# Large deviations and the regularity method for the Erdős–Rényi (hyper)graph

One World Probability Seminar, 2021/06/17

---

Nicholas Cook, Duke University

Based on joint work with Amir Dembo and Huy Tuan Pham

## Extremal vs typical behavior

Many problems in combinatorics are of the form:

*For two functions $f, g : \mathcal{G} \to \mathbb{R}$ on a finite set $\mathcal{G}$,*
*understand the behavior of $f$ under a constraint on $g$.*

Example: Counting subgraphs. $\mathcal{G} = \mathcal{G}_n$ is the set of simple graphs over $[n]$,

$$f(G) = \hom(K_3, G)\,, \qquad g(G) = \hom(K_2, G)$$

where $\hom(H, G) = \#\{\phi : V(H) \to [n] : \ \phi(e) \in E(G) \quad \forall\, e \in E(H)\}$.

**Extremal problem:** Maximize $f$ subject to $g \le L$.

Kruskal–Katona Theorem $\Rightarrow \hom(K_3, G) \le L^{3/2}$. (Saturated by $G = K_{\sqrt{L}}$.)

Extensions to general (hyper)graph embeddings by Alon, Friedgut–Kahn.

Also a consequence of Finner's generalized Hölder inequality.

**Typical behavior:** Typical size of $f(\boldsymbol{G})$ for random $\boldsymbol{G} \in \mathcal{G}$ with $g(\boldsymbol{G}) = L$ (microcanonical ensemble) or $\mathbb{E} g(\boldsymbol{G}) \sim L$ (grand canonical ensemble).

$\boldsymbol{G} = \boldsymbol{G}(n, p)$ : $\ \mathbb{E}\hom(K_2, \boldsymbol{G}) \sim n^2 p\,,\ \mathbb{E}\hom(K_3, \boldsymbol{G}) \sim n^3 p^3 \sim L^{3/2} p^{3/2}$.

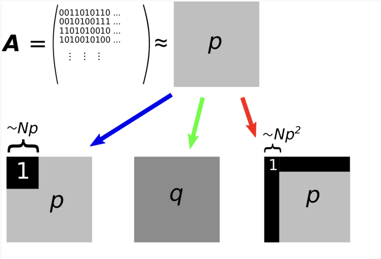
LLN and CLT (Ruciński '88)

# Extremal vs typical vs large deviation regimes

Between the extremal and typical regimes there are large deviation regimes.

**Tail problem:** For $q \in (p, 1)$, estimate
$\mathbb{P}(\hom(K_3, \boldsymbol{G}(n,p)) \sim n^3 q^3)$.

**Conditional structure problem:** Typical structure of $\boldsymbol{G}$ on the (*atypical*) large deviation event? Related to stability of optimizers for the extremal problem (cf. recent works of Keevash on Kruskal–Katona, Ellis et al. on Loomis–Whitney).




There's also the lower tail. **Extremal problem:** Razborov's theorem.
**Large deviations:** Chatterjee–Varadhan '11, Zhao '15, Kozma–Samotij '21.

Large deviations problem inherits some structure of the extremal problem, but there are also new phenomena (and in the hypergraph setting the picture is even more rich, still poorly understood).

**In this talk:** I'll motivate a general study of the high-dimensional geometry of the space of graphs $\mathcal{G}_n \subset \mathbb{R}^{\binom{n}{2}}$, in as seen under various norms (old and new).

# Large deviation principles (LDPs)

In the general setting of a topological measurable space $\mathcal{X}$, an LDP provides a description of the large scale landscape of $\mathcal{X}$ with respect to a sequence of probability measures $\mu_n$.

Informally it says that for large $n$,

$$\mu_n(\mathcal{E}) \approx \exp(-R_n \inf\{J(x) : x \in \mathcal{E}\})$$

for arbitrary $\mathcal{E}$ in the topology, where $R_n$ is the *speed* and $J : \mathcal{X} \to \mathbb{R}_+$ is the *rate function*.

Applying to level sets, yields results for joint tail events for continuous functionals, e.g. of the form

$$\mu_n(\{f_1 > L_1, f_2 > L_2\}) \approx \exp\Big( - R_n \inf \Big\{ J(x) : f_1(x) > L_1,\ f_2(x) > L_2 \Big\} \Big) .$$

Integration by parts gives free energy for Gibbs measures (Varadhan's Lemma).

