Recent Advances in Counting Sparse Graphs

Sourav Chatterjee

The upper tail problem in random graphs

- Let $G(N, p)$ be an Erdős–Rényi random graph. (N vertices, any two connected by an edge with probability p.)

The upper tail problem in random graphs

- Let $G(N, p)$ be an Erdős–Rényi random graph. (N vertices, any two connected by an edge with probability p.)
- Let T be the number of triangles in this graph. What is the behavior of $\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))$ as $N \to \infty$ and δ remains fixed?

The upper tail problem in random graphs

- Let $G(N, p)$ be an Erdős–Rényi random graph. (N vertices, any two connected by an edge with probability p.)
- Let T be the number of triangles in this graph. What is the behavior of $\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))$ as $N \to \infty$ and δ remains fixed?
- Here p may remain fixed or may tend to 0 as $N \to \infty$.

The upper tail problem in random graphs

- Let $G(N, p)$ be an Erdős–Rényi random graph. (N vertices, any two connected by an edge with probability p.)
- Let T be the number of triangles in this graph. What is the behavior of $\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))$ as $N \to \infty$ and δ remains fixed?
- Here p may remain fixed or may tend to 0 as $N \to \infty$.
- Was an open question for a long time. (History in the next two slides.) Still not fully resolved. This deceptively difficult problem falls in the intersection of large deviations, concentration of measure, and random graph theory.

The upper tail problem in random graphs

- Let $G(N, p)$ be an Erdős–Rényi random graph. (N vertices, any two connected by an edge with probability p.)
- Let T be the number of triangles in this graph. What is the behavior of $\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T))$ as $N \to \infty$ and δ remains fixed?
- Here p may remain fixed or may tend to 0 as $N \to \infty$.
- Was an open question for a long time. (History in the next two slides.) Still not fully resolved. This deceptively difficult problem falls in the intersection of large deviations, concentration of measure, and random graph theory.
- This is quite surprising since T is simply a third degree polynomial of independent random variables. We know everything about linear functions of independent random variables, so why all this difficulty for a third degree polynomial?

Upper tails for triangles

- After a long line of successively improving tail bounds by various authors, Kim and Vu (2004) and Janson, Oleszkiewicz and Ruciński (2004) showed that if $p \geq N^{-1} \log N$, then

$$e^{-c_1(\delta) N^2 p^2 \log(1/p)} \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq e^{-c_2(\delta) N^2 p^2},$$

where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend on δ only.

Upper tails for triangles

- After a long line of successively improving tail bounds by various authors, Kim and Vu (2004) and Janson, Oleszkiewicz and Ruciński (2004) showed that if $p \geq N^{-1} \log N$, then

$$e^{-c_1(\delta) N^2 p^2 \log(1/p)} \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq e^{-c_2(\delta) N^2 p^2},$$

  where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend on δ only.
- The logarithmic gap was closed by Chatterjee (2012) and DeMarco and Kahn (2012).

Upper tails for triangles

- After a long line of successively improving tail bounds by various authors, Kim and Vu (2004) and Janson, Oleszkiewicz and Ruciński (2004) showed that if $p \geq N^{-1} \log N$, then

$$e^{-c_1(\delta) N^2 p^2 \log(1/p)} \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq e^{-c_2(\delta) N^2 p^2},$$

  where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend on δ only.
- The logarithmic gap was closed by Chatterjee (2012) and DeMarco and Kahn (2012).
- The computation of the exact constant in the exponent was still open.

Upper tails for triangles

- After a long line of successively improving tail bounds by various authors, Kim and Vu (2004) and Janson, Oleszkiewicz and Ruciński (2004) showed that if $p \geq N^{-1} \log N$, then
$$e^{-c_1(\delta)N^2p^2\log(1/p)} \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq e^{-c_2(\delta)N^2p^2},$$
where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend on δ only.
- The logarithmic gap was closed by Chatterjee (2012) and DeMarco and Kahn (2012).
- The computation of the exact constant in the exponent was still open.
- For a certain range of fixed values of p and δ, Chatterjee and Dey (2009) — using Stein's method for concentration inequalities — computed the exact constant $c(\delta, p)$ such that, as $N \to \infty$,
$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) = e^{-c(\delta,p)N^2(1+o(1))}.$$

Upper tails for triangles

- After a long line of successively improving tail bounds by various authors, Kim and Vu (2004) and Janson, Oleszkiewicz and Ruciński (2004) showed that if $p \geq N^{-1} \log N$, then

$$e^{-c_1(\delta)N^2p^2\log(1/p)} \leq \mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \leq e^{-c_2(\delta)N^2p^2},$$

  where $c_1(\delta)$ and $c_2(\delta)$ are constants that depend on δ only.
- The logarithmic gap was closed by Chatterjee (2012) and DeMarco and Kahn (2012).
- The computation of the exact constant in the exponent was still open.
- For a certain range of fixed values of p and δ, Chatterjee and Dey (2009) — using Stein's method for concentration inequalities — computed the exact constant $c(\delta, p)$ such that, as $N \to \infty$,

$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) = e^{-c(\delta,p)N^2(1+o(1))}.$$

- This result, however, did not cover the full range of δ and p.

Large deviations for triangle counts

- For the case of fixed p, the problem was fully resolved by Chatterjee and Varadhan (2011), who exhibited the constant $c(\delta, p)$ as the solution of a variational problem. This formula was conjectured in an unpublished manuscript of Bolthausen, Comets and Dembo (2003).

Large deviations for triangle counts

- For the case of fixed p, the problem was fully resolved by Chatterjee and Varadhan (2011), who exhibited the constant $c(\delta, p)$ as the solution of a variational problem. This formula was conjectured in an unpublished manuscript of Bolthausen, Comets and Dembo (2003).
- The proof relied on Szemerédi's regularity lemma and the theory of graph limits developed by Lovász and coauthors.

