Large deviations theory for random graphs

Mathematics & Statistics Dept. Colloquium
UNC Greensboro
2023/10/25

Nick Cook, Duke University

Based on joint works with Amir Dembo and Huy Tuan Pham.
In this talk, a graph $G$ is a set of vertices $[n] = \{1, \ldots, n\}$ together with a set $E$ of pairs $\{i, j\} \subset [n]$ called edges.

Write $G_{i,j}$ for the Boolean 0/1 variable that is 1 when $\{i, j\}$ is an edge in $G$.

A random graph is formed by choosing the edge set $E$ in a random way. The $G_{i,j}$ are then $\binom{n}{2}$ (possibly correlated) Bernoulli random variables. In this talk $n$ is large!

Why random graphs?

- Model large networks, statistical estimation for social networks
- Extremal graph theory (probabilistic method of Erdős)
- Mean field models for statistical physics, dynamical systems, constraint satisfaction problems, . . .
Random graph models

**Erdős–Rényi graphs:**

* $G(n, p)$: the edge variables $G_{i,j}$ are independent, $\mathbb{P}(G_{i,j} = 1) = p$.

• $G_{n,m}$: Edge set is uniform random of size $m$.

**Exponential random graph models (ERGMs):** Graph $G$ is chosen with probability proportional to $\exp(H(G))$ for some function ("Hamiltonian") $H$, e.g. the edge-triangle model $H(G) = \alpha e(G) + \beta N_\Delta(G)$, where

$$
e(G) := \sum_{1 \leq i < j \leq n} G_{i,j} = \# \text{ edges}, \quad N_\Delta(G) := \sum_{\{i,j,k\} \subset [n]} G_{i,j}G_{j,k}G_{i,k} = \# \text{ triangles}$$

**Random $d$-regular graphs.** Uniform random under constraint that every vertex has $d$ neighbors. Expanders with high probability.

**Random geometric graphs.** Points $(X_i)_{i=1}^n$ sampled from a distribution/manifold in $\mathbb{R}^d$, connected if sufficiently close.

**Preferential attachment models** (Barabási–Albert). Dynamically generated. Power-law degree distribution, small-world phenomenon.

Source: The Opte Project
Edges and triangles in Erdős–Rényi graphs

Let $G$ random graph from the Erdős–Rényi $G(n, p)$ model. So $\mathbb{P}(G = G) = p^{e(G)}(1 - p)^{\binom{n}{2} - e(G)}$. (Think of $p$ as fixed for now.)

Consider first the random number of edges $e(G) = \sum_{i < j} G_{i,j}$.

- Binomial($\binom{n}{2}, p$) distribution
- Law of averages $\Rightarrow e(G)$ typically $\approx p \binom{n}{2}$.
- Central limit theorem (Laplace): $\frac{e(G) - p \binom{n}{2}}{\sqrt{p(1 - p) \binom{n}{2}}} \Rightarrow$ Normal
- Large deviations (Laplace): for $q \in [0, 1]$, 
  \[
  \log \mathbb{P}(e(G) \sim q \binom{n}{2}) \sim -l_p(q) \binom{n}{2}
  \]
  where $l_p(q) = q \log \frac{q}{p} + (1 - q) \log \frac{1 - q}{1 - p}$.

Now consider the number of triangles $N_\Delta(G) = \sum_{\{i,j,k\} \subset [n]} G_{i,j} G_{j,k} G_{k,\ell}$.

- Cubic polynomial in $\binom{n}{2}$ Bernoulli variables. $\mathbb{E}[N_\Delta(G)] = p^3 \binom{n}{3}$.
- LLN (exercise), CLT (Ruciński ’88)
- Large deviations: The Infamous Upper Tail problem, a driving example for Nonlinear large deviations theory.
  * Other examples: Eigenvalues of random matrices, $k$-term arithmetic progressions in random subsets of $\mathbb{Z}$.
The Infamous Upper Tail (Janson–Ruciński ’02)

**Problem A:** Estimate $\mathbb{P}\{N_\Delta(G) \geq (1 + \delta)\mathbb{E}N_\Delta(G)\}$ for fixed $\delta > 0$.

- Janson–Oleszkiewicz–Ruciński ’04, Kim–Vu ’04
- DeMarco–Kahn ’11, Chatterjee ’11: show
  \[ -\log \mathbb{P}\{N_\Delta(G) \geq (1 + \delta)\mathbb{E}N_\Delta(G)\} \asymp_{\delta} n^2 p^2 \log(1/p) \]

Dependence on $\delta$?