# Large deviation principles (LDPs)

$$\text{LDP:} \qquad \mu_n(\mathcal{E}) \approx \exp(-R_n \cdot \inf\{J(x) : x \in \mathcal{E}\}).$$

Results for tails $\mathbb{P}(\hom(H, \boldsymbol{G}) \geq L)$ are of this form, with $\mathcal{X} = \mathbb{R}$ and $\mu_n$ the distribution of $\hom(H, \boldsymbol{G})$.

However, a more powerful result would be to have an LDP for $\boldsymbol{G}(n, p)$ and deduce tails for functionals $f : \mathcal{G}_n \to \mathbb{R}$, such as embeddings counts.

Problem: This doesn't fit the classical framework: The Erdős–Rényi measures $\mu_{n,p}$ are not a sequence of measures on fixed topological space $\mathcal{X}$, but rather a growing sequence of spaces $\mathcal{G}_n$.

Solution: The topological space $\mathcal{W}$ of *graphons* provides a completion for the space of all graphs of all sizes.

**Theorem (Chatterjee–Varadhan '11)**

For fixed $p \in (0, 1)$ the sequence of Erdős–Rényi measures $\mu_{n,p}$ on $\mathcal{W}$ satisfies an LDP with speed $n^2$ and a certain rate function.

# The regularity method (topological perspective)

Graphon theory provides a topological reformulation of the regularity method from extremal graph theory. In particular we have the following qualitative versions of classic lemmas:

**The Regularity Lemma:** $\mathcal{W}$ is compact (after quotient by vertex relabelings).

**The Counting Lemma:** Homomorphism counts extend to continuous (nonlinear) functionals on $\mathcal{W}$.

The Chatterjee–Varadhan LDP (whose proof hinges on the regularity lemma) combined with the Counting Lemma yields LDPs for homormorphism counts in $\boldsymbol{G}_{n,p}$, reducing the upper tail to an optimization problem for the rate function (analyzed in CV11 and Lubetzky–Zhao '12).

Problem: for $p = o(1)$, graphon space cannot provide an informative LDP.

- ⋆ Can't be rectified by rescaling of subgraph statistics and measures $\mu_{n,p}$, due to a **localization phenomenon**: main contribution to large deviations comes from a vanishing proportion of edges in a dense configuration (recall cliques and hubs).
- ⋆ These structures can occur at various scales.

# Quantitative large deviations for random graphs (C.–Dembo '18)

The CV LDP is really just breaking the space $\mathcal{G}_n \cong \{0,1\}^{\binom{n}{2}}$ into neighborhoods of a bounded number of graphons. Reduces to studying the probability $\boldsymbol{G}(n,p)$ lies in a small ball in the cut norm.

This can be made quantitative. Key fact: In a topological **vector space**, we have the non-asymptotic bound (consequence of minimax theorem)

$$\mu(\mathcal{K}) \leq \exp(-\inf_{x\in\mathcal{K}} J(x))$$

for convex $\mathcal{K}$, appropriate $J$ (convex dual of log-MGF $y \mapsto \log \int e^{\langle y,\cdot\rangle} d\mu$) .

So we can get quantitative tail estimates for $\hom(H, \boldsymbol{G})$ by efficiently covering $\mathcal{X}_n := [0,1]^{\binom{n}{2}}$ with convex sets on which $\hom(H,\cdot)$ does not vary much, where

$$\hom(H,X) = \sum_{\phi: V(H)\to[n]} \prod_{e\in E(H)} X_{\phi(e)}.$$

| | | |
|---|---|---|
| Estimates on covering numbers | $=$ | quantitative compactness. |
| Controlling variation of $\hom(H,\cdot)$ (under some norm) | $=$ | quantitative counting lemma. |

# Quantitative large deviations for random graphs (C.–Dembo 18)

Regard $X \in \mathcal{X}_n = [0,1]^{\binom{n}{2}}$ as adjacency matrices for edge-weighted graphs.

Recall the spectral operator norm: $\|M\|_{\ell_2 \to \ell_2} = \sup_{\|u\|_2, \|v\|_2 = 1} |\langle u, Mv \rangle|$.

