2\right]$?
- Replica symmetric candidate:
  - $G \sim \mathcal{G}(n,M)$ for $M \sim \mathrm{Bin}\left(\binom{n}{2}, p\right)$.
  - $\frac{1}{6}n^3r^3$ triangles with probability $\geq e^{-(1-o(1))I_p(r)}$ for

$$I_p(r) = r\log\frac{r}{p} + (1-r)\log\frac{1-r}{1-p}$$
$$= -\lim_{N\to\infty}\frac{1}{N}\log\mathbb{P}(\mathrm{Bin}(N,p) \geq N\,r)\,.$$

- When does $-\frac{1}{\binom{n}{2}}\log\mathbb{P}\left(X \geq \frac{1}{6}n^3r^3\right) \xrightarrow[n\to\infty]{} I_p(r)$?

Sparse random graphs

- Rate function in the ***sparse*** regime? e.g.,
  - Let $G \sim \mathcal{G}(n,p)$ for $p \ll 1$.
  - Let $X = \#$ triangles in G and write
    $$\mathbb{P}(X \geq 2\,\mathbb{E}X) = \exp[-R(n,p)]\,.$$
  - What is $R(n,p)$? [Generally, $R(n,p,\delta)$ for $X \geq (1+\delta)\,\mathbb{E}X$].
- For intuition, consider lower tails:
  - we saw: $\mathbb{P}(X=0) \geq e^{-c\min\{n^2p,\, n^3p^3\}}$
  - [Janson (1990)]'s Poisson large deviation inequality:
    $$\mathbb{P}(X < (1-\delta)\mathbb{E}X) \leq e^{-c_\delta \frac{(\mathbb{E}X)^2}{\Delta + \mathbb{E}X}}$$
    matching upper bound!
  - $\Rightarrow$ $R(n,p,\delta) \asymp \min\{n^2p,\ n^3p^3\}$ (transition at $p \asymp 1/\sqrt{n}$.)
- Similar treatment for upper tail?

The infamous upper tail

- No known generic tool to tackle the upper tail…

  $\mathbb{P}(X \geq (1+\delta)\,\mathbb{E}X) = \exp[-R(n,p,\delta)]$

  The infamous upper tail

  S Janson, A Ruciński - Random Structures & Algorithms, 2002 - Wiley Online Library

  - [Janson, Oleszkiewicz, Rucinski '04], [Bollobás '81, '85], [Janson Luczak, Rucinski '00], [Janson, Rucinski '02, '04a, '04b], [Vu '01],[Kim, Vu '04], [Chatterjee-Dey '10], … , *via* Hoeffding-Azuma ineq./ Talagrand ineq./ Stein's method/…
  - For $\frac{\log n}{n} \leq p \ll 1$: $n^2p^2 \lesssim R(n,p,\delta) \lesssim n^2p^2\log(1/p)$
  - **Order** of $R(n,p,\delta)$ finally resolved in [Chatterjee '11] and [DeMarco, Kahn '11], independently showing

$$R(n,p,\delta) \asymp n^2p^2\log(1/p)$$

- Leading order asymptotics unknown for any δ…

The Chatterjee–Varadhan solution

- The LDP is reduced to a variational problem on *graphons* $f\colon [0,1]^2 \to [0,1]$ (symmetric measurable):
  - Set:
    - $I_p(f) = \int_{[0,1]^2} I_p\big(f(x,y)\big)dxdy.$
    - Subgraph count (H with $V(H) = [m]$) in f:

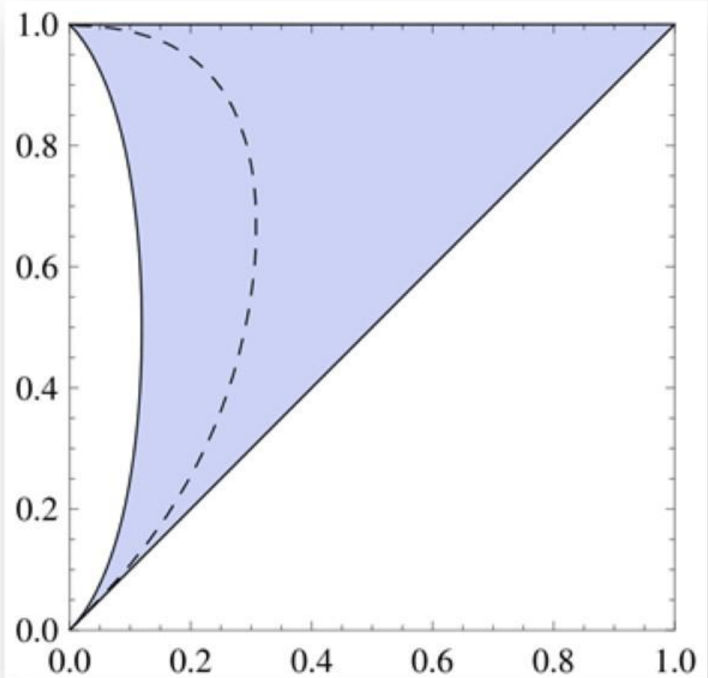
      $$t(H,f) = \int_{[0,1]^m} \prod_{ij\in E(H)} f\big(x_i,x_j\big)dx_1 \cdots dx_m$$
    - Variational problem for upper tails:

