Limit Theorems for the Distribution of Triangles of a Specific Color

Nianzu Xiong

April 2022

A THESIS

PRESENTED TO THE FACULTIES OF MATHEMATICS

AT THE UNIVERSITY OF PENNSYLVANIA

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS

FOR THE DEGREE OF

BACHELOR OF ARTS IN MATHEMATICS

WITH HONORS

Advisors: Professor Bhaswar Bhattacharya and Professor Robin Pemantle

Abstract

In a deterministic sequence of simple undirected graphs whose vertices are colored by choosing a color uniformly from c colors, we provide limit theorems related to the distribution of the number of triangles whose vertices share a specific color. In particular, we prove the Poisson approximation for this distribution as $c \to \infty$, as well as a central limit theorem. We also give applications of the theorems in real life.

Acknowledgement

I would first like to thank my advisor Bhaswar Bhattacharya for his year-long mentorship. He taught me so much about probability theory and graph theory that are beyond helpful in my research in microeconomic theory. I cannot thank him enough for introducing me to this intriguing question. I also cannot thank him enough for his patience when things pile up as I try to graduate early, as well as his encouragement when I need more confidence to move forward. I also thank Professor Robin Pemantle for generously serving as the co-supervisor of this project.

To Yuzhou Lin, I always feel inspired when talking about mathematics, statistics, and life with you.

To Sunwenxi Huang, thank you for your many encouragements and company during COVID.

To Mingyuan Ma, thank you for sharing your great passion in probability with me. I remember climbing up and down the hills at Berkeley with you while I ramble on about my thoughts on statistics and life.

1 Introduction

Given a deterministic sequence of simple undirected graphs $\{G_n\}_{n\in\mathbb{N}}$, we denote the set of vertices in G_n by $V(G_n)$ and the set of edges by $E(G_n)$. We then denote by $T(G_n)$ the set of triangles in G_n. There are c colors, and the probability that a vertex is colored with any of these colors is $\frac{1}{c}$. We assume that each vertex has degree at most D, a universal constant that does not depend on n. We study the number of triangles with a specific color, say color red, in the graph G_n. We denote this number by $N(G_n)$. If a triangle is formed with vertices i, j, k, then we denote the triangle by (i, j, k). Let Y_i, Y_j, Y_k be the colors of the respective vertices. Therefore, we have

$$N(G_n) = \sum_{(i,j,k)\in T(G_n)} \mathbf{1}\{Y_i = Y_j = Y_k = \text{red}\}.$$

The main results of this paper includes a Poisson approximation of the number of triangles as $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \lambda$, as well as a central limit theorem when $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \infty$. In addition, we discuss some real-life applications of the Poisson approximation, including generalizations of the well-known birthday problem.

2 Literature Review

In this section, we discuss some seminal results in the limit theorems of the distribution of monochromatic edges and monochromatic triangles. Bhattacharya, Diaconis, and Mukherjee (2017) study the distribution of monochromatic edges in random graphs. Let

$$N_e(G_n) = \sum_{(i,j)\in E(G_n)} \mathbf{1}\{Y_i = Y_j\}$$

be the number of monochromatic edges in G_n, they use the method of moments to show that as $c = c(n) \to \infty$,

$$N_e(G_n) \xrightarrow{\mathcal{D}} \begin{cases} 0 & \text{if } \frac{|E(G_n)|}{c} \xrightarrow{\mathcal{P}} 0 \\ \text{Poisson}(\lambda) & \text{if } \frac{|E(G_n)|}{c} \xrightarrow{\mathcal{P}} \lambda \\ \infty & \text{if } \frac{|E(G_n)|}{c} \xrightarrow{\mathcal{P}} \infty \end{cases}.$$

They also show that as $c = c(n) \to \infty$ and $\frac{|E(G_n)|}{c} \to \infty$,

$$\frac{N_e(G_n) - \frac{|E(G_n)|}{c}}{\sqrt{\frac{|E(G_n)|}{c}}} \xrightarrow{\mathcal{D}} N(0, 1).$$

They further investigate the limiting distribution of monochromatic edges when the number of colors is fixed and $|E(G_n)| \to \infty$. They show that

$$\frac{N_e(G_n) - \frac{|E(G_n)|}{c}}{\sqrt{\frac{|E(G_n)|}{c}}} \xrightarrow{\mathcal{D}} N(0, 1 - \frac{1}{c})$$

if and only if G_n is asymptotic 4-cycle free (i.e., $N(G_n, C_4) = o(|E(G_n)|^2)$).

Generalizing the results of monochromatic edges to monochromatic triangles turned out to be more computationally involved. Let

$$T_3(G_n) = \sum_{(i,j,k) \in T(G_n)} \mathbf{1}\{Y_i = Y_j = Y_k\}$$

be the number of monochromatic triangles in a deterministic graph G_n. Bhattacharya and Mukherjee (2017) show that for a sequence of determistic graphs $\{G_n\}_{n \in \mathbb{N}}$, if

$$\lim_{n \to \infty} \mathbb{E}(T_3(G_n)) = \frac{|T(G_n)|}{c^2} = \lambda$$

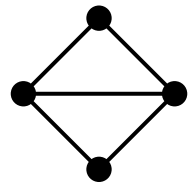
and

$$\lim_{n \to \infty} \mathrm{Var}(T_3(G_n)) = \frac{1}{c^2}(1 - \frac{1}{c^2})T(G_n) + 2(\frac{1}{c^3} - \frac{1}{c^4})N(\triangle^2, G_n) = \lambda,$$

