Large deviations for sparse random graphs

CIRM, Luminy
2019/04/19

Nicholas Cook, Stanford University
Based on joint work with Amir Dembo
Universality for typical behavior: Examples

- **CLT:** for \(X_1, X_2, \ldots \) iid, \(\mathbb{E} X_1 = 0, \mathbb{E} X_1^2 = 1\),
  \[\forall \ a < b, \quad \mathbb{P}\left\{ \frac{X_1 + \cdots + X_N}{\sqrt{N}} \in [a, b] \right\} \longrightarrow \gamma([a, b]) \quad \text{(universal)}\].

- Let \(A = (a_{ij})_{i,j=1}^N\) adjacency matrix for the Erdős–Rényi graph \(G(N, p)\) with \(0 < p \ll 1\), eigenvalues \(\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_N\).

\[\lambda_1\] is asymptotically **Gaussian**.

For \(p \gg N^{-2/3}\): \(\lambda_2, -\lambda_N\) follow the **Tracy–Widom** law [Lee–Schnelli ’16].
(But Gaussian for \(N^{-7/9} \ll p \ll N^{-2/3}\) [Huang–Landon–Yau ’17].)
Large deviations: Beyond universality

**CLT:** for $X_1, X_2, \ldots$ iid, $\mathbb{E} X_1 = 0$, $\mathbb{E} X_1^2 = 1$,

$$\forall a < b, \quad \mathbb{P} \left\{ \frac{X_1 + \cdots + X_N}{\sqrt{N}} \in [a, b] \right\} \longrightarrow \gamma([a, b]) \quad \text{(universal)}.$$

Compare Cramér's **Large deviations principle (LDP):**

$$\forall a < b, \quad \frac{1}{N} \log \mathbb{P} \left\{ \frac{X_1 + \cdots + X_N}{N} \in [a, b] \right\} \longrightarrow - \inf_{x \in [a, b]} J(x),$$

where $J$ is the non-universal rate function depending strongly on the law of $X_1$ (particularly its tail behavior).
How about **nonlinear** functionals?

**Example:** Extreme eigenvalues of random matrices / random graphs.

In this talk we focus on low-degree polynomials of Bernoulli variables. (Tails for eigenvalues will be under the hood.)

Note we consider outliers at scale $Np$ (for LDP at scale of the bulk cf. recent work of Guionnet–Husson).
Subgraph counts in $G(N, p)$

- Let $G \sim G(N, p)$ be an Erdős–Rényi graph on vertices $[N] = \{1, \ldots, N\}$.

- Number of triangles in $G$: $\mathcal{N}_\Delta(G) = \sum_{\{i,j,k\} \subset [N]} a_{ij} a_{jk} a_{ik}$
  (recall the adjacency matrix $A = (a_{ij})_{i,j=1}^N$ with $a_{ij} = 1$ if $\{i,j\}$ is an edge).
  $\mathbb{E} \mathcal{N}_\Delta(G) = \binom{N}{3} p^3$.

- **Question:** Conditional on $G$ having extra triangles, i.e. $\{ \mathcal{N}_\Delta(G) \geq \binom{N}{3} q^3 \}$ for some $q > p$, how are the edges distributed?
  A few possibilities:

  (A) As in $G(N, q)$?

  (B) As in $G(N, p)$ with a small planted clique?

  (C) As in $G(N, p)$ with a small planted hub?

  Answer is (A) for much (but not all!) of $0 < p < q < 1$ fixed. [Chatterjee–Varadhan ’11]+[Lubetzky–Zhao ’12].
**Conjecture:** Let $H$ have max degree $D$. For $N^{-1/D} \ll p \ll 1$, depending on the size of $\delta$,\[
 G \left\{ \mathcal{N}_H(G) \geq (1 + \delta) \mathbb{E} \mathcal{N}_H(G) \right\} \approx G(N, p) + \text{planted clique or hub}.
\]
The “infamous” upper tail for triangle counts [Janson–Ruciński ’02]

- Upper tail up to constant factors in the exponent:
  \[ P\{N_\Delta(G) \geq (1 + \delta) \mathbb{E} N_\Delta(G)\} = p^{\Theta(\delta N^2 p^2)}, \quad p \geq (\log N)/N. \]