      $$\phi(p,r) = \inf\big\{ I_p(f) : t(H,f) \geq r \big\}.$$
- [CV11]: solution gives the rate function; moreover, w.h.p. $(\,\mathcal{G}(n,p) \mid t(H,\cdot) \geq r\,)$ close (in $\delta_\square$) to minimizer.

Phase diagram for triangles

- [CD10, CV11]:
  - $\phi(p,r)$ minimized by $f \equiv r$ (symmetry) if $\left(r^3, I_{p(r)}\right)$ lies on convex-minorant of $x \mapsto I_p\left(x^{1/3}\right)$ (the natural guess).
  - Full phase diagram? One or more phase transitions?
- <u>THEOREM</u>: ([L., Zhao '14+])

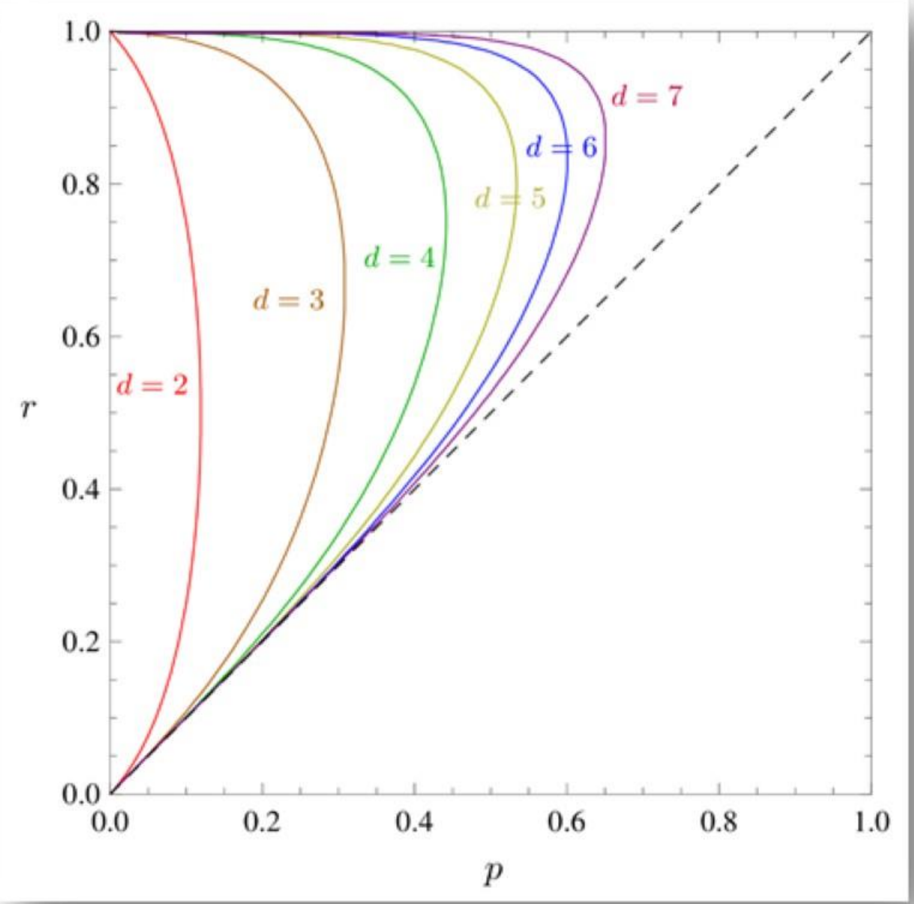
Symmetry replica for upper tails of triangles occurs iff
$p < \left[1 + (r^{-1} - 1)^{1/(1-2r)}\right]^{-1}$