**Spectral regularity lemma**

Denote $R_{n,p} := n^2 p^\Delta \log(1/p)$. Assuming $np^\Delta \gg \log n$, for any fixed $\varepsilon > 0$ (small) and $K \geq 1$ (large) we can cover $\mathcal{X}_n$ with a collection of $e^{o(R_{n,p})}$ balls

$$B(X_j, \varepsilon) = \{X \in \mathcal{X}_n : \|X - X_j\|_{\ell_2 \to \ell_2} \leq \varepsilon n p^{\Delta/2}\}$$

together with a set $\mathcal{E}_0$ of measure $\mu_{n,p}(\mathcal{E}_0) \leq \exp(-cKR_{n,p})$.

**Spectral counting lemma**

Let $p \in (0,1)$, $L \geq 1$ and $\varepsilon > 0$ be arbitrary. Suppose $\mathcal{K} \subseteq \mathcal{X}_n$ is a convex set of diameter at most $\varepsilon n p^{\Delta_\star(H)}$ in the spectral norm, and that for every induced strict subgraph $F \prec H$ there exists $X \in \mathcal{K}$ such that $t_p(F, X) \leq L$. Then

$$|t_p(H, X_1) - t_p(H, X_2)| \lesssim_H L\varepsilon \qquad \forall X_1, X_2 \in \mathcal{K}.$$

Here $\Delta_\star(H) := \frac{1}{2} \max_{\{u,v\} \in E(H)} \deg(u) + \deg(v)$, $\quad t_p(H, X) = \frac{\hom(H,X)}{n^{v(H)} p^{e(H)}}$.

# New norms for detecting localization (C.–Dembo–Pham '21)

To extend this approach to hypergraphs, need different norms (no spectral theory for tensors).

As motivation we first recall the cut norm for matrices:

$$\|A - B\|_{\square} = \max_{U,V\subseteq[n]} \Big| \sum_{i\in U, j\in V} A_{ij} - B_{ij} \Big| = \max_{U,V\subseteq[n]} \big| \langle 1_U \otimes 1_V, A - B\rangle \big|.$$

**Counting lemma:** $|\hom(H, A) - \hom(H, B)| \leq e(H)\|A - B\|_{\square}$.

**Proof:**

To extend this approach to hypergraphs, need different norms (no spectral theory for tensors).

As motivation we first recall the cut norm for matrices:

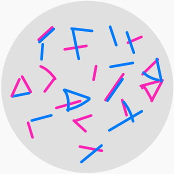
$$\|A - B\|_{\square} = \max_{U,V \subseteq [n]} \Big| \sum_{i \in U, j \in V} A_{ij} - B_{ij} \Big| = \max_{U,V \subseteq [n]} \big| \langle 1_U \otimes 1_V, A - B \rangle \big|.$$

**Counting lemma:** $|\hom(H, A) - \hom(H, B)| \le e(H) \|A - B\|_{\square}$.

**Proof:** $A$, $B$ adjacency matrices of graphs over common vertex set $[n]$.

pink edges

blue edges



# New norms for detecting localization (C.–Dembo–Pham ’21)

To extend this approach to hypergraphs, need different norms (no spectral theory for tensors).

As motivation we first recall the cut norm for matrices:

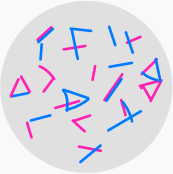
$$\|A - B\|_{\square} = \max_{U,V\subseteq[n]} \Big| \sum_{i\in U, j\in V} A_{ij} - B_{ij} \Big| = \max_{U,V\subseteq[n]} \big|\langle 1_U \otimes 1_V, A - B\rangle\big|.$$

**Counting lemma:** $|\,\mathrm{hom}(H, A) - \mathrm{hom}(H, B)| \le e(H)\|A - B\|_{\square}$.

**Proof:** $A$, $B$ adjacency matrices of graphs over common vertex set $[n]$.

pink edges

blue edges



For $H = K_3$:

$\# \triangle - \# \triangle$

$= (\# \triangle - \# \triangle)$

$+ (\# \triangle - \# \triangle)$

$+ (\# \triangle - \# \triangle)$

## New norms for detecting localization (C.–Dembo–Pham ’21)

To extend this approach to hypergraphs, need different norms (no spectral theory for tensors).