  [Chatterjee ’12], [DeMarco–Kahn ’12]

- Recent work finds the leading exponential order:
  \[ P\{N_\Delta(G) \geq (1 + \delta) \mathbb{E} N_\Delta(G)\} = p^{(1+o(1)) \min\{\frac{\delta^2/3}{2}, \frac{\delta}{3}\} N^2 p^2} \]

(matching probabilities for planted clique or hub of appropriate size) for \(N^{-\kappa} \ll p \ll 1\), with

* \(\kappa = \frac{1}{41} - \epsilon\) [Chatterjee–Dembo ’14] + [Lubetzky–Zhao ’14]
* \(\kappa = \frac{1}{18} - \epsilon\) [Eldan ’16]
* \(\kappa = \frac{1}{3}\) [C.–Dembo ’18] (and \(\frac{1}{2} - \epsilon\) for cycles of length \(\ell \geq 4\)).
* \(\kappa = \frac{1}{2} - \epsilon\) [Augeri ’18] for cycles of length \(\ell \geq 3\).
* \(\kappa = 1 - \epsilon\) [Harel–Mousset–Samotij ’19 (yesterday)].

- How about general subgraphs?
Main result: Upper tail for general subgraph counts

- Let $H = (V, E)$ connected of max degree $D$, and assume $N^{-\kappa(H)} \ll p \ll 1$ for some $\kappa(H) \in (0, 1)$.
- [Chatterjee–Dembo ’14] + [Bhattacharya–Ganguly–Lubetzky–Zhao ’16]:
  \[
P \left( \mathcal{N}_H(G) \geq (1 + \delta) \mathbb{E} \mathcal{N}_H(G) \right) = p^{(1+o(1))c_H(\delta)N^2p^D}
\]
  matching the probability of a planted clique or hub up to sub-exponential factors, assuming $\kappa(H) = \frac{c}{D|E|}$.
  (Formula for $c_H(\delta)$ was obtained by [BGLZ ’16] as solution to LDP variational problem, valid down to $\kappa(H) = 1/D$.)
- [Eldan ’16] + [BGLZ ’16]: can take $\kappa(H) = \frac{1}{6|E|} - \epsilon$.
- [C.–Dembo ’18]: $\kappa(H) = \frac{1}{3D-2} - \epsilon$. 
Main result: Upper tail for general subgraph counts

\textbf{Theorem (C.–Dembo '18)}

Fix $H = (V, E)$ connected of max degree $D \geq 2$. If $N^{-\frac{1}{3D-2}} + \epsilon \leq p \ll 1$ then

$$\mathbb{P}\{N_H(G) \geq (1 + \delta) \mathbb{E}N_H(G)\} = p^{(1+o(1))c_H(\delta)N^2p^D}.$$ 

- This is currently the best result for general $H$, but see
  - [C–D '18], [Augeri '18] for sharpening in case of cycles (exploiting relationship to the spectrum of $A$);
  - very recent improvement to $\kappa(H) = \frac{2}{D} - \epsilon$ for $H$ non-bipartite $D$-regular by [Harel–Mousset–Samotij '19].

- We actually get a sharper $\kappa(H)$ (more complicated formula), in particular $\kappa(H) = 1/(2D - 1)$ for $H$ a star.

- We also get:
  - lower tails (reduction to variational problem – can solve only for Sidorenko graphs);
  - upper tails for $\lambda_1, \lambda_2, -\lambda_N$ (together with subsequent work by [Bhattacharya–Ganguly '18] solving the LDP variational problem).
Further motivation: Exponential random graphs (ERGs)

- *Edge-triangle model* (popular in sociology literature): for \( \alpha, \beta \in \mathbb{R} \),
  \[
  \mathbb{P}(G = G) = \frac{1}{Z_N(\alpha, \beta)} e^{\alpha N_e(G) + \beta \frac{1}{N} N_\Delta(G)}, \quad G \in \mathcal{G}_N.
  \]

- Estimates for upper tails of subgraph counts \( N_H(G_N, p) \) are closely related to estimates for the *partition function* (Varadhan’s Lemma and Bryc’s Theorem).

