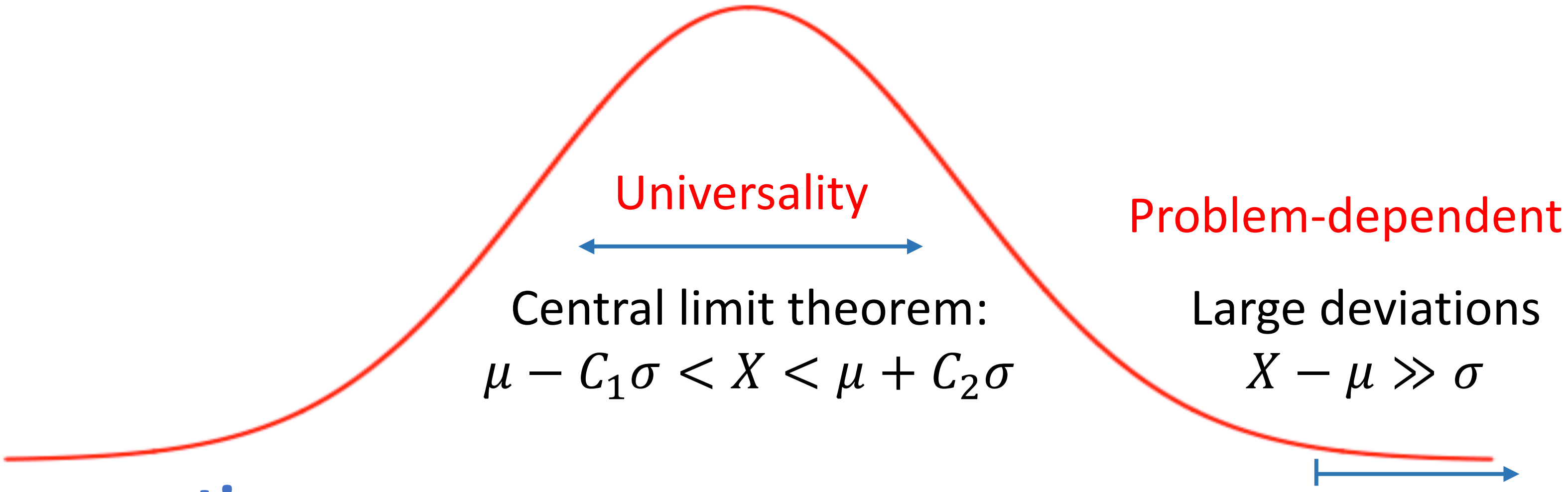
Large Deviations and Exponential Random Graphs

Yufei Zhao

MIT

May 2018



Key questions:

- What is the probability of seeing large deviation? (often exponentially small)
- What does a typical conditioned instance look like?
- How to model/estimate/sample?

Warm up: sum of independent random variables

Let $X = Y_1 + Y_2 + \cdots + Y_N$
Y_i's are i.i.d. random variables with finite variance

- Central Limit Theorem: $\frac{X - \mathbb{E}X}{\mathrm{stdev}\ X} \to \text{Normal as } N \to \infty$
- Large deviation theory (Cramér's theorem):
$$\mathbb{P}(X \geq Nx) \approx e^{-NI(x)}$$
  where $I(x)$ is the *rate function*, which depends on the distribution of the Y_i's
  e.g., if $Y_i \sim \mathrm{Bernoulli}(p)$, then $I(x) = x \log \frac{x}{p} + (1 - x) \log \frac{1-x}{1-p}$

Sums of dependent random variables

E.g., $X = f(Y_1, Y_2, \ldots, Y_N)$
$Y_1, Y_2 \ldots$ i.i.d. Bernoulli random variables
f – a low degree polynomial

- Moments calculation: $\mathbb{E}[X^k]$ often easy to compute
- Central limit theorem: follows with enough control on moments
- Large deviations: ???

The upper tail problem

Let X be the number of triangles in the Erdős–Rényi random graph $\mathbf{G}(n,p)$ (n vertices, every pair is an edge with probability p independently)

$$\mathbb{E}X = \binom{n}{3} p^3$$

Central Limit Theorem (Ruciński '88): X is asymptotially normal, i.e.,

$$\frac{X - \mathbb{E}X}{\sqrt{\mathrm{Var}\, X}} \to \text{Normal}, \qquad \text{as } n \to \infty, \text{provided } np \to \infty, n(1-p) \to \infty$$

Problem: Estimate $\mathbb{P}(X \geq (1+\delta)\mathbb{E}X)$ (fixed $\delta > 0$)

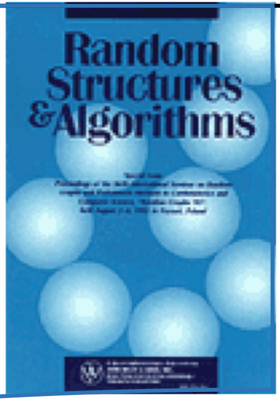
X = # triangles in $\mathbf{G}(n,p)$. $\mathbb{P}(X \geq (1+\delta)\mathbb{E}X) = ?$

[redacted]

The Infamous Upper Tail

Random Structures & Algorithms 2002

Svante Janson, Andrzej Ruciński



Janson, Oleszkiewicz, Rucinski ’04
Bollobás ’81, ’85
Janson, Luczak, Rucinski ’02, ’04
Vu ’01
Kim & Vu ’04
Chatterjee & Dey ’10

Order of

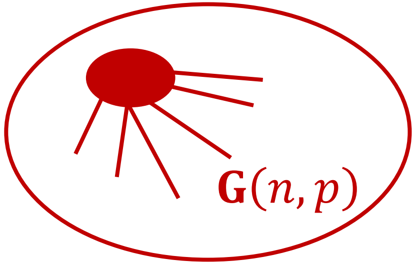
$$\log \mathbb{P}(X \geq (1+\delta)\mathbb{E}X)$$

independently determined by

DeMarco & Kahn ’11
and
Chatterjee ’11

What can “cause” a random graph to have too many triangles?