- Dense case ($p$ fixed): Chatterjee–Varadhan, Lubetzky–Zhao ’11 (more on this soon)
- Sparse case: $\mathbb{P}\{N_\Delta(G) \geq (1 + \delta)\mathbb{E}N_\Delta(G)\} = p^{(1+o(1))c(\delta)n^2p^2}$

  where $c(\delta) = \min\{\frac{\delta^2}{3}, \frac{\delta^3}{2}\}$, assuming $n^{-\kappa} \ll p \ll 1$ with

  * $\kappa = \frac{1}{41}$ [Chatterjee–Dembo ’14] + [Lubetzky–Zhao ’14]
  * $\kappa = \frac{1}{18}$ [Eldan ’16]
  * $\kappa = \frac{1}{3}$ [C.–Dembo ’18]
  * $\kappa = \frac{1}{2}$ [Augeri ’18]
  * $\kappa = 1$ [Harel–Mousset–Samotij ’19].

Also results on the upper tail for general $F$-counts with $\kappa = \kappa(F)$, formulas obtained by Bhattacharya–Ganguly–Lubetzky–Zhao ’16.
The Infamous Upper Tail (Janson–Ruciński ’02)

Problem B: Conditional on \( \{N_\Delta(G) \geq (1 + \delta)\mathbb{E}N_\Delta(G)\} \), what does the graph look like? How are the edges distributed?

Three natural guesses:

1. **Boost in edge density**
   to \( q = (1 + \delta)^{1/3}p \)

2. Appearance of a **clique**
of size \( \sim \delta^{1/3}pn \)

3. Appearance of a **hub** (biclique)
of size \( \sim \frac{1}{3}\delta p^2n \).

Sparse case: For \( n^{-1/2} \ll p \ll 1 \), phase transition from **hub** to **clique** as \( \delta \) crosses \( \frac{27}{8} \).
(LZ14, HMS19, C.–Dembo ’22)

Dense case (\( p \) fixed): \( \{N_\Delta(G) \sim q^3\binom{n}{3}\} \) for \( p < q \leq 1 \).
- Large deviation principle (LDP) for the ER graph
  (Chatterjee–Varadhan ’11)
- LDP optimization problem and characterization of the **symmetric regime** (Lubetzky–Zhao ’12)
- Problem B still open in the **symmetry breaking** regime.
A Large Deviation Principle (LDP) for a sequence of random elements $X_n$ of a compact metric space $\mathcal{X}$ says for large $n$ and small $\varepsilon$,

$$\log \mathbb{P}(X_n \in B(x, \varepsilon)) \approx -r_n J(x)$$

for some speed $r_n$ and rate function $J : \mathcal{X} \to \mathbb{R}^+$.\\

How can we view a sequence $G_n$ of Erdős–Rényi graphs on $[n]$ as elements of a single metric space?\\

We can identify any graph $G$ over $[n]$ with a symmetric step function $g(x, y) = G_{\lfloor xn \rfloor, \lfloor yn \rfloor}$ on the unit square $[0, 1]^2$.

Embeds all finite graphs in the space $\mathcal{W}$ of symmetric functions $g : [0, 1]^2 \to [0, 1]$, equipped with a metric induced by the cut norm $\|f\|_\square = \sup_{S, T \subset [0, 1]} |\int_{S \times T} f|$. This is graphon space (Lovász et al. ’06–’10).
Graphon space provides a topological reformulation of the classic regularity method from extremal graph theory.

