# BIRTHDAY PROBLEM, MONOCHROMATIC SUBGRAPHS & THE SECOND MOMENT PHENOMENON

Somabha Mukherjee[1]

University of Pennsylvania

March 30, 2018

Joint work with Bhaswar B. Bhattacharya[2] and Sumit Mukherjee[3]

[1] Department of Statistics , University of Pennsylvania
[2] Department of Statistics , University of Pennsylvania
[3] Department of Statistics , Columbia University

# Outline


1. Introduction
2. The Second Moment Phenomenon
3. A More General Result
4. Sketch of Proof
5. Erdős-Renyi Random Graphs
6. Conclusion
7. Bibliography

# The Birthday Problem

- **Different Types of Questions on Birthday-Matching:**
  1. In a room, what is the approximate number of people needed to ensure that there will be $s$ people with the same birthday, with probability at least $p$? ($s \geq 2$ and $0 \leq p \leq 1$ are given)
  2. More generally, in a general friendship network with a large number ($n$) of individuals, what is the probability that there will be $s$ friends with the same birthday?
  3. In a group of $n$ boys and $n$ girls, what is the probability that there is a boy-girl birthday match?
- **A General Setup To Analyze the Above Questions:**
  - Each vertex of a graph $G_n$ is colored independently of the others, and uniformly, using one of $c_n = 365$ colors.
  - Question 1 asks what is the approximate value of $n$ needed to ensure that there is a monochromatic $s-$clique in $G_n = K_n$, with probability at least $p$?
  - Question 2 asks for the probability that there will be a monochromatic $s-$clique in a general friendship-network graph $G_n$.
  - Question 3 asks for the probability that there will be a monochromatic edge in the complete bipartite graph $G_n = K_{n,n}$.

# Our Goal

- **To Start With:** Each vertex of a graph $G_n$ is colored independently of the others, and uniformly, using one of $c_n$ $(\to \infty)$ colors.
- **Our Interest:** Generalizing the questions asked in the previous slide: Investigating the limiting distributional behavior of the number of monochromatic copies of a fixed, connected graph $H$ in $G_n$, denoted by $T(H, G_n)$, under suitable assumptions.
- Bhattacharya, Diaconis and Mukherjee [1] showed that under this independent, uniform coloring scheme, the number of monochromatic edges $T(K_2, G_n)$ of a graph $G_n$ converges in distribution to Pois($\lambda$), under the assumption $\mathbb{E}T(K_2, G_n) \to \lambda$.
- **Natural Question:** Does $\mathbb{E}T(H, G_n) \to \lambda$ imply that $T(H, G_n) \xrightarrow{D} \mathrm{Pois}(\lambda)$? NO! The next slide shows a counterexample.

# Convergence of First Moment is Not Enough!

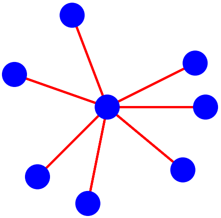

Figure: The 7-star, $K_{1,7}$

- $K_{1,n}$: star-graph with $n+1$ vertices, also called $n-$star.
- Under the independent uniform coloring scheme, suppose that for a fixed $r$,

$$\lim_{n\to\infty} \mathbb{E}T(K_{1,r}, K_{1,n}) \to \lambda.$$

- Then, $T(K_{1,r}, G_n) \xrightarrow{D} \binom{X}{r}$, where $X \sim \mathrm{Pois}((r!\lambda)^{\frac{1}{r}})$ (shown in [2]).

# Outline


1 Introduction

2 The Second Moment Phenomenon

3 A More General Result

4 Sketch of Proof

5 Erdős-Renyi Random Graphs

6 Conclusion

7 Bibliography

# Our Poisson Convergence Result

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

## Theorem (Bhattacharya, Mukherjee, M.)

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

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

*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*

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

The above theorem characterizes the second moment phenomenon for the number of monochromatic subgraphs in a uniform random coloring of a graph sequence.

# Why is the Converse not True for Non-Stars?

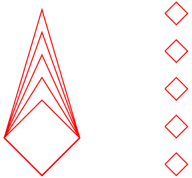

Figure: The graph $G_5(C_4)$

- $P_n(H)$: pyramid of height $n$ formed by merging $n$ copies of $H$ on $|V(H)| - 1$ vertices (called base vertices), corresponding to the isomorphisms.
- $G_n(H) = P_n(H)$ plus $n$ disjoint copies of $H$. Take $c_n = \lfloor n^{1/(|V(H)|-1)} \rfloor$.
- Any copy of $H$ in $P_n(H)$ passes through at least 2 base vertices of $P_n(H)$. So,

$$\mathbb{P}(T(H, P_n(H)) > 0) \leq \binom{|V(H)| - 1}{2} \frac{1}{c_n} \to 0,$$

  i.e. $T(H, P_n(H)) \xrightarrow{P} 0$. Hence, $T(H, G_n(H)) \xrightarrow{D} \mathrm{Pois}(1)$.
- However, $\liminf \mathbb{E} T(H, G_n(H)) = \liminf \mathbb{E} T(H, P_n(H)) + 1 \geq 2$ for every $n$.