- Overall increase in edge density
- Some extra edges forming a clique
- Some some number of vertices forming a hub connecting to everything else
- …



replica symmetry

symmetry breaking

Summary of what we now know/believe

X = # triangles in $\mathbf{G}(n, p)$

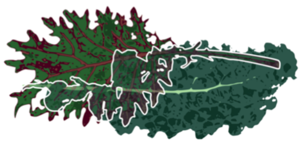
Large deviation: $X \geq (1 + \delta)\mathbb{E}X$ (constant δ)

- **Sparse setting:** $p \to 0$ (not too quickly) as $n \to \infty$
  - If $\delta > 27/8$, plant a clique
  - If $\delta < 27/8$, plant a hub
- **Dense setting:** constant p
  - Some range of δ: replica symmetry (uniform density boost)
  - Outside of this range: symmetry breaking (precise structure unknown)

How to compute large deviations

1. Prove a **large deviation principle (LDP)** that reduces the problem to a variational problem (maximization/minimization problem modeling the “most likely cause”)
2. Solve this variational problem

Review of large deviations



Fixed $0 < p < q < 1$.

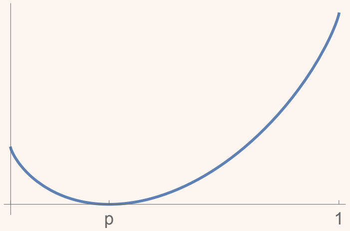
$X \sim \text{Binomial}(n, p)$. $\quad\quad$ $\mathbf{P}(X \geq nq) = ??$

$$\boxed{\log \mathbb{P}(X \geq nq) = -(I_p(q) + o(1))n \qquad \text{as } n \to \infty}$$

"cost of tilting"

Relative entropy (KL divergence):

$$I_p(x) := x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$$



Triangles in G(n,p)

For each pair (i, j) of vertices

- Tilt its probability to some $q_{ij} \geq p$
- Pay $I_p(q_{ij})$ cost in log probability.

Objective: minimize relative entropy cost $\min \sum_{1 \le i < j \le n} I_p(q_{ij})$

Constraint: enough triangles $\sum_{1 \le i < j < k \le n} q_{ij} q_{ik} q_{jk} \geq \binom{n}{3} q^3$

This actually works! The minimum is asymptotically $-\log \mathbb{P}(X \geq \binom{n}{3} q^3)$

Chatterjee—Varadhan ’11	dense setting: p constant
Chatterjee—Dembo ’16	sparse setting: $p \geq n^{-1/42} \log n$
Eldan ’17+	improved: $p \geq n^{-1/18} \log n$

Another interpretation

By Gibbs variational principle, a conditional probability distribution is given by the entropy-maximizing probability distribution subject to the conditions.

Large deviation principle (whenever it holds): For random graphs, we can approximate this distribution by an entropy-maximizing product measure (independent edges)

Graphon variational problem

$W(x,y) = W(y,x)$

- A graphon is a symmetric measurable function $W: [0,1]^2 \to [0,1]$.

Discrete variational problem	Graphon variational problem [Chatterjee—Varadhan]
Minimize $\sum_{1\le i<j\le n} I_p(q_{ij})$ Subject to $\displaystyle\sum_{1\le i<j<k\le n} q_{ij}q_{ik}q_{jk} \ge \binom{n}{3} q^3$	Minimize $\int_{[0,1]^2} I_p(W(x,y))\,dxdy$ Subject to $\displaystyle\int_{[0,1]^3} W(x,y)W(x,z)W(y,z)\,dxdydz \ge q^3$

- Due to compactness of the space of graphons under cut metric (Lovasz—Szegedy), the above minimum is always attained
- In general we do NOT know how to solve the variational problem

What do the minimizing graphons represent?

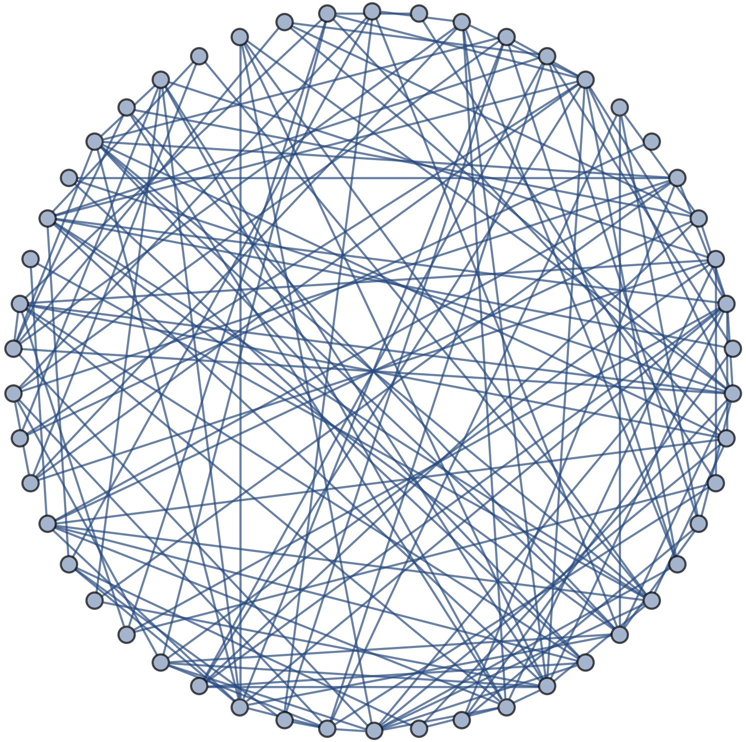
The set of relative entropy minimizing graphons represents the most likely graphs conditioned on the rare event.