Large deviations for triangle counts

- For the case of fixed p, the problem was fully resolved by Chatterjee and Varadhan (2011), who exhibited the constant $c(\delta, p)$ as the solution of a variational problem. This formula was conjectured in an unpublished manuscript of Bolthausen, Comets and Dembo (2003).
- The proof relied on Szemerédi's regularity lemma and the theory of graph limits developed by Lovász and coauthors.
- However, the regularity lemma has no satisfactory analog for sparse graphs, certainly when p is of order $N^{-\alpha}$ for some positive α.

Large deviations for triangle counts

- For the case of fixed p, the problem was fully resolved by Chatterjee and Varadhan (2011), who exhibited the constant $c(\delta, p)$ as the solution of a variational problem. This formula was conjectured in an unpublished manuscript of Bolthausen, Comets and Dembo (2003).
- The proof relied on Szemerédi's regularity lemma and the theory of graph limits developed by Lovász and coauthors.
- However, the regularity lemma has no satisfactory analog for sparse graphs, certainly when p is of order $N^{-\alpha}$ for some positive α.
- This made it impossible to extend the proof to the sparse case.

Partial solution

Theorem (Lubetzky and Zhao (2014))

If $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}$, then

$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) = \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\}N^2p^2\log\frac{1}{p}\right).$$

Partial solution

Theorem (Lubetzky and Zhao (2014))

If $N \to \infty$ *and* $p \to 0$ *slower than* $N^{-1/42}$*, then*

$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) \\ = \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\}N^2 p^2 \log\frac{1}{p}\right).$$

- Conjecture: This formula is valid if $p \to 0$ slower than $N^{-1/2}$.

Partial solution

Theorem (Lubetzky and Zhao (2014))

If $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}$, then

$$\mathbb{P}(T \geq (1+\delta)\mathbb{E}(T)) = \exp\left(-(1+o(1))\min\left\{\frac{\delta^{2/3}}{2}, \frac{\delta}{3}\right\} N^2 p^2 \log \frac{1}{p}\right).$$

- Conjecture: This formula is valid if $p \to 0$ slower than $N^{-1/2}$.
- The proof uses the theory of nonlinear large deviations developed by Chatterjee and Dembo (2014). This will be described later in this talk.

Generalization

- Take any connected finite simple graph H on k vertices with maximum degree $\Delta \geq 2$.

Generalization

- Take any connected finite simple graph H on k vertices with maximum degree $\Delta \geq 2$.
- Let H^* be the induced subgraph of H on all vertices whose degree in H is Δ.

Generalization

- Take any connected finite simple graph H on k vertices with maximum degree $\Delta \geq 2$.
- Let H^* be the induced subgraph of H on all vertices whose degree in H is Δ.
- Define a polynomial $P_{H^*}(x) := \sum_k i_{H^*}(k)x^k$, where $i_{H^*}(k)$ is the number of k-element independent sets in H^*.

Generalization

- Take any connected finite simple graph H on k vertices with maximum degree $\Delta \geq 2$.
- Let H^* be the induced subgraph of H on all vertices whose degree in H is Δ.
- Define a polynomial $P_{H^*}(x) := \sum_k i_{H^*}(k)x^k$, where $i_{H^*}(k)$ is the number of k-element independent sets in H^*.
- Let $H_{N,p}$ be the number of homomorphisms of H into $G(N,p)$.

Generalization

- Take any connected finite simple graph H on k vertices with maximum degree $\Delta \geq 2$.
- Let H^* be the induced subgraph of H on all vertices whose degree in H is Δ.
- Define a polynomial $P_{H^*}(x) := \sum_k i_{H^*}(k)x^k$, where $i_{H^*}(k)$ is the number of k-element independent sets in H^*.
- Let $H_{N,p}$ be the number of homomorphisms of H into $G(N, p)$.
- A homomorphism is a map from the vertex set of H into the vertex set of $G(N, p)$ that preserves edges. For example, if H is a triangle, then $H_{N,p}$ is six times the number of triangles in $G(N, p)$.

Large deviations for subgraph counts

Theorem (Bhattacharya, Ganguly, Lubetzky and Zhao (2015))

For any $\delta > 0$*, there is a unique positive number* $\theta = \theta(H, \delta)$ *that solves* $P_{H^*}(\theta) = 1 + \delta$*. There is a constant* $\alpha_H > 0$ *depending only on* H*, such that if* $N \to \infty$ *and* $p \to 0$ *slower than* $N^{-\alpha_H}$*,*

$$\mathbb{P}(H_{N,p} \geq (1+\delta)\mathbb{E}(H_{N,p})) = \exp\left(-(1+o(1))c(\delta)N^2 p^\Delta \log \frac{1}{p}\right),$$

where

$$c(\delta) = \begin{cases} \min\{\theta, \frac{1}{2}\delta^{2/k}\} & \textit{if } H \textit{ is regular,} \\ \theta & \textit{if } H \textit{ is irregular.} \end{cases}$$

Large deviations for subgraph counts

Theorem (Bhattacharya, Ganguly, Lubetzky and Zhao (2015))

For any $\delta > 0$, there is a unique positive number $\theta = \theta(H, \delta)$ that solves $P_{H^*}(\theta) = 1 + \delta$. There is a constant $\alpha_H > 0$ depending only on H, such that if $N \to \infty$ and $p \to 0$ slower than $N^{-\alpha_H}$,

$$\mathbb{P}(H_{N,p} \geq (1+\delta)\mathbb{E}(H_{N,p})) = \exp\left(-(1+o(1))c(\delta)N^2 p^\Delta \log \frac{1}{p}\right),$$

where

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

Example: If $H = C_4$, then

$$c(\delta) = \begin{cases} \frac{1}{2}\sqrt{\delta} & \text{if } \delta < 16, \\ -1 + \sqrt{1 + \frac{1}{2}\delta} & \text{if } \delta \geq 16. \end{cases}$$

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.
- Let $Y = (Y_1, \ldots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables.

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.
- Let $Y = (Y_1, \ldots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables.
- We want to find an approximation for the upper tail probability $\mathbb{P}(f(Y) \geq t)$ when t is much bigger than $\mathbb{E}(f(Y))$.

