Large deviations for random networks & applications

$G \sim G(n, p)$ a random graph on $n$ vertices where every edge occurs ind. w.p. $p$.

$H = (V(H), E(H))$
- Take $H$ to be $K_3 \Delta$

$X_H$ - no of copies of $H$ in $G$.
$G(X_H) \approx n^3 p^3$

$P \left( \frac{X_H - \mathbb{E}(X_H)}{\mathbb{E}(X_H)} > (1 + \delta) \sigma \right) \leq \delta > 0.$

- Infamous upper tail problem.
  (Janson - Rucinski '02).

Geometric question

What does the random graph $G$ look like given the event $A$.

Fact: $X_H$ is a polynomial of independent Bernoulli variables.
Recall some classical concentration &
large deviations results for linear func-

Azuma–Hoeffding Inequality

\( X_1, X_2, \ldots, X_n \) are independent mean
zero random variables such that

\[ a_i \leq X_i \leq b_i \quad \text{almost surely.} \]

\[ S_n = \sum_{i=1}^{n} X_i \]

\[ P(\mid S_n \mid > t) \]

- strategy is to compute exponential
  moments & then apply Markov.

\[ E(e^{\theta X_i}) \leq e^{\theta^2 (b_i - a_i)^2 / 8} \]

\[ E(e^{\theta S_n}) \leq e^{\theta^2 \sum_{i=1}^{n} c_i^2 / 8} \]

\[ P(S_n > t) \leq e^{\theta^2 \sum_{i=1}^{n} c_i^2 / 8} - \Theta \epsilon \]

- optimize over \( \Theta \).

- \( X_i \) are coin tosses

\( X_i \sim \text{Ber}(p) \quad q > p \)

\[ S_n = np + O(\sqrt{n}) \]
\[ P(S_n > nq) \leq e^{-n I_p(q)} \] (error-free bound).

- Lower bound. - (Tilting)
- Strategy is to do a change of measure which makes the atypical event typical.

To get back to the original measure estimate the R-N derivative between the two measures

\[ P(A) = \int_{A} \log \frac{dP}{dQ} \, dQ \]

Recall \( X_n \) is the number of copies of \( H \) in \( G_n \).
\[ t(H, G) = \frac{1}{n^{\delta(H)}} \sum_{i_1, i_2, \ldots, i_k} \prod_{e \in E(H)} a_{i_j} \delta_{i_j} \]

A graph on \( n \) vertices with adjacency matrix \( A = (a_{ij}) \)

- What does \( G(n, p) \) look like given \( t(H, G) \) is large?

- \( G(n, p) \) continues to look like an Erdős–Rényi-type graph but with different densities (in homogeneous random graph).

It will be convenient to define a metric on graphs \( G \) embed \( H \) in the same space.

**Defn** Let \( \mathcal{W} \) be the set of all

- symmetric functions from \([0,1]^2 \rightarrow [0,1]\)
- Graphons.

Note that any finite graph naturally embeds in \( \mathcal{W} \) as a \( 0,1 \)-valued step function.
For $f, g \in \mathcal{G}$

$$d_\mathcal{G}(f, g) = \sup_{s, t \in \mathcal{G}} |f(s) - g(t)|$$

cut distance.

- Since we don't care about labels of the graphs, we should identify graphs/graphons which are the same up to a relabelling.

6. Measure pres. bijection on $\mathcal{G}$

$f \sim g$ are equivalent if $f, g$ s.t. $d_\mathcal{G}(f, g) = 0$

- We would work with the quotient space $\hat{\mathcal{G}}$.

- \[ G_1 \sim G(n, p) \]

- \[ d(G_1, p) \] is typically small.
A related question is what is the prob the graph looks like.

- Notice that this is exactly the coin tossing problem.

The prob of this, using the same reasoning as the coin tossing is

\[ E\left[ \sum_{(A_1)} I_p(P_1) + \sum_{(A_2)} I_p(P_2) \right] \]

- We want to find the best possible block graph which makes the atypical event typical.

\[ \phi(H, n, p, \delta) = \min \left( \sum_{1 \leq i < j \leq n} I_p(q_{ij}) : t(H, \delta) \geq \text{typ} \right) \]

\[ E_p(t(H, \delta)) \]
\[ Q = (q_{ij}) \]

is the new weighted graph.

- \( \phi(H, r, p, \delta) \) is the best proof one can get by the strategy of considering inhomogeneous random graphs.

- How to prove that this is indeed the optimal strategy.

- Szemerédi's regularity lemma

  - Roughly, this says "any graph can be approximated by a block random graph where the size of blocks is only a function of the error and not the size of the graph."

- (Weak regular lemma - Frieze & Kannan)

  - Given any graph \( G = (V, E) \) there exists a partition of \( V \) into
K classes $A_1, A_2, \ldots, A_K$ such that

$$P_{ij} = \frac{E_{ij}(A_i, A_j)}{|A_i| \cdot |A_j|}$$

$$d_{\Omega}(G, G_p) \leq O\left(\frac{1}{\sqrt{\log k}}\right)$$

One crucial property of the cut dist

- (Counting Lemma)

Fix $H$, and graphs $f, g$

$$|t(H, f) - t(H, g)| \leq O_H(d_\Omega(f, g))$$

- Using the above two facts
  one can compute the probability that $G$ looks like a given block graph and then union bond over all possible choices of block graphs. Consider all possible partition of $V$ into $K$ blocks and all possible edge densities.
$P_{ij}$ up to an error \( e \) with $3e$.

- If the union bd is over a not too big set, the upper bd one gets is $\mathcal{E} \phi(h, \eta, \rho, \delta) + \text{smaller order}$.

- This fails if $p$ is going to zero with a faster than a polylog.

- Full LDP on graphs for a fixed $p$ was proven by Chatterjee & Varadhan (2011).

- The argument above which is more combinatorial. - Lubetzky-Zaho (2015).