Replica symmetry: If minimized (uniquely) by the constant graphon, then the conditioned random graph is close to Erdős–Rényi (in cut distance).

Sparse setting

G(*n*,*p*)

$p = p_n \to 0$ as $n \to \infty$, perhaps slowly



Order of the rate

Theorem (DeMarco—Kahn '11, Chatterjee '11).
Let X denote the number of triangles in $\mathbf{G}(n,p)$.
Fix $\delta > 0$. For $p \gtrsim (\log n)/n$,

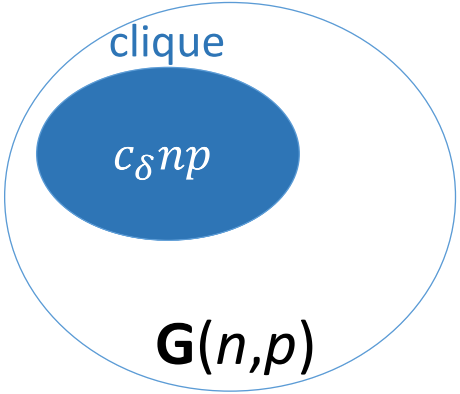
$$\mathbb{P}(X \geq (1+\delta)\mathbb{E}X) = p^{\Theta_\delta(n^2p^2)}$$

Proof of lower bound:

Force a clique on $m = \Theta_\delta(np)$ vertices

Obtain $\binom{m}{3} \geq (1+\delta)\binom{n}{3}p^3$ triangles

Occurs with probability $p^{\binom{m}{2}} = p^{\Theta_\delta(n^2p^2)}$

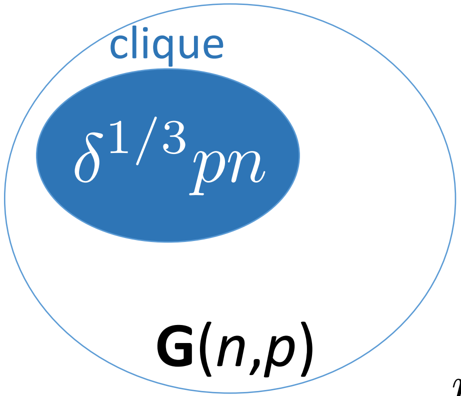


Theorem (Chatterjee—Dembo/Eldan + Lubetzky—Z.).
Let X denote the number of triangles in $\mathbf{G}(n,p)$.
Fix $\delta > 0$. With $p \to 0$ and and $p \geq n^{\boxed{-1/18}} \log n$,

Improve this!

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

Proof of lower bound:



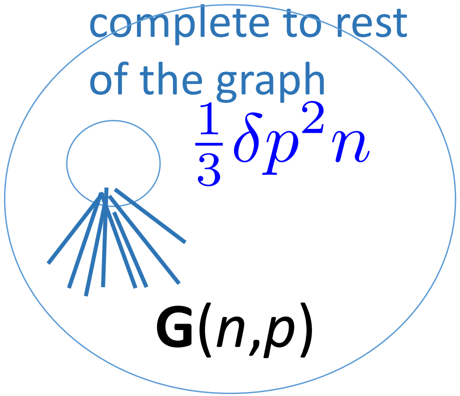
$\sim \delta p^3 \binom{n}{3}$

extra triangles

With probability:

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

Preferred for δ > 27/8



$\sim \delta p^3 \binom{n}{3}$

extra triangles

With probability:

$p^{(1+o(1))\frac{1}{3}\delta p^2 n^2}$

Preferred for δ < 27/8

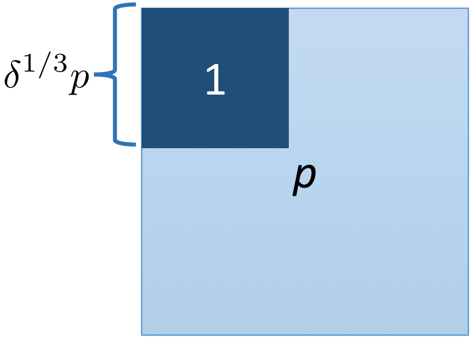
Theorem (Chatterjee—Dembo/Eldan + Lubetzky—Z.).
Let X denote the number of triangles in $\mathbf{G}(n,p)$.
Fix $\delta > 0$. With $p \to 0$ and and $p \geq n^{-1/18} \log n$,

