Limiting Distribution of Quadratic Chaos on Graphs[1]

Somabha Mukherjee

University of Pennsylvania

March 23, 2019

[1] Joint work with Bhaswar B. Bhattacharya, Sumit Mukherjee

Outline

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1 \leq i \leq n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1 \leq i \leq n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.
- $T_n := \sum_{1 \leq u < v \leq n} A_n(u,v) X_u X_v$:- Quadratic Chaos on G_n.

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1 \le i \le n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.
- $T_n := \sum_{1 \le u < v \le n} A_n(u,v) X_u X_v$:- Quadratic Chaos on G_n.
- Goal: Find the limiting distribution of T_n in the sparse regime: $\mathbb{E} T_n = p_n^2 |E(G_n)| = O(1)$.

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1\leq i\leq n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.
- $T_n := \sum_{1\leq u<v\leq n} A_n(u,v)X_uX_v$:- Quadratic Chaos on G_n.
- Goal: Find the limiting distribution of T_n in the sparse regime: $\mathbb{E}T_n = p_n^2|E(G_n)| = O(1)$.

Applications:

- A form similar to T_n arises as the Hamiltonian of the Ising model in Statistical Physics, in absence of an external magnetic field.

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1 \le i \le n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.
- $T_n := \sum_{1 \le u < v \le n} A_n(u,v) X_u X_v$:- Quadratic Chaos on G_n.
- Goal: Find the limiting distribution of T_n in the sparse regime: $\mathbb{E} T_n = p_n^2 |E(G_n)| = O(1)$.

Applications:

- A form similar to T_n arises as the Hamiltonian of the Ising model in Statistical Physics, in absence of an external magnetic field.
- Asymptotics of T_n often used in the study of coincidences, for example, the Birthday problem. What is the probability that there are two friends in a friendship network G_n with birthdays on January 1?

Framework, Goal and Some Connections

- G_n: sequence of graphs on n vertices, with adjacency matrix A_n.
- $\{X_i\}_{1\leq i\leq n}$: sequence of i.i.d. Bernoulli(p_n) random variables, with $\lim_{n\to\infty} p_n = 0$.
- $T_n := \sum_{1\leq u<v\leq n} A_n(u,v)X_uX_v$:- Quadratic Chaos on G_n.
- Goal: Find the limiting distribution of T_n in the sparse regime: $\mathbb{E}T_n = p_n^2|E(G_n)| = O(1)$.

Applications:

- A form similar to T_n arises as the Hamiltonian of the Ising model in Statistical Physics, in absence of an external magnetic field.
- Asymptotics of T_n often used in the study of coincidences, for example, the Birthday problem. What is the probability that there are two friends in a friendship network G_n with birthdays on January 1?
- Various graph based nonparametric two-sample tests in Statistics are functions of T_n. In particular, the Bahadur slopes of these statistics can be computed using a large deviation theory for T_n proved by us.

Outline

Examples:

Let $X_1, X_2, \ldots$ be i.i.d. Bernoulli$\left(\frac{\lambda}{n}\right)$.

- Complete Graph on n Vertices: $T_n \xrightarrow{D} \binom{N}{2}$, where $N \sim$ Poisson(λ).

Examples:

Let $X_1, X_2, \ldots$ be i.i.d. Bernoulli$\left(\frac{\lambda}{n}\right)$.

- Complete Graph on n Vertices: $T_n \xrightarrow{D} \binom{N}{2}$, where $N \sim$ Poisson(λ).
- Dense Erdős-Rényi Graphs: $\mathbb{G}(n, p)$: $T_n \xrightarrow{D} \mathrm{Binomial}\left(\binom{N}{2}, p\right)$, where $N \sim$ Poisson(λ).

Examples:

Let $X_1, X_2, \ldots$ be i.i.d. Bernoulli$(\frac{\lambda}{n})$.