Dense case (\( \alpha, \beta \) fixed): [Bhamidi–Bressler–Sly ’08], [Chatterjee–Diaconis ’11], [Lubetzky–Zhao ’12].

- In progress: Apply our tools to get quantitative estimates on \( Z_N(\alpha, \beta) \) when \( \alpha, \beta \) can grow with \( N \), allowing for sparse ERGs. (following [Chatterjee–Dembo ’14], [Eldan ’16], [Eldan–Gross ’17].)

- Problems with ERGs:
  * For \((\alpha, \beta) \in \mathbb{R} \times \mathbb{R}_+ \) fixed, \( G \) looks like an Erdős–Rényi graph!
  * Degeneracy: for some ranges of \( \alpha, \beta \), \( G \) is close to empty or full.
Previous approaches to upper tails

- [Chatterjee–Dembo ’14]: large deviations for nonlinear functions $f : \{0, 1\}^d \to \mathbb{R}$ through the study of Gibbs measures $\mu$ with density $\mu(\{x\}) \propto e^{h(x)}$ for some Hamiltonian $h : \{0, 1\}^d \to \mathbb{R}$.

- Taking $e^{h(x)}$ as a “smooth” approximation to the indicator function $1_{f(x) \geq t}$, recover estimates on $\mathbb{P}(f(X) \geq t)$ from estimates on the partition function $Z = \sum_{x \in \{0, 1\}^d} e^{h(x)}$.

- C–D obtain conditions for validity of the naïve mean field approximation:

$$\log Z = \sup_{\nu \in M_1(\{0, 1\}^d)} \int h d\nu - H(\nu \| \mu) \approx \sup_{\nu \in M_1(\{0, 1\}^d)} \int h d\nu - H(\nu \| \mu)$$

where $H(\nu \| \mu)$ is the relative entropy.

- Extended and refined by [Yan ’15], [Eldan ’16], [Augeri ’18], [Austin ’18].

- Disadvantage: We incur errors in the passage from indicator functions to smooth approximations. Leads to results in sub-optimal range of sparsity.
Dense case (Chatterjee–Varadhan ’11)

- For a sequence of probability measures $\mu_N$ on a common topological space $\mathcal{X}$, large deviations principle (LDP) yields asymptotics of form

$$\mu_N(\mathcal{E}) \approx \exp \left( - v_N \inf_{x \in \mathcal{E}} J(x) \right), \quad \mathcal{E} \subseteq \mathcal{X},$$

for a rate function $J$ and speed $v_N$.

- In dense case ($p$ fixed), C–V get an LDP for $\mu_N(\cdot) = \mathbb{P}(G \in \cdot)$. What does it mean? $\mu_N$ live on separate spaces $\mathcal{G}_N \cong \{0, 1\}^\begin{pmatrix} N \\ 2 \end{pmatrix}$...

- The space of graphons provides a “completion” of $\bigcup_{N \geq 1} \mathcal{G}_N$:

$$\mathcal{W} := \{ g : [0, 1]^2 \rightarrow [0, 1] \text{ symmetric, Lebesgue measurable} \},$$

equipped with a topology coming from the cut-norm:

$$\|f\|_\Box := \sup_{S, T \subseteq [0, 1]} \left| \int_S \times T f(x, y) dx dy \right| .$$

- Graphons are limits of rescaled adjacency matrices, and $\| \cdot \|_\Box$ extends the matrix cut-norm $\|M\|_\Box = \max_{U, V \subseteq [N]} \left| \sum_{(i, j) \in U \times V} M_{ij} \right|$.
Identify a finite graph $G \in \mathcal{G}_N$ with $g \in \mathcal{W}$ via its adjacency matrix $A$, putting $g(x, y) := A_{[Nx], [Ny]}$. General $g \in \mathcal{W}$ is like a “continuum adjacency matrix”.

$$A = \begin{pmatrix} 0011010110 \ldots \\ 0010100111 \ldots \\ 1101010010 \ldots \\ 1010010100 \ldots \\ \vdots & \vdots & \vdots \end{pmatrix} \approx p$$
Graphon space provides a topological reformulation of the classic regularity method from extremal graph theory.