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.
- Let $Y = (Y_1, \ldots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables.
- We want to find an approximation for the upper tail probability $\mathbb{P}(f(Y) \geq t)$ when t is much bigger than $\mathbb{E}(f(Y))$.
- Classical large deviations theory well-suited for linear f.

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.
- Let $Y = (Y_1, \ldots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables.
- We want to find an approximation for the upper tail probability $\mathbb{P}(f(Y) \geq t)$ when t is much bigger than $\mathbb{E}(f(Y))$.
- Classical large deviations theory well-suited for linear f.
- May be quite nontrivial even for very simple nonlinear f, as we saw in the random graph example. Problems tackled on ad hoc basis.

A general problem

- Take any smooth $f : [0,1]^n \to \mathbb{R}$.
- Let $Y = (Y_1, \ldots, Y_n)$ be a vector of i.i.d. *Bernoulli*(p) random variables.
- We want to find an approximation for the upper tail probability $\mathbb{P}(f(Y) \geq t)$ when t is much bigger than $\mathbb{E}(f(Y))$.
- Classical large deviations theory well-suited for linear f.
- May be quite nontrivial even for very simple nonlinear f, as we saw in the random graph example. Problems tackled on ad hoc basis.
- Can there be a more unified approach?

The naive mean field approximation

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, define

$$I_p(x) := \sum_{i=1}^n \left(x_i \log \frac{x_i}{p} + (1 - x_i) \log \frac{1 - x_i}{1 - p} \right).$$

The naive mean field approximation

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, define

$$I_p(x) := \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{p} + (1 - x_i) \log \frac{1 - x_i}{1 - p} \right).$$

- For each $t \in \mathbb{R}$, define

$$\phi_p(t) := \inf\{I_p(x) \, : \, x \in [0,1]^n \text{ such that } f(x) \geq tn\}.$$

The naive mean field approximation

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, define

$$I_p(x) := \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{p} + (1 - x_i) \log \frac{1 - x_i}{1 - p} \right).$$

- For each $t \in \mathbb{R}$, define

$$\phi_p(t) := \inf\{I_p(x) \,:\, x \in [0,1]^n \text{ such that } f(x) \geq tn\}.$$

- In many problems, it turns out that

$$\mathbb{P}(f(Y) \geq tn) \approx \exp(-\phi_p(t)). \qquad (\star)$$

The naive mean field approximation

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, define

$$I_p(x) := \sum_{i=1}^{n} \left(x_i \log \frac{x_i}{p} + (1 - x_i) \log \frac{1 - x_i}{1 - p} \right).$$

- For each $t \in \mathbb{R}$, define

$$\phi_p(t) := \inf\{I_p(x) \,:\, x \in [0,1]^n \text{ such that } f(x) \geq tn\}.$$

- In many problems, it turns out that

$$\mathbb{P}(f(Y) \geq tn) \approx \exp(-\phi_p(t)). \qquad (\star)$$

- In particular, this is true in great generality for linear functions.

The naive mean field approximation

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, define

$$I_p(x) := \sum_{i=1}^n \left(x_i \log \frac{x_i}{p} + (1 - x_i) \log \frac{1 - x_i}{1 - p} \right).$$

- For each $t \in \mathbb{R}$, define

$$\phi_p(t) := \inf\{ I_p(x) \,:\, x \in [0,1]^n \text{ such that } f(x) \geq tn \}.$$

- In many problems, it turns out that

$$\mathbb{P}(f(Y) \geq tn) \approx \exp(-\phi_p(t)). \qquad (\star)$$

- In particular, this is true in great generality for linear functions.
- I will describe a sufficient condition under which the above approximation is valid for nonlinear maps.

Low complexity gradient condition

Theorem (Chatterjee & Dembo, 2014. Rough statement.)

The approximation ($\star$) *is valid when, in addition to some smoothness conditions on the function* f*, the gradient vector* $\nabla f(x) = (\partial f/\partial x_1, \ldots \partial f/\partial x_n)$ *may be approximately encoded by* $o(n)$ *bits of information.*

Low complexity gradient condition

Theorem (Chatterjee & Dembo, 2014. Rough statement.)

The approximation ($\star$) *is valid when, in addition to some smoothness conditions on the function* f*, the gradient vector* $\nabla f(x) = (\partial f/\partial x_1, \ldots \partial f/\partial x_n)$ *may be approximately encoded by* $o(n)$ *bits of information.*

- We call this the "low complexity gradient" condition.

Low complexity gradient condition

Theorem (Chatterjee & Dembo, 2014. Rough statement.)

The approximation ($\star$) *is valid when, in addition to some smoothness conditions on the function* f*, the gradient vector* $\nabla f(x) = (\partial f/\partial x_1, \ldots \partial f/\partial x_n)$ *may be approximately encoded by* $o(n)$ *bits of information.*

- We call this the "low complexity gradient" condition.
- Actual statement of the theorem involves a messy error term arising out of the smoothness conditions on f.

Low complexity gradient condition

Theorem (Chatterjee & Dembo, 2014. Rough statement.)

The approximation ($\star$) *is valid when, in addition to some smoothness conditions on the function* f*, the gradient vector* $\nabla f(x) = (\partial f/\partial x_1, \ldots \partial f/\partial x_n)$ *may be approximately encoded by* $o(n)$ *bits of information.*

- We call this the "low complexity gradient" condition.
- Actual statement of the theorem involves a messy error term arising out of the smoothness conditions on f.
- Many notable results on sharp upper and lower bounds for tail probabilities of nonlinear functions (Talagrand, Kim, Vu, Latała, ...) that hold up to constant factors in the exponent, but no results about the precise approximation ($\star$).

Low complexity gradient condition

Theorem (Chatterjee & Dembo, 2014. Rough statement.)