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

Similar results for the number of K_t

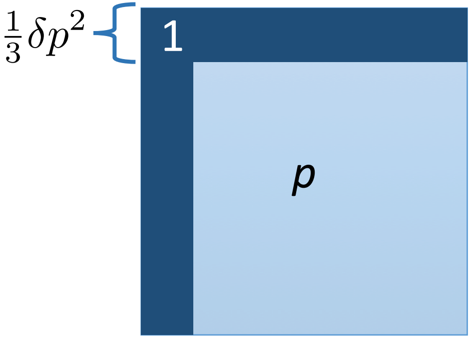
[Bhattacharya, Ganguly, Lubetzky, Z. '17]
Solution for every H

Proof of lower bound:



$(1+o(1))\delta p^3 \binom{n}{3}$
extra triangles

With probability:
$p^{(1+o(1))\frac{1}{2}\delta^{2/3}p^2n^2}$



$(1+o(1))\delta p^3 \binom{n}{3}$
extra triangles

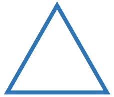
With probability:
$p^{(1+o(1))\frac{1}{3}\delta p^2n^2}$

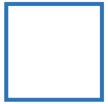
Theorem (Bhattacharya, Ganguly, Lubetzky, Z. '17)**.**
Fix $\delta > 0$ and a graph H. Let X_H = # copies of H in $\mathbf{G}(n,p)$.
With $p \to 0$ and and $p \geq n^{-1/6e(H)} \log n$,

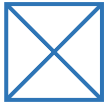
$$\mathbb{P}(X_H \geq (1+\delta)\mathbb{E}X_H) = p^{(c_H(\delta)+o(1))p^{\Delta}n^2}$$

where Δ = max deg H, and $c_H(\delta) > 0$ is an explicit constant ...

For example

For $H = C_3$  $c_H(\delta) = \min\{\frac{1}{2}\delta^{2/3}, \frac{1}{3}\delta\}$

For $H = C_4$  $c_H(\delta) = \min\left\{\frac{1}{2}\delta^{1/2}, -1 + \left(1 + \frac{1}{2}\delta\right)^{1/2}\right\}$

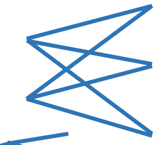
For $H = K_4$  $c_H(\delta) = \min\{\frac{1}{2}\delta^{1/2}, \frac{1}{4}\delta\}$

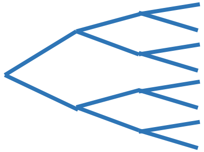
Theorem (Bhattacharya, Ganguly, Lubetzky, Z. '17).
Fix $\delta > 0$ and a graph H. Let X_H = # copies of H in $\mathbf{G}(n,p)$.
With $p \to 0$ and and $p \geq n^{-1/6e(H)} \log n$,

$$\mathbb{P}(X_H \geq (1+\delta)\mathbb{E}X_H) = p^{(c_H(\delta)+o(1))p^{\Delta}n^2}$$

where Δ = max deg H, and $c_H(\delta) > 0$ is an explicit constant …

For example

For $H = K_{2,3}$  $c_H(\delta) = (1+\delta)^{1/2} - 1$

For $H =$  $c_H(\delta) = -\frac{3}{2} + \frac{1}{2}\sqrt{5 + 4\sqrt{1+\delta}}$

Theorem (Bhattacharya, Ganguly, Lubetzky, Z. '17).
Fix $\delta > 0$ and a graph H. Let X_H = # copies of H in $\mathbf{G}(n,p)$.
With $p \to 0$ and and $p \geq n^{-1/6e(H)} \log n$,

$$\mathbb{P}(X_H \geq (1+\delta)\mathbb{E}X_H) = p^{(c_H(\delta)+o(1))p^{\Delta}n^2}$$

where Δ = max deg H, and $c_H(\delta) > 0$ is an explicit constant …

Independence polynomial: $P_H(x) := \sum_{\text{indep set } I} x^{|I|}$

Let H^* denote the subgraph of H induced by its maximum degree vertices.

Let $\theta > 0$ satisfy $P_{H^*}(\theta) = 1 + \delta$. Then, for a connected graph H,

$$c_H(\delta) = \begin{cases} \min\{\theta, \frac{1}{2}\delta^{2/v(H)}\} & \text{if } H \text{ is regular} \\ \theta & \text{if } H \text{ is irregular} \end{cases}$$

Large deviations in random hypergraphs

Ongoing joint work with Yang Liu

- $\mathbf{G}^{(k)}(n,p)$: random *k*-uniform hypergraph, where every triple appears with probability *p* independently
- Given some fixed 3-uniform hypergraph *H*, what can you say about upper tails of *H*-densities in $\mathbf{G}^{(3)}(n,p)$?
- Possible ways to embed extra edges
  - Plant clique: all triples contained in some chosen subset *S* of vertices
  - Plant 2-hub: all triples with at least two vertices in *S*
  - Plant 1-hub: all triples with at least one vertex in *S*
  - *A simultaneous overlay of these constructions*
- Currently we understand what happens when *H* is a clique …

Arithmetic progressions

Theorem (Bhattacharya, Ganguly, Shao, Z.). Fix k and $\delta > 0$. Let X_k denote the number of k-term arithmetic progressions in a random subset of $\{1, 2, \ldots, N\}$ where every element is included with probability p. With $p \to 0$ and $p \geq n^{-1/(6k\lceil(k-2)/2\rceil)} \log n$,

$$\mathbb{P}(X_k \geq (1+\delta)\mathbb{E}X_k) = p^{(1+o(1))\sqrt{\delta p^k n^2}}$$