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

**Theorem (Chatterjee–Varadhan)**

For fixed $p \in (0, 1)$, the sequence of Erdős–Rényi graphs $\{G_n\}_{n \geq 1} \subset \mathcal{W}$ satisfies an LDP of speed $n^2$, with rate function $J(g) = \int_{[0,1]^2} l_p(g(x,y)) \, dx \, dy$.

**Key fact 2:** The triangle-counting function $N_\Delta(\cdot)$ (or more generally the count of any fixed subgraph $F$) extends to a continuous function on $\mathcal{W}$ ($\approx$ the counting lemma).

**Corollary:** Upper tails for subgraph counts (apply the LDP to super-level sets).

**Moral:** The cut-norm topology is the right topology if you’re interested in subgraph counts (for dense graphs at least).
Large deviations for sparse graphs

Problem: for $p = o(1)$, there is 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 are invisible to the cut norm. Related to challenges for developing (useful) sparse graph limit theories.

Quantitative approach: under some norm $\| \cdot \|_*$ on the set $G_n$ of graphs on $[n]$,

- bound covering numbers (compactness)
- bound Lipschitz constants of $F$-count functions $N_F(G)$ (continuity)

C.–Dembo ’18: used the spectral norm (applied to the adjacency matrix), covering $G_n$ with a net of low-rank matrices, together with a tiny “bad” set.

C.–Dembo–Pham ’21: developed generalizations $\| \cdot \|_B$ of the cut-norm to the hypergraph setting. Decomposition of 0/1 tensors as $A = A_{struct} + A_{rand}$, where

- $A_{struct}$ is a short linear combination of “structured” tensors of controlled size under $\| \cdot \|_B$, and
- the pseudorandom remainder $A_{rand}$ is small under the dual norm $\| \cdot \|_B^*$. 
Exponential random graph models (ERGMs)

Recall $G_n$ is the set of graphs over vertex set $[n] = \{1, \ldots, n\}$.

An ERGM is a probability measure on $G_n$ with mass function of the form

$$\frac{1}{Z_n(\alpha, \beta)} e^{n^2 H(G; \beta) - \alpha e(G)}$$

where $\alpha, \beta_1, \ldots, \beta_m \in \mathbb{R}$ are the model parameters, $e(G) = \sum_{1 \leq i < j \leq n} G_{i,j}$, and

$$H(G; \beta) = \sum_{k=1}^m \beta_k f_k(G)$$

for a fixed collection of graph statistics $f_k(G)$. Common choice is the densities of some fixed graphs $F_1, \ldots, F_m$ in $G$.

Ex. 1: Erdős–Rényi distribution. Taking $H \equiv 0$, $\alpha = \log \frac{1-p}{p}$ gives mass function $p^{e(G)} (1-p)^{\binom{n}{2} - e(G)}$.

Ex. 2: Edge-triangle model. $H(G; \beta) = \beta N_\Delta(G)/\binom{n}{3}$
ERGMs: Motivation and challenges

\[
\frac{1}{Z_n(\alpha, \beta)} \exp \left( n^2 \sum_{k=1}^{m} \beta_k f_k(G) - \alpha e(G) \right)
\]

- Introduced in the social sciences literature in the 80s–90s as parametric family of distributions for modeling social networks. [Frank & Strauss '86, Wasserman & Pattison '96].

- Want graphs with transitivity: friends of friends are more likely to be friends.

- The separable form of the Hamiltonian implies that the functions \( e, f_k \) are sufficient statistics for the model parameters \( \alpha, \beta_k \).

- Estimation of model parameters by MLE requires knowledge of the partition function \( Z(\alpha, \beta) \), which is often done by sampling using local MCMC algorithms.
ERGMs: Motivation and challenges

\[
\frac{1}{Z_n(\alpha, \beta)} \exp \left( n^2 \sum_{k=1}^{m} \beta_k f_k(G) - \alpha e(G) \right)
\]

Many problems in practice [Strauss '86, Snijders '02, Handcock '02, '03].

