Limiting Distributions and Large Deviations of Motif Counts in Randomly Colored Graphs[1]

Somabha Mukherjee

University of Pennsylvania

November 16, 2018

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

Outline

Setup

- G_n: sequence of graphs on n vertices.

Setup

- G_n: sequence of graphs on n vertices.
- Each vertex of G_n is colored independently of the others, using one of c_n ($\to \infty$) many colors, chosen uniformly at random.

Setup

- G_n: sequence of graphs on n vertices.
- Each vertex of G_n is colored independently of the others, using one of c_n $(\to \infty)$ many colors, chosen uniformly at random.
- H: fixed, connected graph.

Setup

- G_n: sequence of graphs on n vertices.
- Each vertex of G_n is colored independently of the others, using one of c_n $(\to \infty)$ many colors, chosen uniformly at random.
- H: fixed, connected graph.
- $T(H, G_n)$: Number of monochromatic copies of H in G_n.

Setup

- G_n: sequence of graphs on n vertices.
- Each vertex of G_n is colored independently of the others, using one of c_n $(\to \infty)$ many colors, chosen uniformly at random.
- H: fixed, connected graph.
- $T(H, G_n)$: Number of monochromatic copies of H in G_n.

Connection to the Birthday Problem

- In a friendship network, what is the probability that there are r friends with the same birthday?
- Same as asking for the probability of observing a monochromatic r−clique K_r in the friendship network graph colored with 365 colors.

Setup

- G_n: sequence of graphs on n vertices.
- Each vertex of G_n is colored independently of the others, using one of c_n ($\to \infty$) many colors, chosen uniformly at random.
- H: fixed, connected graph.
- $T(H, G_n)$: Number of monochromatic copies of H in G_n.

Connection to the Birthday Problem

- In a friendship network, what is the probability that there are r friends with the same birthday?
- Same as asking for the probability of observing a monochromatic $r-$clique K_r in the friendship network graph colored with 365 colors.
- In a group of m boys and n girls, what is the probability that there is a boy-girl birthday match?
- Equivalent to asking for the probability of observing a monochromatic edge in the bipartite graph $K_{m,n}$ colored using 365 colors.

Background and a Counterexample

- Bhattacharya, Diaconis and Mukherjee (2017) showed that $T(K_2, G_n) \xrightarrow{D}$ Poisson(λ), if $\mathbb{E}\left[T(K_2, G_n)\right] \to \lambda$.

Background and a Counterexample

- Bhattacharya, Diaconis and Mukherjee (2017) showed that $T(K_2, G_n) \xrightarrow{D}$ Poisson(λ), if $\mathbb{E}\left[T(K_2, G_n)\right] \to \lambda$.
- NOT TRUE FOR GENERAL SUBGRAPHS !

Background and a Counterexample

- Bhattacharya, Diaconis and Mukherjee (2017) showed that $T(K_2, G_n) \xrightarrow{D}$ Poisson(λ), if $\mathbb{E}[T(K_2, G_n)] \to \lambda$.
- NOT TRUE FOR GENERAL SUBGRAPHS !

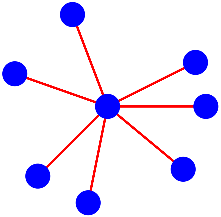

Figure: The 7-star, $K_{1,7}$

- For fixed r, suppose that $\mathbb{E}\, T(K_{1.,r}, K_{1,n}) \to \lambda$ as $n \to \infty$.

Background and a Counterexample

- Bhattacharya, Diaconis and Mukherjee (2017) showed that $T(K_2, G_n) \xrightarrow{D}$ Poisson(λ), if $\mathbb{E}[T(K_2, G_n)] \to \lambda$.
- NOT TRUE FOR GENERAL SUBGRAPHS !