Recent improvement by Briët--Gopi

The order in the exponent was determined by Warnke, and holds for all

$$p \gtrsim \left(\frac{\log n}{n}\right)^{1/(k-1)}$$

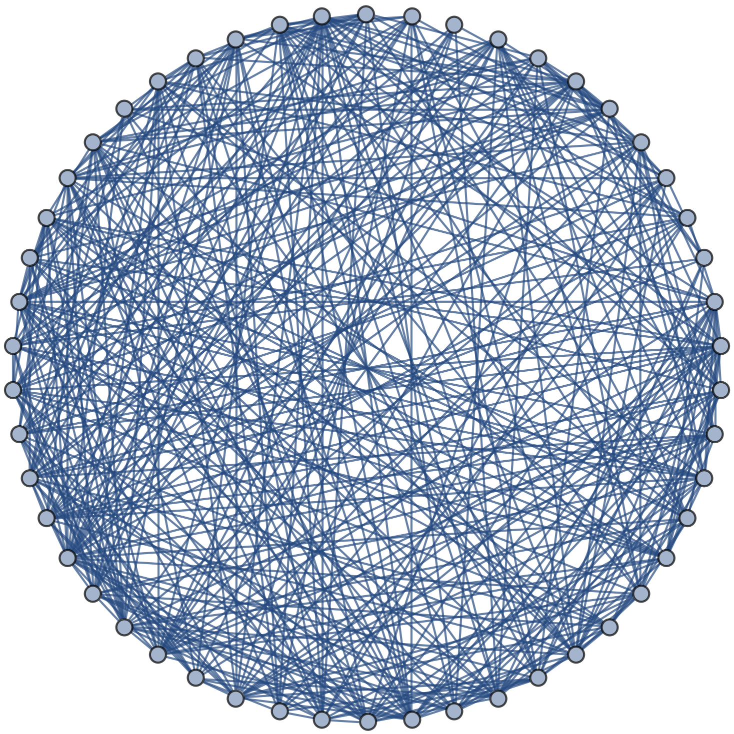
- Proof of lower bound: plant an interval of length $\sim \sqrt{\delta p^k n^2}$

Dense setting

$\mathbf{G}(n,p)$

p constant

$n \to \infty$



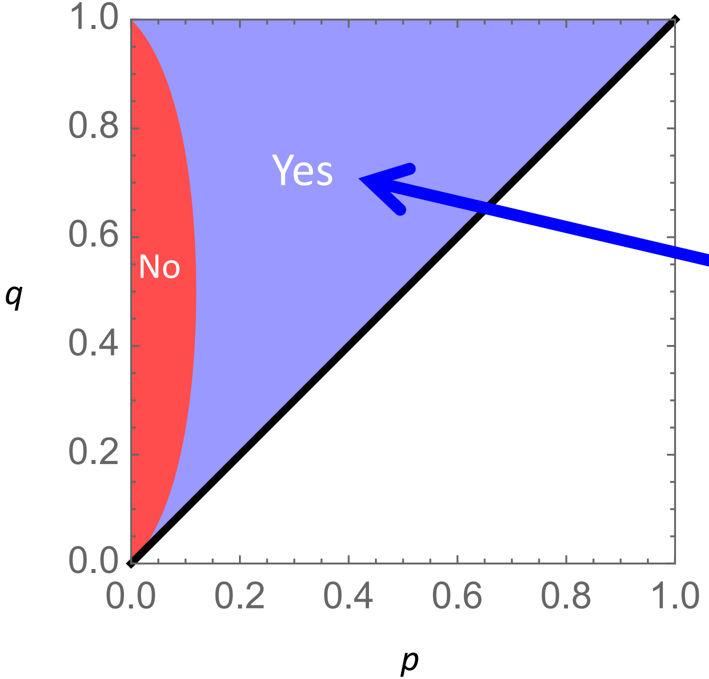
Question (Chatterjee—Varadhan ’11).
Fix $0 < p < q < 1$. Let G be an instance of $\mathbf{G}(n,p)$ conditioned on having at least as many triangles as a typical $\mathbf{G}(n,q)$.
Is $G \approx \mathbf{G}(n,q)$ in cut-distance?

$e_G(U) = \binom{|U|}{2}q + o(n^2)$ for every $U \subset V$.

Possibilities:

- **Yes**: more edges, uniformly distributed (replica symmetry)
- **No**: some other non-uniform distribution of edges (symmetry breaking)

Does **G**(n,p), conditioned on having $\geq \binom{n}{3}q^3$ triangles, look like **G**(n,q)?