The approximation ($\star$) *is valid when, in addition to some smoothness conditions on the function* f*, the gradient vector* $\nabla f(x) = (\partial f/\partial x_1, \ldots \partial f/\partial x_n)$ *may be approximately encoded by* $o(n)$ *bits of information.*

- We call this the "low complexity gradient" condition.
- Actual statement of the theorem involves a messy error term arising out of the smoothness conditions on f.
- Many notable results on sharp upper and lower bounds for tail probabilities of nonlinear functions (Talagrand, Kim, Vu, Latała, ...) that hold up to constant factors in the exponent, but no results about the precise approximation ($\star$).
- Some preliminary work in Chatterjee & Dey (2009).

Example 1: 1D Ising model

- Let

$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1} \, .$$

Example 1: 1D Ising model

- Let
$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1}\,.$$
- Then, for $2 \le i \le n-1$,
$$\frac{\partial f}{\partial x_i} = x_{i-1} + x_{i+1}\,.$$

Example 1: 1D Ising model

- Let
$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1}\,.$$
- Then, for $2 \le i \le n-1$,
$$\frac{\partial f}{\partial x_i} = x_{i-1} + x_{i+1}\,.$$
- Thus, for this f, the gradient vector cannot be approximately encoded by $o(n)$ many bits.

Example 1: 1D Ising model

- Let
$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1} \,.$$
- Then, for $2 \le i \le n-1$,
$$\frac{\partial f}{\partial x_i} = x_{i-1} + x_{i+1} \,.$$
- Thus, for this f, the gradient vector cannot be approximately encoded by $o(n)$ many bits. To know $\nabla f(x)$, even approximately, we need to know the values of all the x_i's.

Example 1: 1D Ising model

- Let
$$f(x) = \sum_{i=1}^{n-1} x_i x_{i+1} \,.$$
- Then, for $2 \le i \le n-1$,
$$\frac{\partial f}{\partial x_i} = x_{i-1} + x_{i+1} \,.$$
- Thus, for this f, the gradient vector cannot be approximately encoded by $o(n)$ many bits. To know $\nabla f(x)$, even approximately, we need to know the values of all the x_i's.
- One can check that the approximation $(\star)$ is not valid for this f.

Example 2: Curie-Weiss model

- Let
$$f(x) = \frac{1}{n} \sum_{1\le i<j\le n} x_i x_j\,.$$

Example 2: Curie-Weiss model

- Let
$$f(x) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j \,.$$
- For each i,
$$\frac{\partial f}{\partial x_i} = \frac{1}{n} \sum_{j \ne i} x_j = -\frac{x_i}{n} + \frac{1}{n} \sum_{j=1}^{n} x_j \,.$$

Example 2: Curie-Weiss model

- Let
$$f(x) = \frac{1}{n} \sum_{1\le i<j\le n} x_i x_j\,.$$
- For each i,
$$\frac{\partial f}{\partial x_i} = \frac{1}{n}\sum_{j\ne i} x_j = -\frac{x_i}{n} + \frac{1}{n}\sum_{j=1}^n x_j\,.$$
- Thus, for this f, the gradient vector is approximately encoded by the single quantity $n^{-1}\sum x_j$.

Example 2: Curie-Weiss model

- Let
$$f(x) = \frac{1}{n} \sum_{1 \le i < j \le n} x_i x_j \,.$$
- For each i,
$$\frac{\partial f}{\partial x_i} = \frac{1}{n} \sum_{j \ne i} x_j = -\frac{x_i}{n} + \frac{1}{n} \sum_{j=1}^{n} x_j \,.$$
- Thus, for this f, the gradient vector is approximately encoded by the single quantity $n^{-1} \sum x_j$.
- The large deviation probabilities for this function satisfy the approximation ($\star$).

Example 3: Subgraph counts in sparse random graphs

- Let T be the number of triangles in an Erdős-Rényi random graph $G(N, p)$.

Example 3: Subgraph counts in sparse random graphs

- Let T be the number of triangles in an Erdős-Rényi random graph $G(N, p)$.
- Then
$$T = \frac{1}{6} \sum_{i,j,k} Y_{ij} Y_{jk} Y_{ki} ,$$
where Y_{ij} is the indicator that edge $\{i, j\}$ is present in the graph.

Example 3: Subgraph counts in sparse random graphs

- Let T be the number of triangles in an Erdős-Rényi random graph $G(N, p)$.
- Then
$$T = \frac{1}{6} \sum_{i,j,k} Y_{ij} Y_{jk} Y_{ki} ,$$
where Y_{ij} is the indicator that edge $\{i, j\}$ is present in the graph.
- Let $n = N(N-1)/2$ and let us agree to denote elements of $\mathbb{R}^n$ as $x = (x_{ij})_{1 \leq i < j \leq N}$, with the convention that $x_{ii} = 0$ and $x_{ji} = x_{ij}$. Define a function $f : \mathbb{R}^n \to \mathbb{R}$ as
$$f(x) = \frac{1}{N} \sum_{i,j,k=1}^{N} x_{ij} x_{jk} x_{ki} .$$

Example 3: Subgraph counts in sparse random graphs

- Let T be the number of triangles in an Erdős-Rényi random graph $G(N, p)$.
- Then
$$T = \frac{1}{6} \sum_{i,j,k} Y_{ij} Y_{jk} Y_{ki} ,$$
where Y_{ij} is the indicator that edge $\{i, j\}$ is present in the graph.
- Let $n = N(N-1)/2$ and let us agree to denote elements of $\mathbb{R}^n$ as $x = (x_{ij})_{1 \le i < j \le N}$, with the convention that $x_{ii} = 0$ and $x_{ji} = x_{ij}$. Define a function $f : \mathbb{R}^n \to \mathbb{R}$ as
$$f(x) = \frac{1}{N} \sum_{i,j,k=1}^{N} x_{ij} x_{jk} x_{ki} .$$
- The plan is to apply the main theorem to this f.

Example 3 contd.

- Note that
$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N}\sum_{k=1}^{N} x_{ik}x_{jk} =: 3a_{ij}(x)\,.$$

Example 3 contd.