# Outline

1 Introduction

2 The Second Moment Phenomenon

3 A More General Result

4 Sketch of Proof

5 Erdős-Renyi Random Graphs

6 Conclusion

7 Bibliography

# Notations Required for Stating the General Result

- $\mathcal{J}_t(H)$: (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:

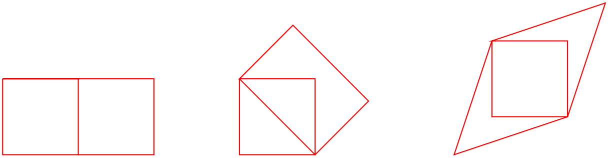

Figure: Graphs in the set $\mathcal{J}_2(C_4)$

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

# Our Poisson-Linear-Combination Result

The following is the most general result proved by us, from which the first theorem follows as a corollary:

## Theorem (Bhattacharya, Mukherjee, M.)

*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*

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

*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


1 Introduction

2 The Second Moment Phenomenon

3 A More General Result

4 Sketch of Proof

5 Erdős-Renyi Random Graphs

6 Conclusion

7 Bibliography

# Outline of Proof of the General Theorem

We followed a method of moments approach, but not directly. In fact, it may not be true that $\mathbb{E}T(H,G_n)^r \to \mathbb{E}\left(\sum_{k=1}^{N(H,K_{|V(H)|})} k\mathrm{Pois}(\lambda_k)\right)^r$ for every $r$.

- **What we did:**
  1. Decomposed the random variable $T(H,G_n)$ as a sum of two quantities: $T^+(H,G_n)$ and $T^-(H,G_n)$.
  2. $T^+(H,G_n)$ was defined by a truncation on $T(H,G_n)$, and $T^-(H,G_n) := T(H,G_n) - T^+(H,G_n)$.
  3. Showed that $T^-(H,G_n)$ converges in $L^1$ to 0.
  4. Showed that $\mathbb{E}T^+(H,G_n)^r \to \mathbb{E}\left(\sum_{k=1}^{N(H,K_{|V(H)|})} k\mathrm{Pois}(\lambda_k)\right)^r$ for every $r$.
- In the next few slides, we sketch the proof for $H = K_3$.
- The general proof exploits an idea essentially similar to the proof for $K_3$, but some steps are more involved, and hard to present.

# Proof of the Poisson Convergence Result for Triangles

- $N(H, G_n)$ : Number of copies of $H$ in $G_n$.
- $\mathbb{E} T(K_3, G_n) \to \lambda \implies \frac{N(K_3, G_n)}{c_n^2} \to \lambda$.
- $\mathrm{Var}\, T(K_3, G_n) = N(K_3, G_n)\frac{1}{c_n^2}\left(1 - \frac{1}{c_n^2}\right) + 2N(D, G_n)\frac{1}{c_n^3}\left(1 - \frac{1}{c_n}\right)$, where $D$ denotes the diamond, i.e. $C_4$ with one diagonal:

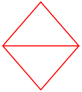

Figure: The Diamond $D$

- $\mathrm{Var}\, T(K_3, G_n) \to \lambda$ and $\mathbb{E} T(K_3, G_n) \to \lambda \implies N(D, G_n) = o(c_n^3)$.
- For each $(u, v) \in E(G_n)$, define $T(u, v)$ to be the number of vertices in $V(G_n) \setminus \{u, v\}$ that are adjacent to both $u$ and $v$, i.e.

$$T(u, v) := \sum_{w \in V(G_n)} A_{uw}(G_n) A_{vw}(G_n).$$

- We call $T(u, v)$ the co-degree of $u$ and $v$.

# Remainder Term Goes to 0 in $L_1$

- $X_v$ : Color of the vertex $v \in V(G_n)$.
- $T_\epsilon^+(K_3, G_n)$ $(\epsilon > 0) :=$
$\sum_{u,v,w \in V(G_n)} A_{uv} A_{vw} A_{uw} \mathbf{1}\left(T(u,v), T(v,w), T(u,w) \leq \epsilon c_n\right) \mathbf{1}(X_u = X_v = X_w)$.
- $\mathbb{E} T_\epsilon^-(K_3, G_n) =$
$\frac{1}{c_n^2} \sum_{u,v,w \in V(G_n)} A_{uv} A_{vw} A_{uw} \mathbf{1}\left(T(u,v) > \epsilon c_n \text{ or } T(v,w) > \epsilon c_n \text{ or } T(u,w) > \epsilon c_n\right)$.
- $\frac{1}{c_n^2} \sum_{u,v,w \in V(G_n)} A_{uv} A_{vw} A_{uw} \mathbf{1}\left(T(u,v) > \epsilon c_n\right)$

$$
\begin{aligned}
&\leq \sum_{u,v \in V(G_n)} A_{uv} \frac{T(u,v)^2}{\epsilon c_n^3} \\
&= \sum_{u,v \in V(G_n)} A_{uv} \frac{2\binom{T(u,v)}{2} + T(u,v)}{\epsilon c_n^3} \\
&\leq C \cdot \frac{2N(D,G_n) + N(K_3,G_n)}{\epsilon c_n^3} \\
&= o(1), \text{ as } n \to \infty, \text{ for a fixed } \epsilon > 0.
\end{aligned}
$$