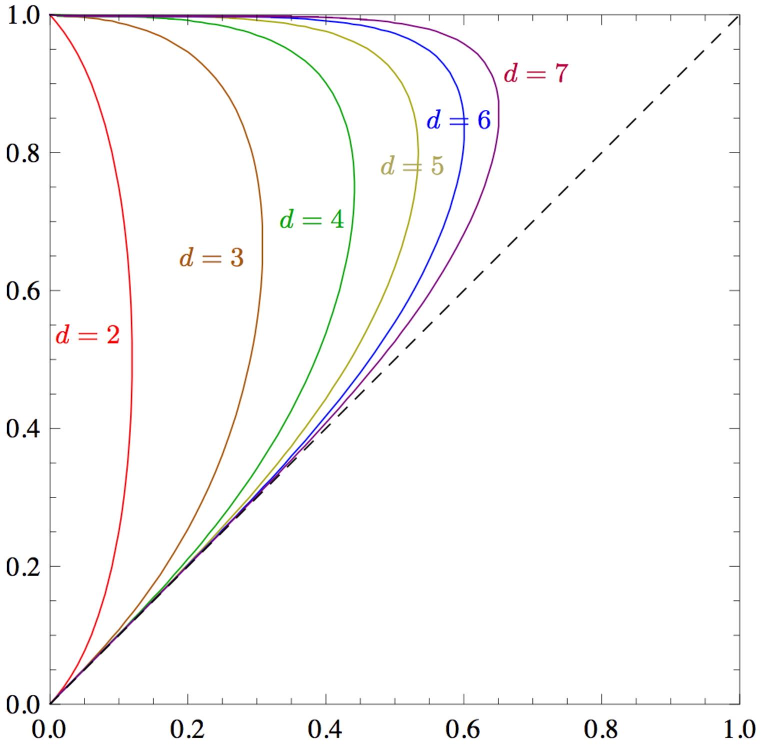


Theorem (Lubetzky—Z. ’15). Replica symmetry phase:

$$p \geq \left(1 + (q^{-1} - 1)^{1/(1-2q)}\right)^{-1}$$

Earlier partial results:
[Chatterjee & Dey ’10] [Chatterjee & Varadhan ’11]

Upper tail of H-density



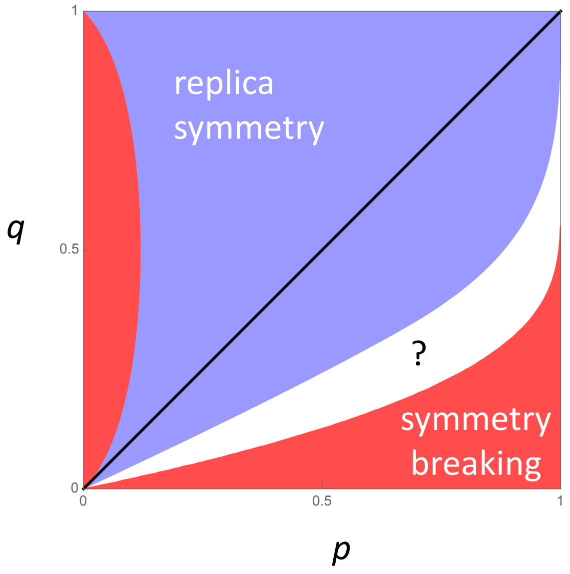
[Lubetzky—Z. '15] Identified the phase diagram for H-density if H is d-regular.

The phase diagram depends only on d.

Also: upper tail large deviation of the top eigenvalue of $\mathbf{G}(n,p)$.
(Top eigvalue typically $\approx np$; what if $\geq nq$?)
Same diagram as $d = 2$

Open: any irregular H,
e.g., a path of two edges

Lower tail



[Z. 2017]

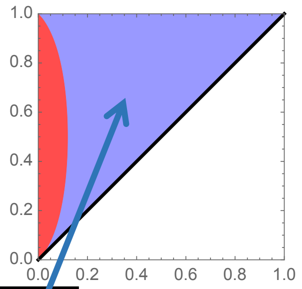
$X \leq (1-\delta)\mathbb{E}X$ as $p \to 0$

$\delta = 0.01$ Replica symmetry

δ^* critical ???

$\delta = 0.99$ Symmetry breaking

Upper tail phase diagram

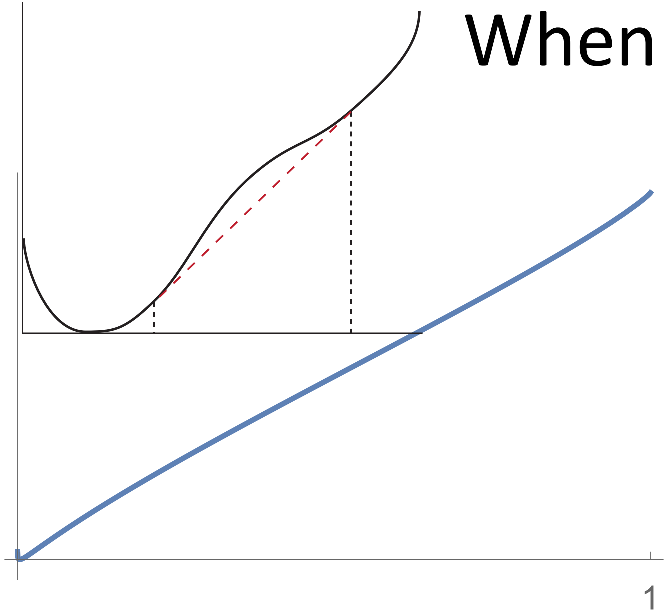


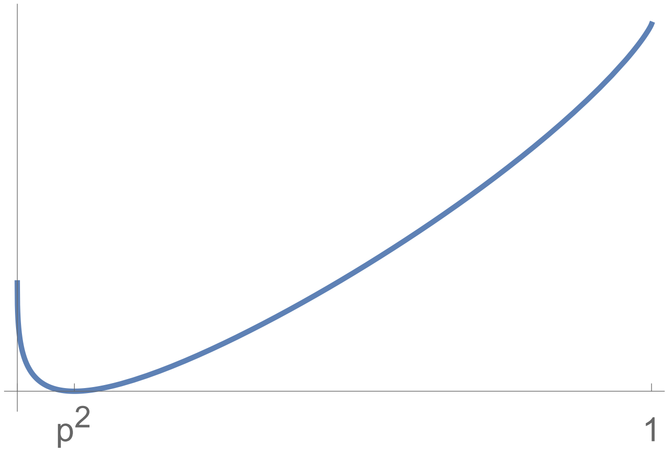
Theorem (Lubetzky—Z. ’15)**.**
Let $0 < p < q < 1$. The constant graphon $W \equiv q$ minimizes $\int_{[0,1]^2} I_p\big(W(x,y)\big)\,dxdy$ subject to