- Note that
$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N}\sum_{k=1}^{N} x_{ik}x_{jk} =: 3a_{ij}(x)\,.$$
- To apply the main theorem, we need to show that $a_{ij}(x)$'s may be approximately encoded by $o(N^2)$ bits.

Example 3 contd.

- Note that

$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N}\sum_{k=1}^{N} x_{ik}x_{jk} =: 3a_{ij}(x)\,.$$

- To apply the main theorem, we need to show that $a_{ij}(x)$'s may be approximately encoded by $o(N^2)$ bits.
- Key observations: (1) If two symmetric matrices $x = (x_{ij})$ and $y = (y_{ij})$ are close in operator norm, then $a_{ij}(x) \approx a_{ij}(y)$ for almost all i, j.

Example 3 contd.

- Note that
$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N}\sum_{k=1}^{N} x_{ik}x_{jk} =: 3a_{ij}(x)\,.$$
- To apply the main theorem, we need to show that $a_{ij}(x)$'s may be approximately encoded by $o(N^2)$ bits.
- Key observations: (1) If two symmetric matrices $x = (x_{ij})$ and $y = (y_{ij})$ are close in operator norm, then $a_{ij}(x) \approx a_{ij}(y)$ for almost all i, j. (2) The space of matrices with entries in $[0, 1]$ has low entropy in operator norm, that is $e^{o(N^2)}$ matrices can well-approximate all matrices in operator norm.

Example 3 contd.

- Note that
$$\frac{\partial f}{\partial x_{ij}} = \frac{3}{N}\sum_{k=1}^{N} x_{ik}x_{jk} =: 3a_{ij}(x)\,.$$
- To apply the main theorem, we need to show that $a_{ij}(x)$'s may be approximately encoded by $o(N^2)$ bits.
- Key observations: (1) If two symmetric matrices $x = (x_{ij})$ and $y = (y_{ij})$ are close in operator norm, then $a_{ij}(x) \approx a_{ij}(y)$ for almost all i, j. (2) The space of matrices with entries in $[0, 1]$ has low entropy in operator norm, that is $e^{o(N^2)}$ matrices can well-approximate all matrices in operator norm.
- These two observations allow us to prove the low complexity condition for ∇f.

Example 3 contd.

- What is the precise result for triangle counts?

Example 3 contd.

- What is the precise result for triangle counts?
- For $x = (x_{ij})_{1 \le i < j \le N}$, define

$$I_p(x) := \sum_{1 \le i < j \le N} \left(x_{ij} \log \frac{x_{ij}}{p} + (1 - x_{ij}) \log \frac{1 - x_{ij}}{1 - p} \right)$$

Example 3 contd.

- What is the precise result for triangle counts?
- For $x = (x_{ij})_{1\le i<j\le N}$, define

$$I_p(x) := \sum_{1\le i<j\le N} \left(x_{ij} \log \frac{x_{ij}}{p} + (1-x_{ij}) \log \frac{1-x_{ij}}{1-p} \right)$$

and

$$T(x) := \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{ki} \,,$$

where $x_{ij} \in [0,1]$ and $x_{ji} = x_{ij}$, $x_{ii} = 0$.

Example 3 contd.

- What is the precise result for triangle counts?
- For $x = (x_{ij})_{1 \le i < j \le N}$, define

$$I_p(x) := \sum_{1 \le i < j \le N} \left(x_{ij} \log \frac{x_{ij}}{p} + (1 - x_{ij}) \log \frac{1 - x_{ij}}{1 - p} \right)$$

and

$$T(x) := \frac{1}{6} \sum_{i,j,k} x_{ij} x_{jk} x_{ki} \, ,$$

where $x_{ij} \in [0, 1]$ and $x_{ji} = x_{ij}$, $x_{ii} = 0$.
- For $u > 1$ define

$$\psi_p(u) := \inf\{ I_p(x) : T(x) \ge u \, \mathbb{E}(T) \} \, ,$$

where T is the number of triangles in $G(N, p)$.

Example 3 contd.

Theorem (Chatterjee & Dembo, 2014.)

For $u > 1$ and N sufficiently large (depending only on u),

$$1 - \frac{c \log N}{N^{1/6} p^2} \leq \frac{\psi_p(u)}{-\log \mathbb{P}(T \geq u\,\mathbb{E}(T))} \leq 1 + \frac{C(\log N)^{33/29}}{N^{1/29} p^{42/29}},$$

where c and C are constants that depend only on u.

Example 3 contd.

Theorem (Chatterjee & Dembo, 2014.)

For $u > 1$ and N sufficiently large (depending only on u),

$$1 - \frac{c \log N}{N^{1/6} p^2} \leq \frac{\psi_p(u)}{-\log \mathbb{P}(T \geq u\,\mathbb{E}(T))} \leq 1 + \frac{C(\log N)^{33/29}}{N^{1/29} p^{42/29}},$$

where c and C are constants that depend only on u.

- In particular,

$$\frac{\psi_p(u)}{-\log \mathbb{P}(T \geq u\,\mathbb{E}(T))} \to 1$$

  if $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}(\log N)^{11/14}$.

Example 3 contd.

Theorem (Chatterjee & Dembo, 2014.)

For $u > 1$ and N sufficiently large (depending only on u),

$$1 - \frac{c \log N}{N^{1/6} p^2} \leq \frac{\psi_p(u)}{-\log \mathbb{P}(T \geq u\,\mathbb{E}(T))} \leq 1 + \frac{C(\log N)^{33/29}}{N^{1/29} p^{42/29}},$$

where c and C are constants that depend only on u.

- In particular,
$$\frac{\psi_p(u)}{-\log \mathbb{P}(T \geq u\,\mathbb{E}(T))} \to 1$$
if $N \to \infty$ and $p \to 0$ slower than $N^{-1/42}(\log N)^{11/14}$.
- Lubetzky and Zhao (2014) studied the function $\psi_p(u)$ to obtain their large deviation result.