  - Coincides with the convex minorant of $x \mapsto I_p\left(x^{1/2}\right)$

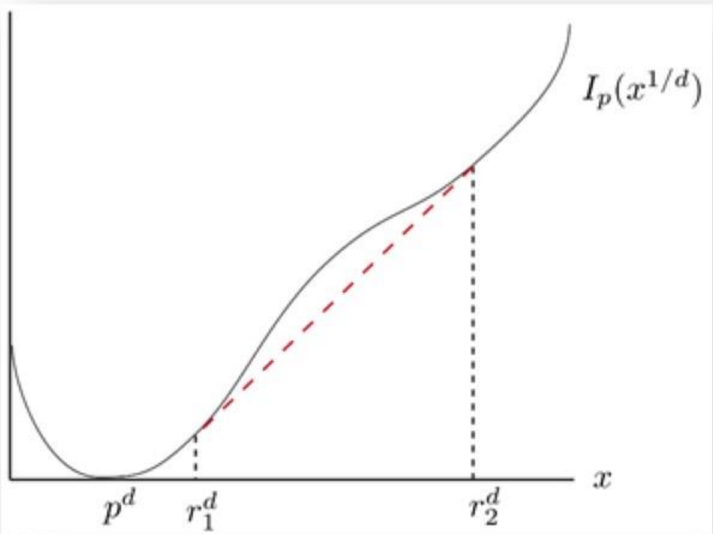
Phase diagram for regular graphs

- More generally:
  - Fix $0 < p < r < 1$ and a d-regular graph H.
  - Minimizer is $f \equiv r$ $\Leftrightarrow (r^d, I_p(r))$ lies on the convex-minorant of $x \mapsto I_p\left(x^{1/d}\right)$.