- Complete Graph on n Vertices: $T_n \xrightarrow{D} \binom{N}{2}$, where $N \sim$ Poisson(λ).
- Dense Erdős-Rényi Graphs: $\mathbb{G}(n, p)$: $T_n \xrightarrow{D} \mathrm{Binomial}\left(\binom{N}{2}, p\right)$, where $N \sim$ Poisson(λ).
- Dense Stochastic Block Models: $\mathrm{SBM}(n_1, n - n_1, p, q)$: If $n_1/n \to \alpha$, then

$$T_n \xrightarrow{D} \mathrm{Binomial}\left(\binom{N_1}{2}, p\right) + \mathrm{Binomial}\left(\binom{N_2}{2}, p\right) + \mathrm{Binomial}\left(N_1 N_2, q\right) ,$$

where $N_1 \sim \mathrm{Poisson}(\alpha\lambda)$, $N_2 \sim \mathrm{Poisson}((1-\alpha)\lambda)$ are independent, and the three summands above are independent, given N_1, N_2.

Examples:

Let $X_1, X_2, \dots$ be i.i.d. Bernoulli$(\frac{\lambda}{n})$.

- Complete Graph on n Vertices: $T_n \xrightarrow{D} \binom{N}{2}$, where $N \sim$ Poisson(λ).
- Dense Erdős-Rényi Graphs: $\mathbb{G}(n,p)$: $T_n \xrightarrow{D} \mathrm{Binomial}\left(\binom{N}{2}, p\right)$, where $N \sim$ Poisson(λ).
- Dense Stochastic Block Models: $\mathrm{SBM}(n_1, n-n_1, p, q)$: If $n_1/n \to \alpha$, then

$$T_n \xrightarrow{D} \mathrm{Binomial}\left(\binom{N_1}{2}, p\right) + \mathrm{Binomial}\left(\binom{N_2}{2}, p\right) + \mathrm{Binomial}\left(N_1 N_2, q\right) ,$$

  where $N_1 \sim \mathrm{Poisson}(\alpha\lambda)$, $N_2 \sim \mathrm{Poisson}((1-\alpha)\lambda)$ are independent, and the three summands above are independent, given N_1, N_2.
- Disjoint Union of n many $n-$stars: (A Sparse Graph Example):
  $T_n \xrightarrow{D} \mathrm{Poisson}(N)$, where $N \sim \mathrm{Poisson}(\lambda)$.

Outline

Assumptions:

- Stretched Graphon associated with G_n: $W_n : [0, np_n]^2 \mapsto [0, 1]$ given by:

$$W_n(x, y) = A_n\left(\lceil x/p_n \rceil, \lceil y/p_n \rceil\right) \ .$$

Assumptions:

- Stretched Graphon associated with G_n: $W_n : [0, np_n]^2 \mapsto [0, 1]$ given by:

$$W_n(x, y) = A_n\left(\lceil x/p_n \rceil, \lceil y/p_n \rceil\right) \ .$$

- Stretched Degree Function of G_n: $d_n(x) := \int_0^\infty W_n(x, y)dy$.

Assumptions:

- Stretched Graphon associated with G_n: $W_n : [0, np_n]^2 \mapsto [0,1]$ given by:

$$W_n(x,y) = A_n\left(\lceil x/p_n \rceil, \lceil y/p_n \rceil\right) \; .$$

- Stretched Degree Function of G_n: $d_n(x) := \int_0^\infty W_n(x,y)dy$.
- W_n converges in cut norm to a symmetric, integrable function $W : [0,\infty)^2 \mapsto [0,1]$, i.e. for all $K \in \mathbb{N}$ large enough,

$$\sup_{f,g:[0,K]\mapsto[-1,1]} \left| \int_{[0,K]^2} \left(W_n(x,y) - W(x,y)\right) f(x)g(y)dxdy \right| \to 0 \; .$$

Assumptions:

- Stretched Graphon associated with G_n: $W_n : [0, np_n]^2 \mapsto [0,1]$ given by:

$$W_n(x,y) = A_n\left(\lceil x/p_n \rceil, \lceil y/p_n \rceil\right) .$$

- Stretched Degree Function of G_n: $d_n(x) := \int_0^\infty W_n(x,y)dy$.
- W_n converges in cut norm to a symmetric, integrable function $W : [0,\infty)^2 \mapsto [0,1]$, i.e. for all $K \in \mathbb{N}$ large enough,

$$\sup_{f,g:[0,K]\mapsto[-1,1]} \left| \int_{[0,K]^2} \left(W_n(x,y) - W(x,y)\right) f(x)g(y)dxdy \right| \to 0 .$$

- $\lim_{K\to\infty} \lim_{n\to\infty} \frac{1}{2} \int_K^\infty \int_K^\infty W_n(x,y)dxdy = \lambda$.