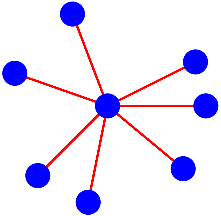

Figure: The 7-star, $K_{1,7}$

- For fixed r, suppose that $\mathbb{E}\, T(K_{1.,r}, K_{1,n}) \to \lambda$ as $n \to \infty$.
- Then, $T(K_{1,r}, G_n) \xrightarrow{D} \binom{X}{r}$, where $X \sim \mathrm{Pois}((r!\lambda)^{\frac{1}{r}})$.

Outline

The Second Moment Phenomenon

- G_n: growing sequence of graphs with vertices colored independently and uniformly using c_n ($\to \infty$) colors.

The Second Moment Phenomenon

- G_n: growing sequence of graphs with vertices colored independently and uniformly using c_n ($\to \infty$) colors.
- H: fixed, connected graph.

The Second Moment Phenomenon

- G_n: growing sequence of graphs with vertices colored independently and uniformly using c_n ($\to \infty$) colors.
- H: fixed, connected graph.

Theorem (Bhattacharya, Mukherjee, M. (2017))

If G_n and H are as above, and $\lambda > 0$, then

$$\lim_{n\to\infty} \mathbb{E}\,T(H, G_n) = \lambda \quad \textit{and} \quad \lim_{n\to\infty} \operatorname{Var} T(H, G_n) = \lambda \implies T(H, G_n) \xrightarrow{D} \operatorname{Pois}(\lambda).$$

Further, the converse is true if and only if H is a star-graph.
In fact, if H is not a star-graph, then for every $\lambda > 0$, $\exists$ a sequence of graphs $G_n(H)$ and a sequence $c_n \to \infty$, such that

$$T(H, G_n(H)) \xrightarrow{D} \operatorname{Pois}(\lambda) \quad \textit{but} \quad \mathbb{E}\,T(H, G_n(H)) \nrightarrow \lambda.$$

Towards a More General Result: A Notation

- $\mathcal{J}_t(H)$: (finite) set of all non-isomorphic graphs obtained by merging two copies of H in exactly t vertices $(1 \leq t \leq |V(H)|)$.

Towards a More General Result: A Notation

- $\mathcal{J}_t(H)$: (finite) set of all non-isomorphic graphs obtained by merging two copies of H in exactly t vertices $(1 \leq t \leq |V(H)|)$.
- For $H = C_4$, the 4−cycle, the sets $\mathcal{J}_2(H)$ and $\mathcal{J}_4(H)$ are illustrated below:

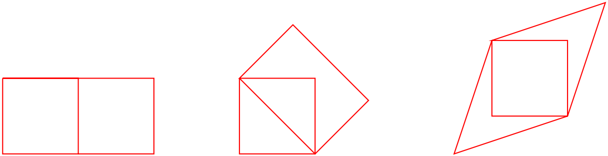

Figure: Graphs in the set $\mathcal{J}_2(C_4)$

Figure: Graphs in the set $\mathcal{J}_4(C_4)$

General Result: Linear Combination of Poissons

Theorem (Bhattacharya, Mukherjee, M. (2017))

Let G_n be a sequence of graphs colored uniformly with c_n $(\to \infty)$ colors, such that:

- *For every $k \in [1, N(H, K_{|V(H)|})]$, there exists $\lambda_k \geq 0$ such that*

$$\lim_{n\to\infty} \frac{\sum_{F \supseteq H : |V(F)| = |V(H)|,\ N(H,F)=k} N_{\mathrm{ind}}(F, G_n)}{c_n^{|V(H)|-1}} = \lambda_k,$$

- *For $t \in [2, |V(H)| - 1]$ and every $F \in \mathcal{J}_t(H)$, $N(F, G_n) = o(c_n^{2|V(H)|-t-1})$.*

Then

$$T(H, G_n) \xrightarrow{D} \sum_{k=1}^{N(H, K_{|V(H)|})} kX_k,$$

where $X_k \sim \mathrm{Pois}(\lambda_k)$ and the collection $\{X_k : 1 \leq k \leq N(H, K_{|V(H)|})\}$ is independent.