As motivation we first recall the cut norm for matrices:

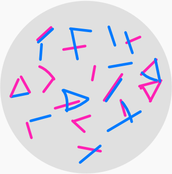
$$\|A-B\|_{\square} = \max_{U,V\subseteq[n]} \Big| \sum_{i\in U, j\in V} A_{ij} - B_{ij} \Big| = \max_{U,V\subseteq[n]} \big| \langle 1_U \otimes 1_V, A-B\rangle \big|.$$

**Counting lemma:** $|\hom(H,A) - \hom(H,B)| \le e(H)\|A-B\|_{\square}$.

**Proof:** $A$, $B$ adjacency matrices of graphs over common vertex set $[n]$.

pink edges

blue edges



For $H = K_3$:

$\# \triangle - \# \triangle$

$= (\# \triangle - \# \triangle)$

$+ (\# \triangle - \# \triangle)$

$+ (\# \triangle - \# \triangle)$

$$(\# \triangle - \# \triangle) = \sum_{i,j,k=1}^{n} A_{ij} A_{jk} (A_{ik} - B_{ik})$$

$$= \sum_{i,k=1}^{n} (A_{ik} - B_{ik}) \underbrace{\sum_{j=1}^{n} A_{ij} A_{jk}}_{=: T_{ik}}$$

$$= \langle A-B, T \rangle$$

# New norms for detecting localization (C.–Dembo–Pham '21)

The cut-norm counting lemma combines well with the **Frieze–Kannan weak regularity lemma:** every matrix $M$ with bounded entries can be decomposed as

$$M = M_{struct} + M_{rand}$$

where $\|M_{rand}\|_\square \le \varepsilon n^2$ and $M_{struct}$ is a linear combination of $O(1/\varepsilon^2)$ cut matrices $T_k = 1_{U_k} \otimes 1_{V_k}$.

**Proof sketch:** (Energy increment argument) Having found $T_1, \dots, T_k$, let $M^{(k)}_{struct}$ be the projection of $M$ to their linear span. If $\|M - M^{(k)}_{struct}\|_\square \le \varepsilon n^2$ we stop; otherwise there is a cut matrix $T_{k+1}$ with $|\langle T_{k+1}, M - M^{(k)}_{struct}\rangle| > \varepsilon n^2$. Must stop in $O(1/\varepsilon^2)$ steps by Pythagoras's theorem and $\|M\|_2^2 = O(n^2)$. $\square$

Basic argument has been put in very general setting by Gowers: For a finite collection $\mathcal{T}$ of "structured" elements spanning a vector space $V$, define

$$\|f\| = \inf\left\{ \sum_j |\lambda_j| : f = \sum_j \lambda_j T_j \right\}, \quad f \in V.$$

Then there is a decomposition $f = f_{struct} + f_{rand}$ with $\|f_{struct}\| + \|f_{rand}\|^* \le \|f\|_2$. Stronger variants give $f = f_{struct} + f_{rand} + f_{small}$.

# New norms for detecting localization (C.–Dembo–Pham ’21)

Write $\mathcal{T} = \{1_U \otimes 1_V : U, V \subseteq [n]\}$ for the set of cut matrices ("test tensors").
In place of

$$\|A - B\|_\square = \sup_{T \in \mathcal{T}} |\langle T, A - B \rangle|$$

we use

$$\|A - B\|_{\mathrm{B}}^* := \sup_{T \in \mathcal{T}} \frac{|\langle T, A - B \rangle|}{\|T\|_{\mathrm{B}}}$$

where the "size" of a cut matrix $T = 1_U \otimes 1_V$ is

$$\|T\|_{\mathrm{B}} := (|U| \vee n_0) \cdot (|V| \vee n_0)$$

with the cutoff scale $n_0 := np^{\Delta - 1}$.

**Ex:** Planted clique on $\asymp np^{\Delta/2}$ vertices is detected by cut $1_U \otimes 1_U$ with $|U|$ above the cutoff scale.

vs. Planted hub on $U \times [n]$ with $|U| \asymp np^{\Delta} \ll n_0$.