1. No transitivity!
2. **Degeneracy**: typical samples are either nearly empty or nearly full (edge density \(\sim 0\) or \(\sim 1\)).
3. Slow convergence of sampling algorithms in some parameter regimes

Bhamidi–Bressler–Sly '08: characterization of high/low-temperature regimes in “ferromagnetic” case \(\beta_k > 0\) when \(f_k(G)\) are subgraph densities.

- Low-temperature: Exponential convergence time for MCMC
- High-temperature: Polynomial convergence, but typical samples resemble Erdős–Rényi graphs (no transitivity!).

Chatterjee–Diaconis '12 (using Chatterjee–Varadhan '11 LDP) show ferromagnetic ERGMs are \(\approx\) mixtures of Erdős–Rényi graphs.
Consider distributions of the general form \( P(G = G) \propto \exp(r_n, p H(G) - \alpha e(G)) \),

\[
H(G) = h \left( \frac{t(F_1, G)}{p^{e(F_1)}}, \ldots, \frac{t(F_m, G)}{p^{e(F_m)}} \right)
\]

for a fixed continuous, non-decreasing \( h : \mathbb{R}_+^m \rightarrow \mathbb{R} \), and graphs \( F_1, \ldots, F_m \) of max-degree \( d \geq 2 \), where \( t(F_k, G) \) is the density of \( F_k \) in \( G \).

Let \( G_n(a, b) \) be the set of \( G \) with an almost-clique \( I \) and an almost-hub \( J \), for some \( I, J \subset [n] \) of sizes \( |I| \sim \sqrt{ap^{d/2}} n, |J| \sim bp^d n \).

We show under growth and decay conditions on \( h \) and \( p \), with high probability, \( G \in G_n(a, b) \) for some \( (a, b) \) in the set of optimizers for

\[
\sup_{a, b \geq 0} \left\{ h(T_1(a, b), \ldots, T_m(a, b)) - \frac{1}{2} a - b \right\}
\]

for some explicit functions \( T_k \) determined by \( F_k \).
On the 2D manifold of “clique-hub” graphs (up to relabeling vertices), level sets of subgraph-counting functions (green/blue/yellow) and relative entropy (red) are ≈smooth curves.

Upper tail event is light-blue region.

Points \((a, b)\) minimizing the entropy \(\frac{1}{2}a + b\) are circled in red.

Here \(\delta_3 = 100\) and \((\delta_1, \delta_2)\) is A. \((3, 24)\), B. \((4, 25)\), C. \((4, 31.5)\).
Example: (Tamed) Edge-Triangle Model

Let $H(G; \beta) = \beta \left( \frac{t(\Delta, G)}{\rho^3} - 1 \right)^{1/3}$.

1. For fixed $\beta \in (0, \frac{16}{9})$, we have $G \in G_n(0, \frac{1}{3} \beta^{3/2})$ with high prob.

2. For fixed $\beta \in (\frac{16}{9}, \infty)$, we have $G \in G_n(\beta^2, 0)$ with high prob.
Example: Edge-$K_3-P_3$ model

$$H(G; \beta_1, \beta_2) = \beta_1 \left( \frac{t(\triangle_1; G)}{p^3} - 1 \right)^{1/3} + \beta_2 \left( \frac{t(\triangle_2, G)}{p^2} - 1 \right)^{1/3}$$

\[ \frac{16}{9} \Rightarrow \beta_1 \]

\[ \beta_2 \uparrow \]
Directions for the future

- Taming growth of Hamiltonian has “cured” the worst form of degeneracy for ERGMs, but clique-hub graphs still don’t look much like social networks. Might get richer structure from degree constraints, antiferromagnetic models, other statistics $f_k(G), \ldots$

- In C.–Dembo–Pham ’21 we get quantitative LDPs for random hypergraphs, but explicit upper-tail formulas are only known in a few cases, such as clique counts (Liu–Zhao ’19)

- ERHMs?

- LDPs for random regular graphs: Bhattacharya–Dembo ’19, Gunby ’21. LDPs mostly open for:
  * Random geometric graphs (Chatterjee–Harel ’21),
  * Random simplicial complexes (Samorodnitsky–Owada ’22)
Thanks for your attention!