Nonlinear large deviations: The main result

- For any $f : [0,1]^n \to \mathbb{R}$, let $\|f\|$ denote the supremum norm of f.

Nonlinear large deviations: The main result

- For any $f : [0,1]^n \to \mathbb{R}$, let $\|f\|$ denote the supremum norm of f.
- Let
$$f_i := \frac{\partial f}{\partial x_i} \quad \text{and} \quad f_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$

Nonlinear large deviations: The main result

- For any $f : [0,1]^n \to \mathbb{R}$, let $\|f\|$ denote the supremum norm of f.
- Let
$$f_i := \frac{\partial f}{\partial x_i} \quad \text{and} \quad f_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$
- Define
$$a := \|f\|, \quad b_i := \|f_i\| \quad \text{and} \quad c_{ij} := \|f_{ij}\|.$$

Nonlinear large deviations: The main result

- For any $f : [0,1]^n \to \mathbb{R}$, let $\|f\|$ denote the supremum norm of f.
- Let
$$f_i := \frac{\partial f}{\partial x_i} \quad \text{and} \quad f_{ij} := \frac{\partial^2 f}{\partial x_i \partial x_j}.$$
- Define
$$a := \|f\|, \quad b_i := \|f_i\| \quad \text{and} \quad c_{ij} := \|f_{ij}\|.$$
- Given $\epsilon > 0$, let $\mathcal{D}(\epsilon)$ be a finite subset of $\mathbb{R}^n$ such that for all $x \in \{0,1\}^n$, there exists $d = (d_1, \ldots, d_n) \in \mathcal{D}(\epsilon)$ such that
$$\sum_{i=1}^n (f_i(x) - d_i)^2 \leq n\epsilon^2.$$

Main result, contd.

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, let

$$I(x) := \sum_{i=1}^n (x_i \log x_i + (1 - x_i) \log(1 - x_i)) \,.$$

Main result, contd.

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, let

$$I(x) := \sum_{i=1}^{n} (x_i \log x_i + (1 - x_i) \log(1 - x_i)) .$$

- Let

$$F := \log \sum_{x \in \{0,1\}^n} e^{f(x)} .$$

Main result, contd.

- For $x = (x_1, \ldots, x_n) \in [0,1]^n$, let
$$I(x) := \sum_{i=1}^{n} (x_i \log x_i + (1-x_i)\log(1-x_i)).$$
- Let
$$F := \log \sum_{x\in\{0,1\}^n} e^{f(x)}.$$
- Given $\epsilon > 0$, define
$$\text{complexity term} := \frac{1}{4}\left(n\sum_{i=1}^{n} b_i^2\right)^{1/2}\epsilon + 3n\epsilon + \log|\mathcal{D}(\epsilon)|, \quad \text{and}$$
$$\text{smoothness term} := 4\left(\sum_{i=1}^{n}(ac_{ii} + b_i^2) + \frac{1}{4}\sum_{i,j=1}^{n}\left(ac_{ij}^2 + b_ib_jc_{ij} + 4b_ic_{ij}\right)\right)^{1/2}$$
$$+ \frac{1}{4}\left(\sum_{i=1}^{n} b_i^2\right)^{1/2}\left(\sum_{i=1}^{n} c_{ii}^2\right)^{1/2} + 3\sum_{i=1}^{n} c_{ii} + \log 2.$$

Statement of the main result

Theorem (Chatterjee & Dembo, 2014)

For any $\epsilon > 0$,

$$F \le \sup_{x \in [0,1]^n} (f(x) - I(x)) + \text{complexity term} + \text{smoothness term}\,,$$

and

$$F \ge \sup_{x \in [0,1]^n} (f(x) - I(x)) - \frac{1}{2}\sum_{i=1}^{n} c_{ii}\,.$$

Proof sketch

- Let $X = (X_1, \ldots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0, 1\}^n$ with respect to the counting measure.

Proof sketch

- Let $X = (X_1, \ldots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0,1\}^n$ with respect to the counting measure.
- For each i, define a function $\hat{x}_i : [0,1]^n \to [0,1]$ as

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j,\, 1 \le j \le n,\ j \ne i).$$

Proof sketch

- Let $X = (X_1, \ldots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0,1\}^n$ with respect to the counting measure.
- For each i, define a function $\hat{x}_i : [0,1]^n \to [0,1]$ as

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j,\, 1 \le j \le n,\ j \neq i).$$

- Let $\hat{x} : [0,1]^n \to [0,1]^n$ be the vector-valued function whose ith coordinate function is $\hat{x}_i$.

Proof sketch

- Let $X = (X_1, \ldots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0,1\}^n$ with respect to the counting measure.
- For each i, define a function $\hat{x}_i : [0,1]^n \to [0,1]$ as

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j,\, 1 \le j \le n,\ j \neq i).$$

- Let $\hat{x} : [0,1]^n \to [0,1]^n$ be the vector-valued function whose ith coordinate function is $\hat{x}_i$.
- Let $\hat{X} = \hat{x}(X)$.

Proof sketch

- Let $X = (X_1, \ldots, X_n)$ be a random vector that has probability density proportional to $e^{f(x)}$ on $\{0,1\}^n$ with respect to the counting measure.
- For each i, define a function $\hat{x}_i : [0,1]^n \to [0,1]$ as

$$\hat{x}_i(x) = \mathbb{E}(X_i \mid X_j = x_j,\, 1 \le j \le n,\ j \ne i).$$

- Let $\hat{x} : [0,1]^n \to [0,1]^n$ be the vector-valued function whose ith coordinate function is $\hat{x}_i$.
- Let $\hat{X} = \hat{x}(X)$.
- The first step in the proof is to show that if the smoothness term is small, then

$$f(X) \approx f(\hat{X}) \ \text{ with high probability.}$$

Proof sketch contd.

- Next, define a function $g : [0,1]^n \times [0,1]^n \to \mathbb{R}$ as

$$g(x,y) := \sum_{i=1}^n (x_i \log y_i + (1-x_i)\log(1-y_i)) .$$

Proof sketch contd.