Assumptions:

- Stretched Graphon associated with G_n: $W_n : [0, np_n]^2 \mapsto [0,1]$ given by:

$$W_n(x,y) = A_n\left(\lceil x/p_n \rceil, \lceil y/p_n \rceil\right) \ .$$

- Stretched Degree Function of G_n: $d_n(x) := \int_0^\infty W_n(x,y)dy$.
- W_n converges in cut norm to a symmetric, integrable function $W : [0,\infty)^2 \mapsto [0,1]$, i.e. for all $K \in \mathbb{N}$ large enough,

$$\sup_{f,g:[0,K]\mapsto[-1,1]} \left| \int_{[0,K]^2} \left(W_n(x,y) - W(x,y)\right) f(x)g(y)dxdy \right| \to 0 \ .$$

- $\lim_{K\to\infty} \lim_{n\to\infty} \frac{1}{2} \int_K^\infty \int_K^\infty W_n(x,y)dxdy = \lambda$.
- There exists an integrable function $d : [0,\infty) \mapsto [0,\infty)$, such that for all $K \in \mathbb{N}$ large enough,

$$(U_K, d_n(U_K)) \xrightarrow{D} (U_K, d(U_K)) \ ,$$

where U_K is a uniform random variable on the interval $[0, K]$.

Main Result:

Under the assumptions in the previous slide, $T_n \xrightarrow{D} Q_1 + Q_2 + Q_3$, where

- $Q_3 \sim \text{Poisson}(\lambda)$, and is independent of (Q_1, Q_2) ,
- Q_2 conditional on some random variable R_2 has a Poisson(R_2) distribution,
- The joint moment generating function of Q_1 and R_2 is given by:

$$\mathbb{E}e^{-sQ_1 - tR_2} = \mathbb{E}\exp\left\{\frac{1}{2}\int_0^\infty \int_0^\infty \phi_{W,s}(x,y)dN(x)dN(y) - t\int_0^\infty \Delta(x)dN(x)\right\}$$

where

1. $\phi_{W,s}(x,y) = \log\left(1 - W(x,y) + W(x,y)e^{-s}\right)$,
2. $\Delta(x) = d(x) - \int_0^\infty W(x,y)dy$ and
3. $\{N(t) : t \geq 0\}$ is a homogeneous Poisson process of rate 1.

Main Result:

Under the assumptions in the previous slide, $T_n \xrightarrow{D} Q_1 + Q_2 + Q_3$, where

- $Q_3 \sim \text{Poisson}(\lambda)$, and is independent of (Q_1, Q_2) ,
- Q_2 conditional on some random variable R_2 has a Poisson(R_2) distribution,
- The joint moment generating function of Q_1 and R_2 is given by:

$$\mathbb{E}e^{-sQ_1 - tR_2} = \mathbb{E}\exp\left\{\frac{1}{2}\int_0^\infty \int_0^\infty \phi_{W,s}(x,y)dN(x)dN(y) - t\int_0^\infty \Delta(x)dN(x)\right\}$$

where

1. $\phi_{W,s}(x,y) = \log\left(1 - W(x,y) + W(x,y)e^{-s}\right)$,
2. $\Delta(x) = d(x) - \int_0^\infty W(x,y)dy$ and
3. $\{N(t) : t \geq 0\}$ is a homogeneous Poisson process of rate 1.

Dense Graphs and a Converse

If G_n is a sequence of dense graphs converging in cut distance to a graphon W, then $Q_2 = Q_3 = 0$ and $\Delta \equiv 0$. Conversely, for any sequence of dense graphs with $p_n^2|E(G_n)| = O(1)$, the limiting distribution of T_n, if exists, must be of the above form for some symmetric, measurable, integrable $W : [0,\infty)^2 \mapsto [0,1]$.

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.
- It holds when $X_1, X_2, \ldots$ are i.i.d. non-negative integer valued random variables with $p_n := \mathbb{P}(X_1 = 1) \to 0$, such that $|\mathbb{E}(G_n)|p_n^2 = \Theta(1)$ and $\lim_{n\to\infty} \frac{\mathbb{E}X_1}{p_n} = 1$.