Outline

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.
- If H is unbalanced, define $\lambda(H) = \min_{H_1 \subset H : \alpha(H_1) > 0} \frac{|V(H)| - |V(H_1)|}{\alpha(H_1)}$, where

$$\alpha(H_1) := |E(H_1)|(|V(H)| - 1) + |E(H)|(|V(H_1)| - 1).$$

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.
- If H is unbalanced, define $\lambda(H) = \min_{H_1 \subset H : \alpha(H_1) > 0} \frac{|V(H)| - |V(H_1)|}{\alpha(H_1)}$, where

$$\alpha(H_1) := |E(H_1)|(|V(H)| - 1) + |E(H)|(|V(H_1)| - 1).$$

Theorem (Bhattacharya, Mukherjee, M. (2017))

Let H be a simple connected graph, and $G_n \sim G(n, p(n))$ be the Erdős-Rényi random graph with $p(n) \in (0,1)$ colored with $c_n(\to \infty)$ colors, such that $\mathbb{E}T(H, G_n) \to \lambda$.

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.
- If H is unbalanced, define $\lambda(H) = \min_{H_1 \subset H : \alpha(H_1) > 0} \frac{|V(H)| - |V(H_1)|}{\alpha(H_1)}$, where

$$\alpha(H_1) := |E(H_1)|(|V(H)| - 1) + |E(H)|(|V(H_1)| - 1).$$

Theorem (Bhattacharya, Mukherjee, M. (2017))

Let H be a simple connected graph, and $G_n \sim G(n, p(n))$ be the Erdős-Rényi random graph with $p(n) \in (0,1)$ colored with $c_n(\to \infty)$ colors, such that $\mathbb{E} T(H, G_n) \to \lambda$.

- *If $p(n) \to 0$ and $p(n) << n^{-\lambda(H)}$, then $T(H, G_n) \xrightarrow{P} 0$.*

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.
- If H is unbalanced, define $\lambda(H) = \min_{H_1 \subset H : \alpha(H_1) > 0} \frac{|V(H)| - |V(H_1)|}{\alpha(H_1)}$, where

$$\alpha(H_1) := |E(H_1)|(|V(H)| - 1) + |E(H)|(|V(H_1)| - 1).$$

Theorem (Bhattacharya, Mukherjee, M. (2017))

Let H be a simple connected graph, and $G_n \sim G(n, p(n))$ be the Erdős-Rényi random graph with $p(n) \in (0,1)$ colored with $c_n(\to \infty)$ colors, such that $\mathbb{E}T(H, G_n) \to \lambda$.

- *If $p(n) \to 0$ and $p(n) << n^{-\lambda(H)}$, then $T(H, G_n) \xrightarrow{P} 0$.*
- *If $p(n) \to 0$ and $p(n) >> n^{-\lambda(H)}$, then $T(H, G_n) \xrightarrow{D} \mathrm{Pois}(\lambda)$.*

A Phase Transition in Erdős-Rényi Random Graphs

Our results can be applied to Erdős-Rényi random graphs.

- If the fixed graph H is balanced, define $\lambda(H) = \frac{|V(H)|}{|E(H)|}$.
- If H is unbalanced, define $\lambda(H) = \min_{H_1 \subset H : \alpha(H_1) > 0} \frac{|V(H)| - |V(H_1)|}{\alpha(H_1)}$, where

$$\alpha(H_1) := |E(H_1)|(|V(H)| - 1) + |E(H)|(|V(H_1)| - 1).$$

Theorem (Bhattacharya, Mukherjee, M. (2017))

Let H be a simple connected graph, and $G_n \sim G(n, p(n))$ be the Erdős-Rényi random graph with $p(n) \in (0,1)$ colored with $c_n(\to \infty)$ colors, such that $\mathbb{E} T(H, G_n) \to \lambda$.