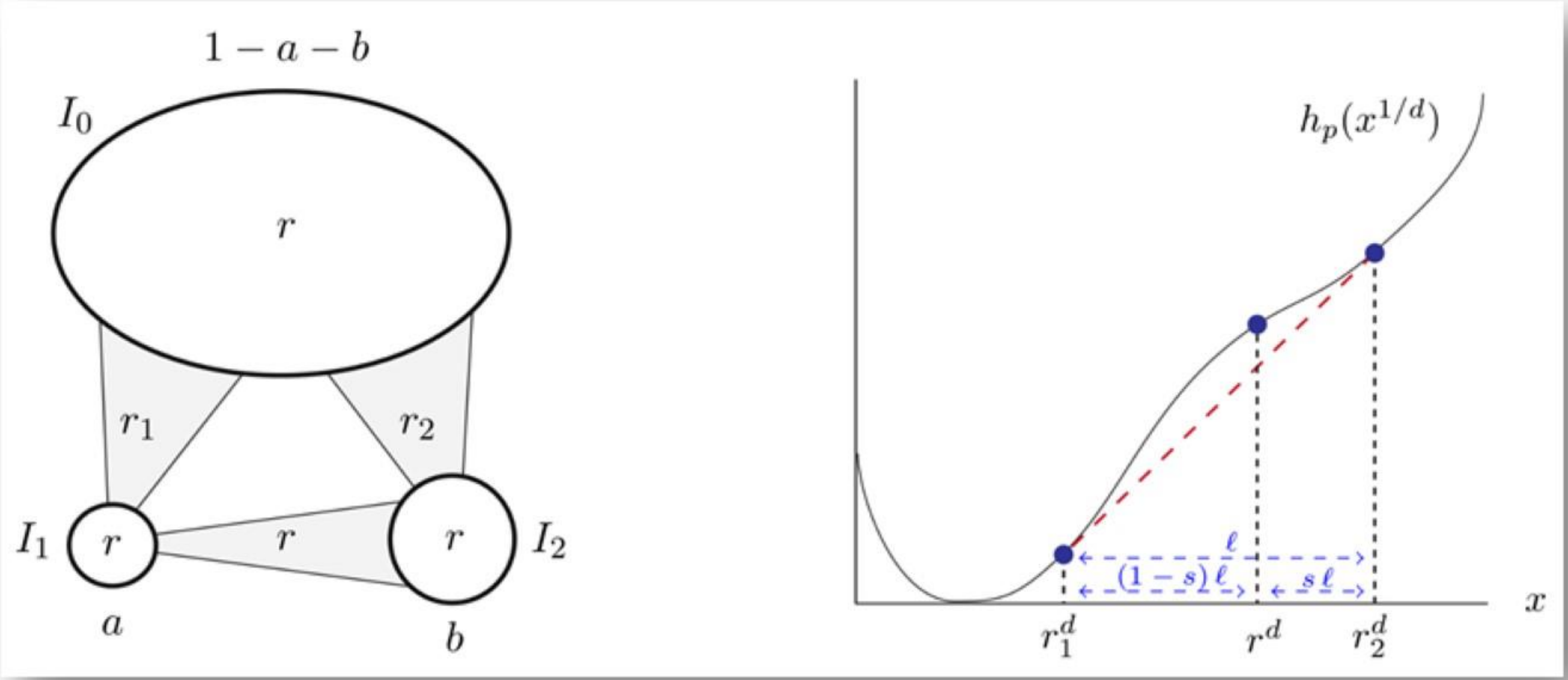
Key to upper bound

- Where does the convex-minorant enter?
  - Let $\psi(x) = I_p(x^{1/d})$ and $\hat{\psi}$ be its convex-minorant.
  - Then by Jensen:
    $I_p(f) = \int \psi(f^d) dxdy \geq \int \hat{\psi}\,(f^d) dxdy \geq \hat{\psi}(\int f^d)$.
  - So, if $\int f^d \geq r^d$ and $\psi(r^d) = \hat{\psi}(r^d)$ then
    $I_p(f) \geq \psi(r^d) = I_p(r)$
- Want $\int f^d \geq r^d$ given LD:
  - [CD10], [CV11]: Hölder; yet one can exploit subgraph structure…
  - Generalized Hölder [Finner '92] gives the correct phase.



A matching lower bound

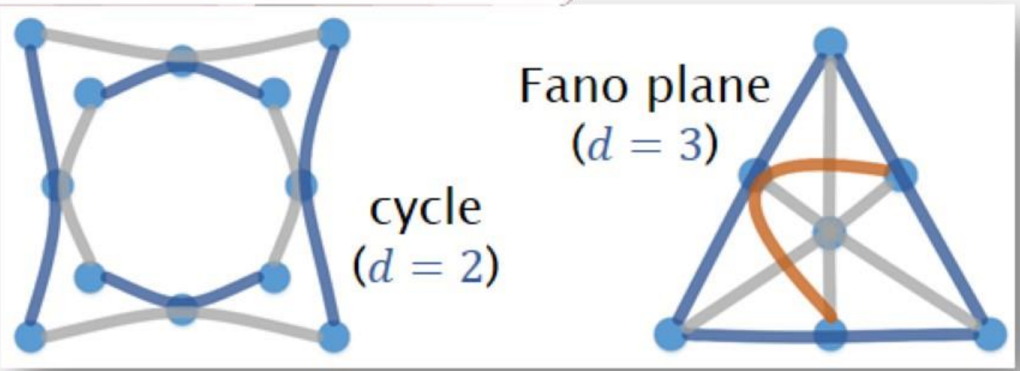
- Tri-partite construction:



  - Choice of $a = s\varepsilon^2$ and $b = (1-s)\varepsilon^2 + \varepsilon^3$ for small enough ε beats the constant function $f \equiv r$.

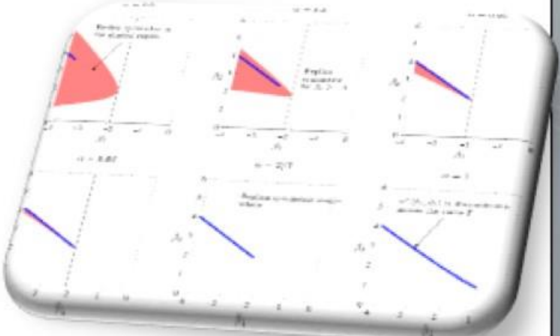
Analogs of phase-diagram result

- d-regular linear hypergraphs:



- Leading eigenvalue:
  $G \sim \mathcal{G}(n,p)$ conditioned on $\lambda_1(G) \geq nr$.
  - phase diagram coincides with $d = 2$.