- Next, define a function $g : [0,1]^n \times [0,1]^n \to \mathbb{R}$ as

$$g(x, y) := \sum_{i=1}^{n} (x_i \log y_i + (1 - x_i) \log(1 - y_i)) .$$

- The second step is to show that with high probability,

$$g(X, \hat{X}) \approx g(\hat{X}, \hat{X}) = I(\hat{X}) .$$

Proof sketch contd.

- Next, define a function $g : [0,1]^n \times [0,1]^n \to \mathbb{R}$ as

$$g(x,y) := \sum_{i=1}^{n} (x_i \log y_i + (1-x_i)\log(1-y_i))\,.$$

- The second step is to show that with high probability,

$$g(X,\hat{X}) \approx g(\hat{X},\hat{X}) = I(\hat{X})\,.$$

- Suppose that these two steps have been proved.

Proof sketch contd.

- Next, define a function $g : [0,1]^n \times [0,1]^n \to \mathbb{R}$ as

$$g(x,y) := \sum_{i=1}^{n} (x_i \log y_i + (1-x_i)\log(1-y_i)) \,.$$

- The second step is to show that with high probability,

$$g(X,\hat{X}) \approx g(\hat{X},\hat{X}) = I(\hat{X}) \,.$$

- Suppose that these two steps have been proved.
- Let A be the set of all x where $f(x) \approx f(\hat{x}(x))$ and $g(x,\hat{x}(x)) \approx I(\hat{x}(x))$.

Proof sketch contd.

- Next, define a function $g : [0,1]^n \times [0,1]^n \to \mathbb{R}$ as

$$g(x,y) := \sum_{i=1}^{n} (x_i \log y_i + (1-x_i)\log(1-y_i))\,.$$

- The second step is to show that with high probability,

$$g(X,\hat{X}) \approx g(\hat{X},\hat{X}) = I(\hat{X})\,.$$

- Suppose that these two steps have been proved.
- Let A be the set of all x where $f(x) \approx f(\hat{x}(x))$ and $g(x,\hat{x}(x)) \approx I(\hat{x}(x))$.
- Since $X \in A$ with high probability,

$$\frac{\sum_{x\in A} e^{f(x)}}{\sum_{x\in\{0,1\}^n} e^{f(x)}} \approx 1\,.$$

Proof sketch contd.

- Therefore

$$F = \log \sum_{x \in \{0,1\}^n} e^{f(x)} \approx \log \sum_{x \in A} e^{f(x)}$$
$$\approx \log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))} .$$

Proof sketch contd.

- Therefore

$$F = \log \sum_{x \in \{0,1\}^n} e^{f(x)} \approx \log \sum_{x \in A} e^{f(x)}$$
$$\approx \log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))} .$$

- Now let ϵ be a small positive number.

Proof sketch contd.

- Therefore

$$F = \log \sum_{x \in \{0,1\}^n} e^{f(x)} \approx \log \sum_{x \in A} e^{f(x)}$$

$$\approx \log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))} .$$

- Now let ϵ be a small positive number.
- Using the low complexity gradient condition, it is possible to construct a set $\mathcal{D}'(\epsilon)$ of size $e^{o(n)}$ such that for each $x \in [0,1]^n$ there exists $p \in \mathcal{D}'(\epsilon)$ such that $\hat{x}(x) \approx p$.

Proof sketch contd.

- Therefore

$$F = \log \sum_{x\in\{0,1\}^n} e^{f(x)} \approx \log \sum_{x\in A} e^{f(x)}$$
$$\approx \log \sum_{x\in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x,\hat{x}(x))} .$$

- Now let ϵ be a small positive number.
- Using the low complexity gradient condition, it is possible to construct a set $\mathcal{D}'(\epsilon)$ of size $e^{o(n)}$ such that for each $x \in [0,1]^n$ there exists $p \in \mathcal{D}'(\epsilon)$ such that $\hat{x}(x) \approx p$.
- For each $p \in \mathcal{D}'(\epsilon)$ let $\mathcal{P}(p)$ be the set of all $x \in \{0,1\}^n$ such that $\hat{x}(x) \approx p$.

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,
$$\sum_{x \in \{0,1\}^n} e^{g(x,p)} = 1.$$

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,

$$\sum_{x \in \{0,1\}^n} e^{g(x,p)} = 1.$$

- Therefore,

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,

$$\sum_{x \in \{0,1\}^n} e^{g(x,p)} = 1.$$

- Therefore,

$$\log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))}$$
$$\leq \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))}$$

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,

$$\sum_{x \in \{0,1\}^n} e^{g(x,p)} = 1.$$

- Therefore,

$$\log \sum_{x \in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))}$$

$$\leq \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x, \hat{x}(x))}$$

$$\approx \log \sum_{p \in \mathcal{D}'(\epsilon)} \sum_{x \in \mathcal{P}(p)} e^{f(p) - I(p) + g(x,p)}$$

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,

$$\sum_{x\in\{0,1\}^n} e^{g(x,p)} = 1.$$

- Therefore,

$$\log \sum_{x\in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x,\hat{x}(x))}$$

$$\leq \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x,\hat{x}(x))}$$

$$\approx \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(p) - I(p) + g(x,p)}$$

$$\leq \log \sum_{p\in\mathcal{D}'(\epsilon)} e^{f(p) - I(p)}$$

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,

$$\sum_{x\in\{0,1\}^n} e^{g(x,p)} = 1.$$

- Therefore,

$$\log \sum_{x\in A} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x,\hat{x}(x))}$$

$$\le \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(\hat{x}(x)) - I(\hat{x}(x)) + g(x,\hat{x}(x))}$$

$$\approx \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(p) - I(p) + g(x,p)}$$

$$\le \log \sum_{p\in\mathcal{D}'(\epsilon)} e^{f(p)-I(p)} \le \log|\mathcal{D}'(\epsilon)| + \sup_{p\in[0,1]^n} (f(p) - I(p)).$$

Proof sketch contd.