- *If $p(n) \to 0$ and $p(n) << n^{-\lambda(H)}$, then $T(H, G_n) \xrightarrow{P} 0$.*
- *If $p(n) \to 0$ and $p(n) >> n^{-\lambda(H)}$, then $T(H, G_n) \xrightarrow{D} \mathrm{Pois}(\lambda)$.*
- *If $p(n) = p \in (0,1)$ is fixed, then*

$$T(H, G_n) \xrightarrow{D} \sum_{F \supseteq H : |V(F)| = |V(H)|} N(H, F) X_F \ ,$$

where $X_F \sim \mathrm{Pois}\left(\lambda \cdot \frac{|Aut(H)|}{|Aut(F)|} p^{|E(F)| - |E(H)|} (1-p)^{\binom{|V(H)|}{2} - |E(F)|}\right)$, independent.

Outline

Large Deviation in a Different Scenario

- Vertices of G_n are colored red with probability p, independently.

Large Deviation in a Different Scenario

- Vertices of G_n are colored red with probability p, independently.
- $S(H, G_n)$: Number of copies of H in G_n with all vertices colored red.

Large Deviation in a Different Scenario

- Vertices of G_n are colored red with probability p, independently.
- $S(H, G_n)$: Number of copies of H in G_n with all vertices colored red.
- Assume that the number of copies of H in G_n is $N(H, G_n) = \Omega\left(n^{|V(H)|}\right)$.

Large Deviation in a Different Scenario

- Vertices of G_n are colored red with probability p, independently.
- $S(H, G_n)$: Number of copies of H in G_n with all vertices colored red.
- Assume that the number of copies of H in G_n is $N(H, G_n) = \Omega\left(n^{|V(H)|}\right)$.

Theorem (Bhattacharya, M. (2018))

Suppose that $p = \Omega(n^{-\alpha})$ *for some* $0 < \alpha < 1/[6(2|V(H)|^2 + |V(H)|)]$. *Define:*

$$f(y_1, \ldots, y_n) = \frac{1}{p^{|V(H)|} N(H, G_n) |Aut(H)|} \sum_{\mathbf{u} \in V(G_n)_{|V(H)|}} \prod_{i=1}^{|V(H)|} y_i \prod_{(a,b) \in E(H)} a_{u_a, u_b}(G_n) ,$$

$$\phi_p(t) = \inf_{\mathbf{y} \in [0,1]^n} \left\{ \sum_{i=1}^{n} y_i \log \frac{y_i}{p} + (1 - y_i) \log \frac{1 - y_i}{1 - p} : f(\mathbf{y}) \geq t \right\} .$$

Then, for $t \in (1, \liminf_{n\to\infty} p^{-|V(H)|})$,

$$\lim_{n\to\infty} \frac{\log \mathbb{P}\left(S(H, G_n) \geq t\, \mathbb{E}\left[S(H, G_n)\right]\right)}{-\phi_p(t)} = 1 .$$

Making the Variational Problem $n-$Free:

- If G_n converges in cut metric to a graphon W, and p is fixed,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S(H,G_n) \geq t\, \mathbb{E}\left[S(H,G_n)\right]\right)$$

$$= \; - \inf_{h:[0,1]\mapsto[0,1],\mathrm{m'ble}} \left\{ \int_0^1 I_p(h) : t(F,W,h) \geq tp^{|V(F)|} t(F,W) \right\},$$

where $I_p(h) := h \log \frac{h}{p} + (1-h)\log\frac{1-h}{1-p}$,

$$t(F,W,h) = \int_{[0,1]^{|V(F)|}} \prod_{(i,j)\in\in E(F)} W(x_i,x_j) \prod_{i=1}^{|V(F)|} h(x_i)\; dx_1 \ldots dx_{|V(F)|}\;,$$

and $t(F,W) = t(F,W,\mathbf{1})$, where $\mathbf{1}$ denotes the constant function 1.