- Exponential random graphs
  $\mathbb{P}(G) \propto e^{\binom{n}{2}(\beta_1 t(K_2,G)+\beta_2 t(K_3,G))}$
  (building on [Chatterjee-Diaconis'13]



Revisiting the sparse regime

$$\mathbb{P}(X \geq (1+\delta)\,\mathbb{E}X) = \exp[-R(n,p,\delta)]$$

- [Chatterjee, Dembo '14]: recent new breakthrough: for $p \gg n^{-\alpha}$ one has $R(n,p,\delta) \sim \phi(n,p,\delta)$ where

$$\phi(n,p,\delta) = \inf\{\, I_p(G) : t(K_3,G) \geq (1+\delta)p^3 \,\}$$

  over $G \in \mathfrak{G}_n$, *weighted undirected graphs* on n vertices.
- Plausibly: extends throughout $\frac{\log n}{n} \ll p \ll 1$.
- (for $p \geq (\log n)^{-c}$: follows from weak regularity.)
- Opens the door to first asymptotic LDP results for the sparse random graph…

New results in the sparse regime

- THEOREM: ([L., Zhao '15+])

Fix $\delta > 0$. *If* $n^{-1/2} \ll p \ll 1$ *then*

$$\lim_{n\to\infty} \frac{\phi(n,p,\delta)}{n^2p^2\log(1/p)} = \min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\}$$

whereas if $n^{-1} \ll p \ll n^{-1/2}$ *then*

$$\lim_{n\to\infty} \frac{\phi(n,p,\delta)}{n^2p^2\log(1/p)} = \frac{\delta^{2/3}}{2}.$$

- COROLLARY:

For any $\delta > 0$, *if* $n^{-1/42}\log n \le p \ll 1$ *then*

$$\mathbb{P}\big(X \ge (1+\delta)p^3\big) = e^{-(1-o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} n^2p^2\log\left(\frac{1}{p}\right)}$$

Ideas from the proofs

- For the lower bound on

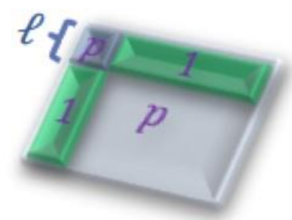
$$\mathbb{P}\big(X \geq (1+\delta)p^3\big) = e^{-(1-o(1))\min\left\{\frac{\delta^{2/3}}{2},\frac{\delta}{3}\right\} n^2p^2\log\left(\frac{1}{p}\right)}$$

  - Take an arbitrary set on $k = \delta^{1/3}np$ vertices and force it to be a *clique* :

$$p^{\binom{k}{2}} = p^{\left(\delta^{2/3}/2+o(1)\right)n^2p^2}$$

  - Or, a set of $\ell = \frac{1}{3}\delta np^2$ vertices and force it to be connected to all other vertices:

$$p^{\ell(n-\ell)} = p^{(\delta/3+o(1))n^2p^2}$$



  - Latter is preferable iff $\delta < 27/8$.
- For the upper bound: reduce to a continuous variational problem; divide and conquer…

Extension to cliques

- THEOREM: ([L., Zhao '15+])

Fix $\delta > 0$ *and* $k \geq 3$. *If* $n^{-1/(k-1)} \ll p \ll 1$ *then*

$$\lim_{n\to\infty} \frac{\phi_{K_k}(n,p,\delta)}{n^2 p^{k-1}\log(1/p)} = \min\left\{\frac{\delta^{2/k}}{2}, \frac{\delta}{k}\right\}$$

whereas if $n^{-2/(k-1)} \ll p \ll n^{-1/(k-1)}$ *then*

$$\lim_{n\to\infty} \frac{\phi_{K_k}(n,p,\delta)}{n^2 p^{k-1}\log(1/p)} = \frac{\delta^{2/k}}{2}.$$

- COROLLARY:

$\forall k \geq 3\ \exists \alpha_k > 0$: *For any* $\delta > 0$, *if* $n^{-\alpha_k} \leq p \ll 1$ *then*

$$\mathbb{P}\left(X_k \geq (1+\delta)p^{\binom{k}{2}}\right) = e^{-(1-o(1))\min\left\{\frac{\delta^{2/k}}{2}, \frac{\delta}{k}\right\} n^2 p^{k-1}\log\left(\frac{1}{p}\right)}$$

Some upper tail open problems

- Dense regime:
  - ? Phase diagram for general (non-regular) graphs.
  - ? What is the solution in a single point within the symmetry breaking regime?
    (at no such pt. can we calculate the rate function…)
  - ? Uniqueness symmetry-breaking solution?
  - ? Are the symmetry-breaking solutions bipartite? or ∃ countable # phase transitions (# parts)?
- Sparse regime:
  - ? Push the [Chatterjee-Dembo'14] result to $p \geq \frac{\log n}{n}$.