- The crucial fact is that for any $p \in [0,1]^n$,
$$\sum_{x\in\{0,1\}^n} e^{g(x,p)} = 1.$$
- Therefore,
$$\begin{aligned}
&\log \sum_{x\in A} e^{f(\hat{x}(x))-I(\hat{x}(x))+g(x,\hat{x}(x))}\\
&\le \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(\hat{x}(x))-I(\hat{x}(x))+g(x,\hat{x}(x))}\\
&\approx \log \sum_{p\in\mathcal{D}'(\epsilon)} \sum_{x\in\mathcal{P}(p)} e^{f(p)-I(p)+g(x,p)}\\
&\le \log \sum_{p\in\mathcal{D}'(\epsilon)} e^{f(p)-I(p)} \le \log|\mathcal{D}'(\epsilon)| + \sup_{p\in[0,1]^n} (f(p)-I(p))\,.
\end{aligned}$$
- This completes the proof sketch for the upper bound, modulo the two unproved steps. I will now sketch why $f(X) \approx f(\hat{X})$.

Proof of $f(X) \approx f(\hat{X})$

- To show this, define

$$h(x) := f(x) - f(\hat{x}(x)) \,.$$

Proof of $f(X) \approx f(\hat{X})$

- To show this, define

$$h(x) := f(x) - f(\hat{x}(x)) .$$

- Let $u_i(t,x) := f_i(tx + (1-t)\hat{x}(x))$, so that

$$h(x) = \int_0^1 \sum_{i=1}^n (x_i - \hat{x}_i(x)) u_i(t,x)\, dt.$$

Proof of $f(X) \approx f(\hat{X})$

- To show this, define

$$h(x) := f(x) - f(\hat{x}(x))\,.$$

- Let $u_i(t,x) := f_i(tx + (1-t)\hat{x}(x))$, so that

$$h(x) = \int_0^1 \sum_{i=1}^n (x_i - \hat{x}_i(x)) u_i(t,x)\,dt.$$

- Thus, if $D := f(X) - f(\hat{X})$, then

$$\mathbb{E}(D^2) = \int_0^1 \sum_{i=1}^n \mathbb{E}((X_i - \hat{X}_i) u_i(t,X) D)\,dt\,. \qquad (\dagger)$$

Proof of $f(X) \approx f(\hat{X})$, contd.

- Let $X^{(i)}$ denote the random vector $(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and let $D_i := h(X^{(i)})$.

Proof of $f(X) \approx f(\hat{X})$, contd.

- Let $X^{(i)}$ denote the random vector $(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and let $D_i := h(X^{(i)})$.
- Then note that $u_i(t, X^{(i)})D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only.

Proof of $f(X) \approx f(\hat{X})$, contd.

- Let $X^{(i)}$ denote the random vector $(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and let $D_i := h(X^{(i)})$.
- Then note that $u_i(t, X^{(i)})D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only.
- Therefore since $\hat{X}_i = \mathbb{E}(X_i \mid (X_j)_{j \neq i})$,

$$\mathbb{E}((X_i - \hat{X}_i)u_i(t, X^{(i)})D_i) = 0.$$

Proof of $f(X) \approx f(\hat{X})$, contd.

- Let $X^{(i)}$ denote the random vector $(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and let $D_i := h(X^{(i)})$.
- Then note that $u_i(t, X^{(i)})D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only.
- Therefore since $\hat{X}_i = \mathbb{E}(X_i \mid (X_j)_{j \neq i})$,

$$\mathbb{E}((X_i - \hat{X}_i)u_i(t, X^{(i)})D_i) = 0.$$

- Thus,

$$\begin{aligned}&\mathbb{E}((X_i - \hat{X}_i)u_i(t, X)D)\\ &= \mathbb{E}((X_i - \hat{X}_i)u_i(t, X)D) - \mathbb{E}((X_i - \hat{X}_i)u_i(t, X^{(i)})D_i)\,.\end{aligned}$$

Proof of $f(X) \approx f(\hat{X})$, contd.

- Let $X^{(i)}$ denote the random vector $(X_1, \ldots, X_{i-1}, 0, X_{i+1}, \ldots, X_n)$ and let $D_i := h(X^{(i)})$.
- Then note that $u_i(t, X^{(i)})D_i$ is a function of the random variables $(X_j)_{j \neq i}$ only.
- Therefore since $\hat{X}_i = \mathbb{E}(X_i \mid (X_j)_{j \neq i})$,

$$\mathbb{E}((X_i - \hat{X}_i)u_i(t, X^{(i)})D_i) = 0.$$

- Thus,

$$\begin{aligned} &\mathbb{E}((X_i - \hat{X}_i)u_i(t, X)D) \\ &= \mathbb{E}((X_i - \hat{X}_i)u_i(t, X)D) - \mathbb{E}((X_i - \hat{X}_i)u_i(t, X^{(i)})D_i) \,. \end{aligned}$$

- If the smoothness term is small, then $u_i(t, X) \approx u_i(t, X^{(i)})$ and $D \approx D_i$. Together with the identity (†), this shows that $f(X) \approx f(\hat{X})$ with high probability.

Summary

- Large deviations for sparse random graphs cannot be tackled by techniques based on graph limit theory and Szemerédi's regularity lemma.

Summary

- Large deviations for sparse random graphs cannot be tackled by techniques based on graph limit theory and Szemerédi's regularity lemma.
- The recently developed theory of large deviations for nonlinear functions of Bernoulli random variables has seen some successful applications in this class of problems, leading to the solutions of some longstanding questions.

Summary

- Large deviations for sparse random graphs cannot be tackled by techniques based on graph limit theory and Szemerédi's regularity lemma.
- The recently developed theory of large deviations for nonlinear functions of Bernoulli random variables has seen some successful applications in this class of problems, leading to the solutions of some longstanding questions.
- The degree of sparsity that is allowed by these theorems is less than optimal. New breakthroughs are required.