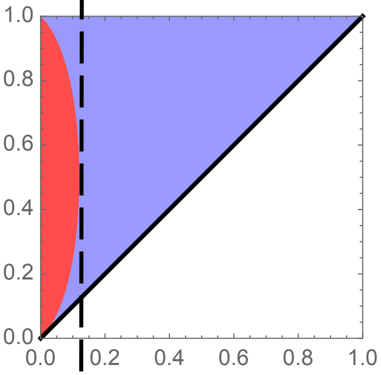
$$\int_{[0,1]^3} W(x,y)W(x,z)W(y,z)\,dxdydz \geq q^3$$

if and only if the point $(q^2, I_p(q))$ lies on the convex minorant of $x \mapsto I_p(\sqrt{x})$.

When is $x \mapsto I_p(\sqrt{x})$ convex?







Not convex for

$$p < \frac{1}{1+e^2}$$

Always convex for

$$p \geq \frac{1}{1+e^2} \approx 0.12$$

Exponential random graph model (ERGM)

A random graph *G* on *n* vertices, where *G* is chosen with probability proportional to $e^{h(G)}$

Examples:

- $h(G) \equiv 1$ same as **G**(*n*, 1/2)
- $h(G) = \beta|E(G)|$ same as **G**(*n*, *p*) for some $p = p(\beta)$
- $h(G) = \beta|E(G)| + \gamma|T(G)|$
  - $\gamma > 0$ prefer more triangles
  - $\gamma < 0$ prefer fewer triangles

T(*G*) = triangles in *G*

Exponential random graph models

- MCMC: Glauber dynamics by flipping a random edge according to its conditional probability
- Does it converge to desired distribution? How quickly?
- [Bhamidi, Bresler, Sly ’08] For the “dense” ERGM

$$p(G) = \frac{1}{Z} \exp\left[\binom{n}{2}\left(\beta_1 t(K_2, G) + \beta_2 t(K_3, G)\right)\right]$$

  with $\beta_2 \geq 0$
  - High temperature regime: mixing time $\Theta(n^2 \log n)$
    “not appreciably different from Erdős–Rényi random graph”
  - Lower temperature regime: mixing time $e^{\Omega(n)}$
- [Chatterjee, Diaconis ’13] Dense ERGMs can be analyzed via the graphon variational problem:

$$\text{Maximize} \quad \underbrace{h(W)}_{\substack{\text{Hamiltonian} \\ \text{(normalized)}}} + \underbrace{I(W)}_{\text{entropy}} \quad \text{over graphons } W$$

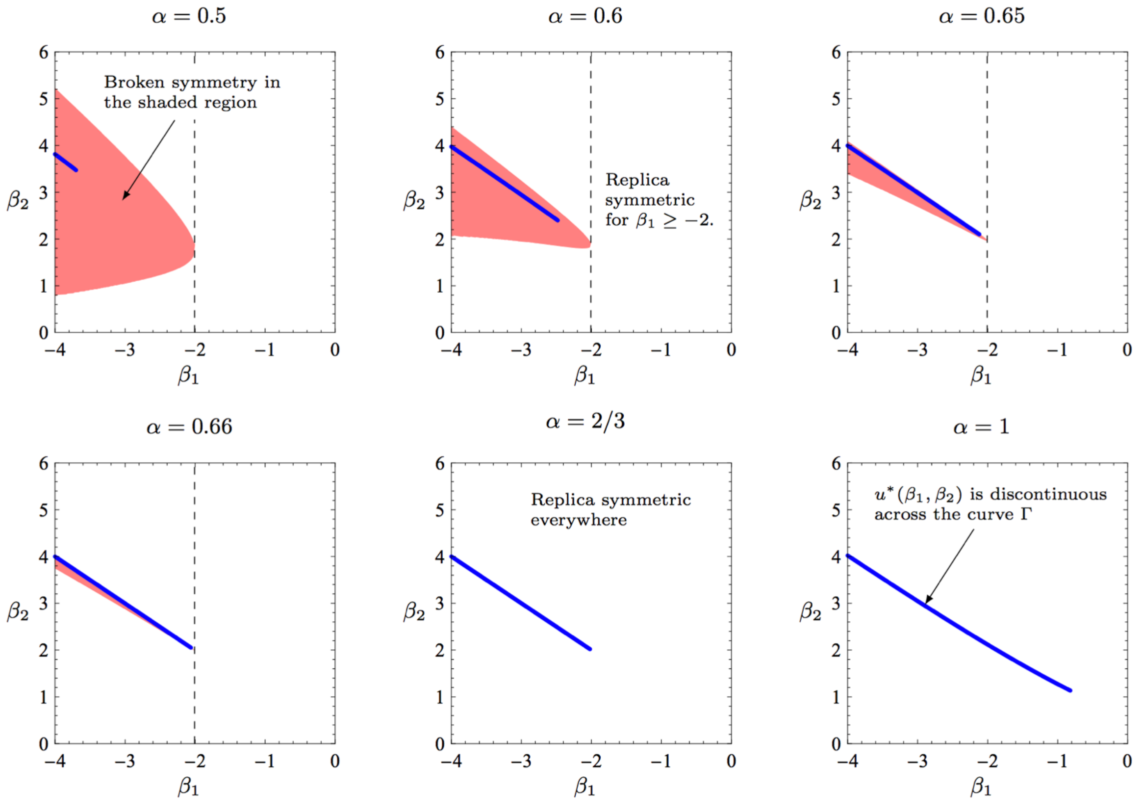
With $\beta_2 \geq 0$ always maximized by constant graphon

Weakness of model?