## The decomposition theorem

**Decomposition theorem (matrix case)**

Assuming $np^{\Delta+1} \gg \log n$, for any fixed $\varepsilon > 0$ (small) and $K \geq 1$ (large), outside of an exceptional set $\mathcal{E}_\star \subset \{0,1\}^{\binom{n}{2}}$ of measure $\mu_{n,p}(\mathcal{E}_\star) \leq \exp(-cKn^2 p^\Delta \log(1/p))$, every adjacency matrix $A \in \{0,1\}^{\binom{n}{2}}$ can be decomposed as

$$A = A_{struct} + A_{rand}$$

where $A_{struct}$ is a linear combination of $k = O(K\varepsilon^{-2}p^{-\Delta})$ cut matrices $T_1, \ldots T_k$, with

$$\sum_{i=1}^{k} \|T_i\|_{\mathrm{B}} \leq K\varepsilon^{-2} n^2 p^{\Delta-2}$$

and

$$\|A_{rand}\|_{\mathrm{B}}^{*} \leq \varepsilon p.$$

Proof goes by energy increment argument, bounding the probability that the algorithm runs very long – key point is that with (very) high probability we can stop much sooner than in the Frieze–Kannan regularity lemma (worst case).

**B$^*$-norm counting lemma**

Let $p \in (0,1)$, $L \geq 1$ and $\varepsilon > 0$ be arbitrary. Suppose $\mathcal{E} \subseteq \mathcal{A}_n$ is a set of diameter at most $\varepsilon p$ in the B$^*$-norm, and that there exists $A_0 \in \mathcal{E}$ such that $t_p(F, A_0) \leq L$ for every proper subgraph $F \subset H$. Then for every $X_1, X_2$ in the convex hull of $\mathcal{E}$,

$$|t_p(H, X_1) - t_p(H, X_2)| \lesssim_H L\,\varepsilon.$$

- ⋆ This is a deterministic statement – no Erdős–Rényi measure here.
- ⋆ Proof goes by induction on the number of edges, using the B$^*$-norm to control the error of one swap in terms of a sum over subgraphs, for which we apply inductive control and the assumption on $A_0$.
- ⋆ Statement is identical for hypergraphs – I just need to tell you what the B$^*$-norm is for $r$-tensors.

# Large deviations for random hypergraphs (C.–Dembo–Pham '21)

Consider now the Erdős–Rényi r-uniform hypergraph ($r$-graph) $\boldsymbol{G} = \boldsymbol{G}^{(r)}(n, p)$, identified with its symmetric adjacency tensor

$$\boldsymbol{A}_{i_1,\dots,i_r} = \mathbb{1}(\{i_1, \dots, i_r\} \in E(\boldsymbol{G})).$$

As for graphs we denote the normalized $H$-homomorphism counts

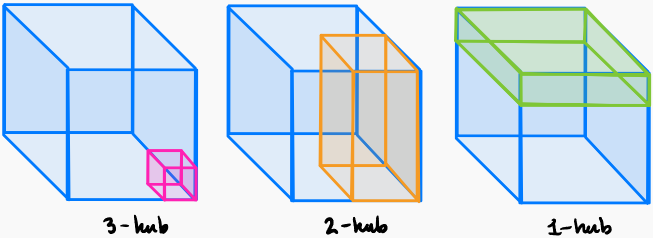
$$t_p(H, \boldsymbol{G}) = \frac{1}{n^{v(H)} p^{e(H)}} \sum_{\phi: V(H) \to [n]} \prod_{e \in E(H)} \boldsymbol{A}_{\phi(e)}.$$

Previous works on the upper tail:

- Lubetzky–Zhao '12: the case $p$ is fixed and $H$ is *linear* (all edge overlaps are of size 1). Proof follows CV11. Key: cut-norm counting lemma extends straightforwardly to $r > 2$ for linear $H$.
- Liu–Zhao '19: $n^{-1/6e(H)} \log n \ll p \ll 1$, $H$ a clique, or a certain linear 3-graph on 6 vertices.

For a fixed $r$-graph $H$, what is the least unlikely way for $\boldsymbol{G}$ to have many $H$-homomorphisms?