**Key fact 1:** The space of graphons with cut-norm topology is compact ($\approx$ Szemerédi’s regularity lemma).

<table>
<thead>
<tr>
<th>Theorem (Chatterjee–Varadhan)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Fix $p \in (0, 1)$ and for $N \geq 1$ let $G_N \sim G(N, p)$. The sequence of probability measures $\mu_N(\cdot) = \mathbb{P}(G_N \in \cdot)$ on the topological space of graphons satisfies an LDP (of speed $N^2$, with explicit rate function).</td>
</tr>
</tbody>
</table>

**Key fact 2:** the subgraph counting functions $\mathcal{N}_H(G)$, suitably extended to graphons, are continuous in the cut-norm topology. ($\approx$ the counting lemma).

**Corollary:** upper tails for subgraph counts $\mathcal{N}_H(G)$ (just apply the LDP to super-level sets).

**Moral:** the cut-norm topology is the right topology if you’re interested in subgraph counts.
Sparse case: Sharpening the regularity method

- Regularity and counting lemmas aren’t accurate enough to analyze sparse graphs (and unfortunately they’re sharp).

- Existing sparse graph limit theories, such as $L^p$-graphons [Borgs–Chayes–Cohn–Zhao ’14], lack a strong enough counting lemma.

- We are able to establish drastically improved regularity and counting lemmas after cutting out appropriate small “bad” events (involving outlier eigenvalues).
Spectral regularity lemma for random graphs

Write $\mathcal{A}_N = \{0, 1\}^{\binom{N}{2}}$ for the space of adjacency matrices and $\mathcal{X}_N = [0, 1]^{\binom{N}{2}}$ for its convex hull (weighted adjacency matrices).

**Proposition (Quantitative compactness for $\mathcal{A}_N$)**

Let $N \in \mathbb{N}$, $K \geq 1$, $p \in (0, 1)$ with $Np \geq \log N$, and $1 \leq R \leq Np$. There exists a partition $\mathcal{A}_N = \bigsqcup_{j=0}^{J} \mathcal{E}_j$ with the following properties:

(a) $\log J \lesssim RN \log(3 + \frac{R}{Kp})$;

(b) $\mathbb{P}\{A_N, p \in \mathcal{E}_0\} \lesssim \exp(-cK^2N^2p^2)$;

(c) For each $1 \leq j \leq J$, there exists $Y_j \in \mathcal{X}_N$ of rank at most $R$ such that $\|A - Y_j\|_{\text{op}} \lesssim \frac{KNP}{\sqrt{R}}$ for all $A \in \mathcal{E}_j$. 
Proposition (Lipschitz continuity for homomorphism counts)
Let $H = (V, E)$ of max degree $D$.
Let $N \in \mathbb{N}$ and $p \in (0, 1)$. For $K \geq 1$ set

$$
\mathcal{E}_H(K) = \left\{ X \in \mathcal{X}_N : \exists F \leq H \text{ with } \text{hom}(F, X) > KN |V_F| p |E_F| \right\}.
$$

(a) If $N^{-1/D} < p < 1$, then for any $K \geq 2$,

$$
\mathbb{P} \left\{ A_{N, p} \in \mathcal{E}_H(K) \right\} \lesssim_H \exp \left( -c(H)K^{1/|V|}N^2 p^D \right).
$$

(b) For any $X, Y \in \mathcal{X}_N$ with $X \notin \mathcal{E}_H(K)$, for all $F \leq H$,

$$
| \text{hom}(F, X) - \text{hom}(F, Y) | \lesssim_H KN |V_F| p |E_F| \frac{\|X - Y\|_{op}}{Np^D}.
$$
Future directions

- Could possibly push down to $p \gg N^{-1/D}$ with an improved counting lemma (our regularity lemma is essentially optimal).
- To take $p \ll N^{-1/D}$ would require better understanding of the geometry of level sets for subgraph counting functionals (recently accomplished for case of $H$ non-bipartite and $D$-regular by [Harel–Mousset–Samotij ’19]).
- Improved estimates for the partition function of sparse Exponential Random Graphs of various types.
- More general classes of random graphs, e.g. Stochastic Block Model.
- Other applications of new regularity and counting lemmas in random graph theory?

Thanks for your attention!