- For the ERGM
$$p(G)=\frac{1}{Z}\exp\left[\binom{n}{2}\left(\beta_1 t(K_2,G)+\beta_2 t(K_3,G)\right)\right]$$
with $\beta_2 \geq 0$ (similar if allow more terms), the graphon that maximizes the variational problem is the constant graphon, so ERGM ≈ **G**(n, p) in this case, so ERGM does not accomplish the goal of modeling triangle clustering

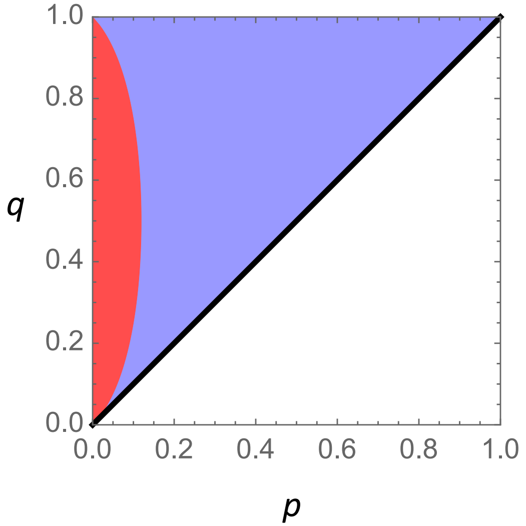
- [Lubetzky, Z. ’15] Modify the model as
$$p(G)=\frac{1}{Z}\exp\left[\binom{n}{2}\left(\beta_1 t(K_2,G)+\beta_2 t(K_3,G)^{\alpha}\right)\right]$$
For $\alpha < 2/3$ we get non-Erdős–Rényi behavior

ERGM $p(G) = \frac{1}{Z}\exp\left[\binom{n}{2}\left(\beta_1 t(K_2,G) + \beta_2 t(K_3,G)^{\alpha}\right)\right]$



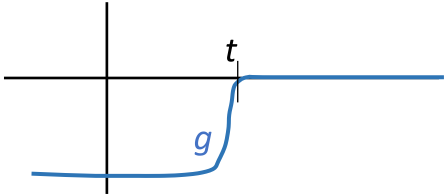
Large deviations:

$t(K_3,\mathbf{G}(n,p)) \geq \binom{n}{3} q^3$



Partition function of ERGM → LDP

- Estimating the partition function $Z = \sum_G e^{h(G)}$ is closely related to sampling
- Estimating the partition function also leads to large deviation principles. Take g to be the function



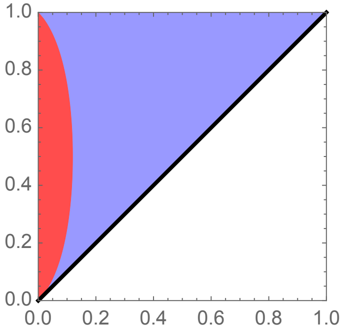
- Then large deviation $T(G) > t$corresponds to computing

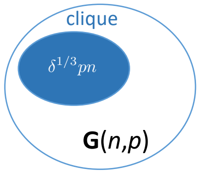
$$\sum_{G:T(G)>t} p^{|E(G)|}(1-p)^{|E(G)|} \approx \sum_G p^{|E(G)|}(1-p)^{|E(G)|} e^{g(T(G))} = \sum_G e^{h(G)} = Z_h$$

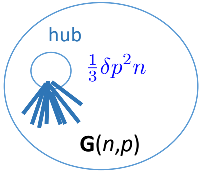
  for some appropriate *h*

- Recent advances give better methods for estimating the partition function, allowing somewhat sparser graphs
  - [Chatterjee—Dembo ’15] Stein’s method [Eldan ’17+] stochastic calculus and control

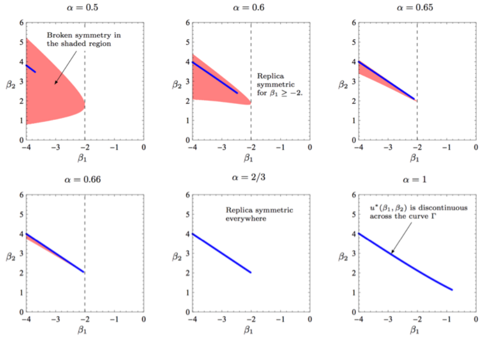
Summary







- Large deviation principles
- Variational problem
- Exponential random graphs
- Large deviations of triangle counts in **G**(n, p)
  - Constant p: replica symmetry vs. symmetry breaking
  - Sparse $p \to 0$: planting cliques or hubs



- Exponential random graphs
  - Adding an exponent introduces non-Erdős–Rényi behavior

Thank you!