Liu & Zhao's mixed hubs




(and symmetrized)

# Large deviations for random hypergraphs (C.–Dembo–Pham '21)

**Theorem (Joint tails for hypergraph homomorphisms)**

Let $H_1, \ldots, H_m$ be $r$-uniform hypergraphs of maximum degree $\Delta$. For any fixed $\delta_1, \ldots, \delta_m > 0$, assuming $p \gg n^{-1/(\Delta+1)}$,

$$\log \mathbb{P}\Big(t_p(H_k, \boldsymbol{G}) > 1 + \delta_k \, , 1 \le k \le m\Big) = -(1 + o(1))\Phi_{n,p}\big(\underline{H}, \underline{\delta} + o(1)\big) \, ,$$

where

$$\Phi_{n,p}(\underline{H}, \underline{\delta}) := \inf_X \Big\{ D(\mu_X \| \mu_{n,p}) : t_p(H_k, X) \ge 1 + \delta_k \, , \, 1 \le k \le m \Big\}.$$

- Note $p$ may be fixed.
- Lower bound holds for $p \gg n^{-1/\Delta}$, and upper bound holds in wider range for certain $H_k$ (e.g. linear hypergraphs, sunflowers).

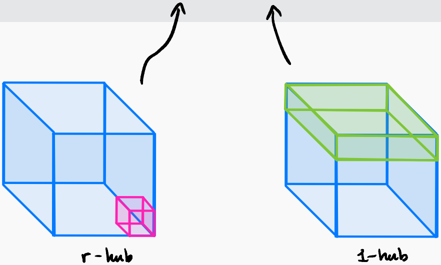
## Large deviations for random hypergraphs (C.–Dembo–Pham '21)

For the case that $H$ is a clique, we get an explicit asymptotic when $p = o(1)$ by combining with a result of Liu–Zhao '19 on the entropic optimization problem.

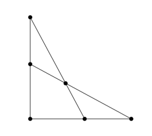
**Corollary A: The upper tail for clique homomorphisms**

For $H = K_k^{(r)}$ the $r$-uniform clique on $k$ vertices and $n^{-1/(\binom{k-1}{r-1}+1)} \ll p \ll 1$,

$$\log \mathbb{P}\Big(\hom(H, \boldsymbol{G}) \ge (1+\delta) n^k p^{\binom{k}{r}}\Big)$$
$$= -(1+o(1)) \min\left\{\frac{\delta^{r/k}}{r!}, \frac{\delta}{(r-1)!k}\right\} n^r p^{\binom{k-1}{r-1}} \log(1/p).$$



Liu and Zhao also considered a certain linear 3-graph – a simple case showing the geometry of the hypergraph LDP landscape is more complicated than in the 2-graphs case!



**Corollary B**

With $H$ the above 3-graph, for $n^{-1/2} \ll p \ll 1$,

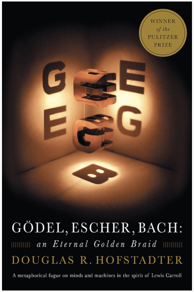
$$\log \mathbb{P}\Big( \hom(H, \boldsymbol{G}) \geq (1+\delta) n^k p^{\binom{k}{r}} \Big) = -\Big(\frac{1}{6} + o(1)\Big) \min\Big\{\sqrt{9+3\delta} - 3, \sqrt{\delta}\Big\} n^3 p^2 \log(1/p)\,.$$

## The $B^*$-norms for tensors

For tensors of higher rank, in place of cuts we take the more general class of Bernoulli test tensors of the form

$$T = \prod_{b \in B} \tau_b \qquad \text{(entrywise product)}$$

where the B is a family of proper subsets $b \subset [r]$, and $\tau_b$ are Bernoulli tensors depending only on coordinates in b. The "base" B is user-specified: for the counting lemma we just need that it "dominates" all edge overlaps.




The class $B = \binom{[r]}{r-1}$ always works, but sometimes there is significant benefit to using a more efficient base (e.g. for sunflowers can take a single b).

As before, $\|Z\|_B^* = \max_{T \in \mathcal{T}} \frac{|\langle Z, T\rangle|}{\|T\|_B}$.

The size $\|T\|_B$ of a test tensor is now a weighted combination of its volume and the volume of its factors: for the case $B = \binom{[r]}{r-1}$,

$$\|T\|_B = n^r p^{r(\Delta-1)} + \|T\|_1 + \sum_{b \in B} np^{\Delta-1} \|\tau_b\|_1 .$$

**Thanks for your attention!**