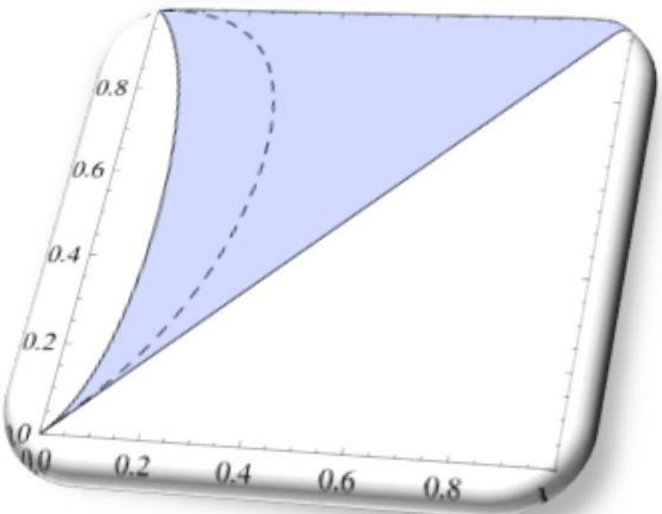
Some lower tail open problems

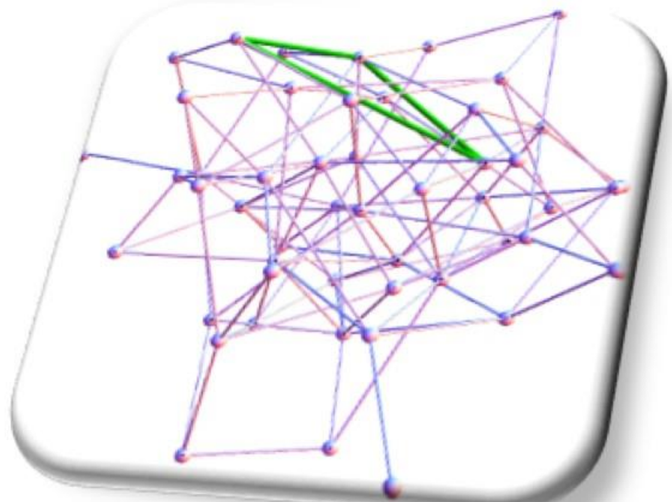
- For some graphs, lower tail is simple
  - *e.g.*: if H is bipartite & Sidorenko's conjecture holds for it, then there is *symmetry everywhere in the dense regime*.
- Triangles: ***inglorious lower tail***: very little is known:
  - Dense regime: phase diagram for symmetry replica?
  - Sparse regime: solution of the variational problem?

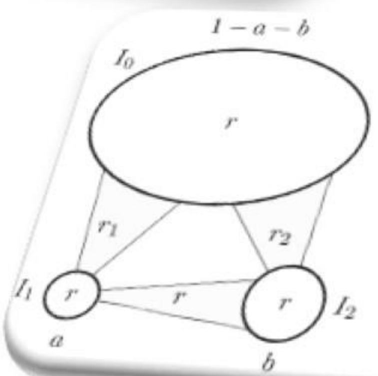
In addition to:

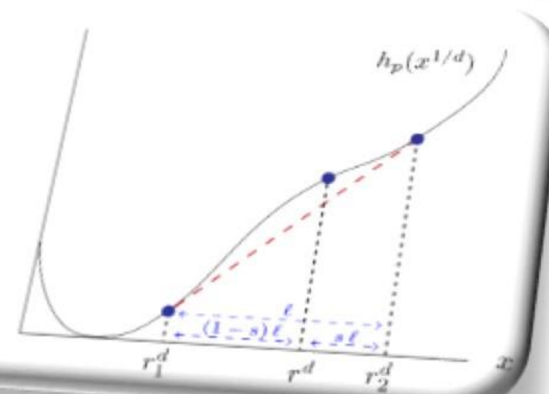
- all of the above open problems for the upper tail…

Thank you


0.8
0.6
0.4
0.2
0
0.0
0.2
0.4
0.6
0.8




$1-a-b$
I_0
r
r_1
r_2
I_1
r
r
r
I_2
a
b


$h_p(x^{1/d})$
ℓ
$(1-s)\ell$
$s\ell$
r_1^d
r^d
r_2^d
x