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.
- It holds when $X_1, X_2, \ldots$ are i.i.d. non-negative integer valued random variables with $p_n := \mathbb{P}(X_1 = 1) \to 0$, such that $|\mathbb{E}(G_n)|p_n^2 = \Theta(1)$ and $\lim_{n\to\infty} \frac{\mathbb{E}X_1}{p_n} = 1$.

Examples of Such a Distribution

- Sparse Binomial: $\text{Bin}(n, \theta_n)$, where $n\theta_n \to 0$.

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.
- It holds when $X_1, X_2, \ldots$ are i.i.d. non-negative integer valued random variables with $p_n := \mathbb{P}(X_1 = 1) \to 0$, such that $|\mathbb{E}(G_n)|p_n^2 = \Theta(1)$ and $\lim_{n\to\infty} \frac{\mathbb{E}X_1}{p_n} = 1$.

Examples of Such a Distribution

- Sparse Binomial: $\text{Bin}(n, \theta_n)$, where $n\theta_n \to 0$.
- Sparse Poisson: $\text{Pois}(\theta_n)$, where $\theta_n \to 0$.

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.
- It holds when $X_1, X_2, \ldots$ are i.i.d. non-negative integer valued random variables with $p_n := \mathbb{P}(X_1 = 1) \to 0$, such that $|\mathbb{E}(G_n)| p_n^2 = \Theta(1)$ and $\lim_{n\to\infty} \frac{\mathbb{E}X_1}{p_n} = 1$.

Examples of Such a Distribution

- Sparse Binomial: $\text{Bin}(n, \theta_n)$, where $n\theta_n \to 0$.
- Sparse Poisson: $\text{Pois}(\theta_n)$, where $\theta_n \to 0$.
- Sparse Negative Binomial: $\text{NB}(n, \theta_n)$, where $n\theta_n \to 0$. (Number of successes in independent trials, with probability θ_n of success, observed until n failures.)

A Universality Result:

- Our main result holds when the X_i's come from distributions slightly more general than Bernoulli with mean going to 0.
- It holds when $X_1, X_2, \ldots$ are i.i.d. non-negative integer valued random variables with $p_n := \mathbb{P}(X_1 = 1) \to 0$, such that $|\mathbb{E}(G_n)| p_n^2 = \Theta(1)$ and $\lim_{n\to\infty} \frac{\mathbb{E}X_1}{p_n} = 1$.

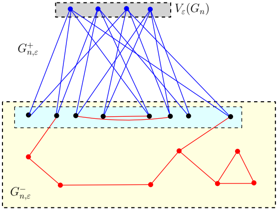
Examples of Such a Distribution

- Sparse Binomial: $\mathrm{Bin}(n, \theta_n)$, where $n\theta_n \to 0$.
- Sparse Poisson: $\mathrm{Pois}(\theta_n)$, where $\theta_n \to 0$.
- Sparse Negative Binomial: $\mathrm{NB}(n, \theta_n)$, where $n\theta_n \to 0$. (Number of successes in independent trials, with probability θ_n of success, observed until n failures.)
- Sparse Hypergeometric: $\mathrm{HG}(N_n, K_n, m_n)$, where $m_n K_n / N_n \to 0$. (Number of successes in m_n draws WOR from a population of size N_n having exactly K_n success states.)

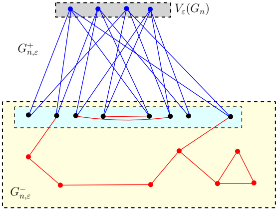
Outline

Answer: By Partitioning the Graph

- $V_\varepsilon(G_n) := \left\{ v \in V(G_n) : \deg(v) > \frac{\varepsilon}{p_n} \right\}.$

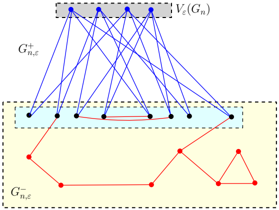


Answer: By Partitioning the Graph



- $V_\varepsilon(G_n) := \left\{ v \in V(G_n) : \deg(v) > \frac{\varepsilon}{p_n} \right\}$.
- $T_{n,1} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)} A_n(u,v) X_u X_v,$
- $T_{n,2} := \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v,$
- $T_{n,3} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n)^c, v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v.$