# Main Term Goes to $\mathrm{Pois}(\lambda)$ in Distribution

- This is the more involved part of the proof, at the heart of which lies subgraph counting arguments. We sketch the main idea below.
- Define a collection $\{Z_{uvw} : u, v, w \in V(G_n), \text{ distinct}\}$ of independent $\mathrm{Ber}\left(\frac{1}{c_n^2}\right)$ random variables.
- $W_\epsilon(G_n) := \sum_{u,v,w \in V(G_n)} A_{uv} A_{vw} A_{uw} \mathbf{1}\left(T(u,v), T(v,w), T(u,w) \leq \epsilon c_n\right) Z_{uvw}$.
- **What did we show?**

$$\left|\mathbb{E} T_\epsilon^+(H, G_n)^r - \mathbb{E} W_\epsilon(G_n)^r\right| \leq C\epsilon(1 - c_n^{-\alpha})$$

  for some constants $C$ and $\alpha > 0$, not depending on $n$ or $\epsilon$.
- Hence, $\lim_{\epsilon \to 0} \lim_{n \to \infty} \left|\mathbb{E} T_\epsilon^+(H, G_n)^r - \mathbb{E} W_\epsilon(G_n)^r\right| = 0$.
- $W_\epsilon(G_n)$ converges in distribution and in all moments to $\mathrm{Pois}(\lambda)$, since it has a Binomial distribution with mean converging to $\lambda$.

# Outline


1. Introduction
2. The Second Moment Phenomenon
3. A More General Result
4. Sketch of Proof
5. Erdős-Renyi Random Graphs
6. Conclusion
7. Bibliography

# Results for Erdős-Rényi Random Graphs

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

- 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

*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, the*

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

*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


1 Introduction

2 The Second Moment Phenomenon

3 A More General Result

4 Sketch of Proof

5 Erdős-Renyi Random Graphs

6 Conclusion

7 Bibliography

# Summary of What We Showed:

- If for a connected graph $H$ and a growing sequence of graphs $G_n$, we have: $\mathbb{E}(T(H,G_n)) \to \lambda$ and $\mathrm{Var}(T(H,G_n)) \to \lambda$ as $n \to \infty$, then $T(H,G_n) \xrightarrow{d} \mathrm{Pois}(\lambda)$ (generalizing a previous result of Bhattacharya, Diaconis, Mukherjee).
- In fact, the above result is just a corollary of a more general "Poisson linear combination" convergence theorem proved by us.
- Weak convergence of $T(H,G_n)$ to $\mathrm{Pois}(\lambda)$ implies conergence of the first two moments to the corresponding moments of $\mathrm{Pois}(\lambda)$ if and only if $H$ is a star-graph.
- As an application, we derive the limiting distribution of $T(H,G_n)$ when $G_n \sim G(n,p)$. Multiple phase transitions arise as $p$ varies from 0 to 1, depending on whether the graph is balanced or not.

# Some Questions:

- What can be said about the limiting distribution of "color-1" edges in the uniform coloring setup?
- We are in the process of characterizing this limiting distribution under some additional assumptions on the graph $G_n$.
- Does the number of monochromatic / color-1 copies of connected subgraphs satisfy some large deviation principle?
- We have a positive answer for the case of color-1 subgraphs, although the large deviation variational problem is yet to be solved.
- Does the number of monochromatic copies of a connected subgraph satisfy a Central Limit Theorem under the only assumption that its expected value goes to $\infty$?
- We have a negative answer for all non-star graphs, but do not yet know what happens for star-graphs.

# Outline


1. Introduction
2. The Second Moment Phenomenon
3. A More General Result
4. Sketch of Proof
5. Erdős-Renyi Random Graphs
6. Conclusion
7. Bibliography

# Reference


BHATTACHARYA, B. B., DIACONIS, P., MUKHERJEE, S., *Universal Poisson and Normal limit theorems in graph coloring problems with connections to extremal combinatorics,* Annals of Applied Probability, Vol. 27 (1), 337-394, 2017.

BHATTACHARYA, B. B., MUKHERJEE, S., *Limit Theorems for Monochromatic 2-Stars and Triangles* arXiv:1704.04674v1, 2017.