Making the Variational Problem $n-$Free:

- If G_n converges in cut metric to a graphon W, and p is fixed,

$$\lim_{n\to\infty} \frac{1}{n} \log \mathbb{P}\left(S(H,G_n) \geq t\ \mathbb{E}\left[S(H,G_n)\right]\right)$$
$$= \quad - \inf_{h:[0,1]\mapsto[0,1],\mathrm{m'ble}} \left\{ \int_0^1 I_p(h) : t(F,W,h) \geq tp^{|V(F)|} t(F,W) \right\},$$

where $I_p(h) := h \log \frac{h}{p} + (1-h) \log \frac{1-h}{1-p}$,

$$t(F,W,h) = \int_{[0,1]^{|V(F)|}} \prod_{(i,j)\in\in E(F)} W(x_i,x_j) \prod_{i=1}^{|V(F)|} h(x_i)\ dx_1 \ldots dx_{|V(F)|}\ ,$$

and $t(F,W) = t(F,W,\mathbf{1})$, where $\mathbf{1}$ denotes the constant function 1.

- If the limiting graphon is in the block form $\sum_{i=1}^d \sum_{j=1}^d W_{i,j} \mathbb{1}_{A_i \times A_j}$ for some measurable partition $A_1, \ldots, A_d$ of $[0,1]$, then the $n-$ *free* variational problem above is actually a $d-$ dimensional optimization problem.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.
- If the variational problem has no replica symmetric solution, we call the scenario replica symmetry breaking.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.
- If the variational problem has no replica symmetric solution, we call the scenario replica symmetry breaking.
- In what follows, assume that the graphs G_n are regular, $H = K_2$, $r := \sqrt{t}p$, $I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.
- If the variational problem has no replica symmetric solution, we call the scenario replica symmetry breaking.
- In what follows, assume that the graphs G_n are regular, $H = K_2$, $r := \sqrt{tp}$, $I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$.
- If the point $\left(r^2, I_p(r)\right)$ lies on the convex minorant of the function $J_p(x) := I_p\left(\sqrt{x}\right)$, there is replica symmetry.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.
- If the variational problem has no replica symmetric solution, we call the scenario replica symmetry breaking.
- In what follows, assume that the graphs G_n are regular, $H = K_2$, $r := \sqrt{t}p$, $I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$.
- If the point $\left(r^2, I_p(r)\right)$ lies on the convex minorant of the function $J_p(x) := I_p\left(\sqrt{x}\right)$, there is replica symmetry.
- The function $J_p(x)$ is convex if and only if $p \geq \left(1 + e^2\right)^{-1}$.

Replica Symmetry and Breaking Phenomena:

- A solution f to the infinite dimensional variational problem is called replica symmetric, if f is the constant function.
- If the variational problem has no replica symmetric solution, we call the scenario replica symmetry breaking.
- In what follows, assume that the graphs G_n are regular, $H = K_2$, $r := \sqrt{t}p$, $I_p(x) = x \log \frac{x}{p} + (1-x) \log \frac{1-x}{1-p}$.
- If the point $\left(r^2, I_p(r)\right)$ lies on the convex minorant of the function $J_p(x) := I_p\left(\sqrt{x}\right)$, there is replica symmetry.
- The function $J_p(x)$ is convex if and only if $p \geq \left(1 + e^2\right)^{-1}$.
- If $p < \left(1 + e^2\right)^{-1}$ and the point $\left(r^2, I_p(r)\right)$ does not lie on the convex minorant of the function $J_p(x)$, we can construct examples of dense, regular graphs G_n giving rise to replica symmetry breaking.

Outline

References

1. S. Chatterjee and A. Dembo, **Nonlinear large deviations**, *Advances in Mathematics*, 2016.
2. R. Eldan, **Gaussian-width gradient complexity, reverse log-Sobolev inequalities and nonlinear large deviations**, *arXiv preprint*, 2016.
3. E. Lubetzky, Y. Zhao, **On replica symmetry of large deviations in random graphs**, Random Structures & Algorithms, 2014.
4. C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vestergombi, **Convergent sequence of dense graphs I: Subgraph frequencies, metric properties and testing**, *Advances in Mathematics*, 2009.
5. C. Borgs, J.T. Chayes, L. Lovász, V.T. Sós, and K. Vestergombi, **Convergent sequence of dense graphs II: Multiway cuts and statistical physics**, *Annals of Mathematics*, 2012.