Answer: By Partitioning the Graph

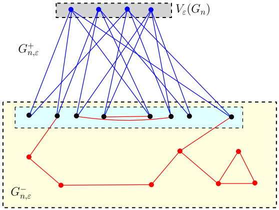


- $V_\varepsilon(G_n) := \left\{ v \in V(G_n) : \deg(v) > \frac{\varepsilon}{p_n} \right\}.$
- $T_{n,1} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)} A_n(u,v) X_u X_v,$
- $T_{n,2} := \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v,$
- $T_{n,3} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n)^c, v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v.$
- Step 1: $T_{n,3}$ is asymptotically independent of $(T_{n,1}, T_{n,2})$ in terms of mixed moments, i.e.

$$\mathbb{E}(T_{n,1}^a T_{n,2}^b T_{n,3}^c) - \mathbb{E}(T_{n,1}^a T_{n,2}^b)\mathbb{E}(T_{n,3}^c) \to 0$$

as $n \to \infty$ followed by $\varepsilon \to 0$.

Answer: By Partitioning the Graph

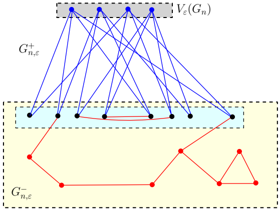


- $V_\varepsilon(G_n) := \left\{ v \in V(G_n) : \deg(v) > \frac{\varepsilon}{p_n} \right\}$.
- $T_{n,1} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)} A_n(u,v) X_u X_v$,
- $T_{n,2} := \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v$,
- $T_{n,3} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n)^c, v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v$.
- Step 1: $T_{n,3}$ is asymptotically independent of $(T_{n,1}, T_{n,2})$ in terms of mixed moments, i.e.

$$\mathbb{E}(T^a_{n,1} T^b_{n,2} T^c_{n,3}) - \mathbb{E}(T^a_{n,1} T^b_{n,2}) \mathbb{E}(T^c_{n,3}) \to 0$$

  as $n \to \infty$ followed by $\varepsilon \to 0$.
- The Poisson term Q_3 comes as a limit of $T_{n,3}$.

Answer: By Partitioning the Graph



- $V_\varepsilon(G_n) := \left\{v \in V(G_n) : \deg(v) > \frac{\varepsilon}{p_n}\right\}$.
- $T_{n,1} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)} A_n(u,v) X_u X_v$,
- $T_{n,2} := \sum_{u \in V_\varepsilon(G_n), v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v$,
- $T_{n,3} := \frac{1}{2} \sum_{u \in V_\varepsilon(G_n)^c, v \in V_\varepsilon(G_n)^c} A_n(u,v) X_u X_v$.
- Step 1: $T_{n,3}$ is asymptotically independent of $(T_{n,1}, T_{n,2})$ in terms of mixed moments, i.e.

$$\mathbb{E}(T^a_{n,1} T^b_{n,2} T^c_{n,3}) - \mathbb{E}(T^a_{n,1} T^b_{n,2})\mathbb{E}(T^c_{n,3}) \to 0$$

  as $n \to \infty$ followed by $\varepsilon \to 0$.
- The Poisson term Q_3 comes as a limit of $T_{n,3}$.
- (Q_1, Q_2) appears as the limit of $(T_{n,1}, T_{n,2})$, by an edge-independent random graph planting argument.

Outline

References

1. C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi, Convergent sequences of dense graphs I: Subgraph frequencies, metric properties and testing, *Advances in Mathematics*, Vol. 219, 1801–1851, 2009.

2. C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vesztergombi, Convergent sequences of dense graphs II. Multiway cuts and statistical physics, *Annals of Mathematics*, Vol. 176, 151–219, 2012.

3. L. Lovász, *Large Networks and Graph Limits*, Colloquium Publications, Vol. 60, 2012.

4. B. B. Bhattacharya, S. Mukherjee, and S. Mukherjee, Birthday paradox, monochromatic subgraphs, and the second moment phenomenon, `arXiv:1711.01465`, 2017.