then

$$T_3(G_n) \xrightarrow{\mathcal{D}} \text{Poisson}(\lambda).$$

In one of the expressions above, $N(\triangle^2, G_n)$ refers to the number of copies of the graph below in G_n:



a graph of $\triangle^2$

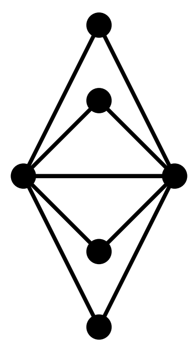
Furthermore, let us define

$$Z_3(G_n) := \frac{T_3(G_n) - \frac{|T(G_n)|}{c^2}}{\sqrt{\frac{1}{c^2}(1-\frac{1}{c^2})T(G_n) + 2(\frac{1}{c^3}-\frac{1}{c^4})N(\triangle^2, G_n)}}.$$

Bhattacharya, Fang, and Yan (2021) show that

$$\sup_{x\in\mathbb{R}} |\mathbb{P}(Z_3(G_n) \le x) - \Phi(x)| \le K \cdot \{[\frac{1+N(\triangle^4, G_n)}{(|T(G_n)| + N(\triangle^2, G_n))^2}]^{\frac{1}{4}} + \frac{b(G_n)}{(|T(G_n)| + N(\triangle^2, G_n))^2}\}^{\frac{1}{5}}.$$

In the expression above, $N(\triangle^4, G_n)$ is the number of copies of the graph below in G_n,



a graph of $\triangle^4$

and

$$b(G_n) := \sum_{1\le i\le j\le k\le l\le |V(G_n)|} (d_{ij}d_{jk}d_{kl}d_{li} + d_{ij}d_{jl}d_{lk}d_{ki} + d_{ik}d_{kj}d_{jl}d_{li}),$$

where d_{ij} is the number of triangles in G_n that uses (i, j) as an edge and other terms are defined similarly. K is a universal constant only depending on c.

A natural extension to the study of the distribution of monochromatic triangles is to analyze the limiting distribution of triangles whose vertices are all colored with a specific color. In the first theorem of this paper, we characterize the limiting distributions of $N(G_n)$ as $c \to \infty$, depending on the value of $\frac{|T(G_n)|}{c^3}$. We undertake a similar approach as the one used to derive the convergence to Poisson in the case of monochromatic edge, and then offer an alternative proof using Stein's method. Although the method of moments works fine in the case of monochromatic edges, the desired conclusion in the case of triangles of a specific color could be proved more directly using some other inequalities that involve calculating the Wasserstein distance between some random variables.

3 Main Results

Theorem 1: Given a deterministic sequence of simple undirected graphs $\{G_n\}_{n\in\mathbb{N}}$, as $c = c(n) \to \infty$, we have $N(G_n) \xrightarrow{\mathcal{D}} \begin{cases} 0 & \text{if } \frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} 0 \\ \text{Poisson}(\lambda) & \text{if } \frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \lambda \\ \infty & \text{if } \frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \infty \end{cases}$.

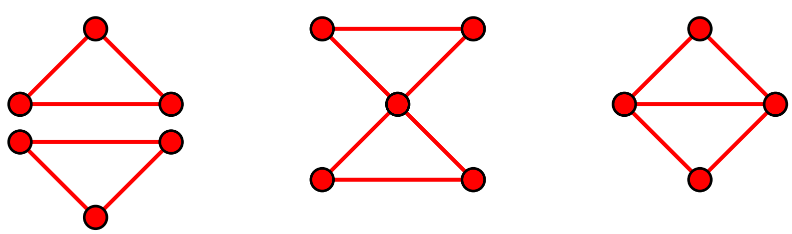
Proof. **Case 1**: $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} 0$.

Here notice

$$\mathbb{P}(N(G_n) > 0|G_n) \leq \mathbb{E}(N(G_n)|G_n) = \frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} 0.$$

Then $N(G_n) \xrightarrow{\mathcal{D}} 0$.

Case 2: $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \infty$. We begin by noticing that there are three possible relationships between any two triangles, as determined by the number of shared vertices:



no shared vertex 1 shared vertex 2 shared vertices

Relationship 1: (i_1, j_1, k_1) and (i_2, j_2, k_2) have no shared vertex (we denote such a pair as $\triangle_0$), then

$$\mathbb{E}(\mathbf{1}\{Y_{i_1} = Y_{j_1} = Y_{k_1} = \text{red}\} \cdot \mathbf{1}\{Y_{i_2} = Y_{j_2} = Y_{k_2} = \text{red}\}) = \frac{1}{c^6}.$$

Relationship 2: if (i_1, j_1, k_1) and (i_2, j_2, k_2) have 1 shared vertex (we denote such a pair as $\triangle_1$), then without loss of generality, say $i_1 = i_2$, we have

$$\mathbb{E}(\mathbf{1}\{Y_{i_1} = Y_{j_1} = Y_{k_1} = \text{red}\} \cdot \mathbf{1}\{Y_{i_2} = Y_{j_2} = Y_{k_2} = \text{red}\}) = \frac{1}{c^5}.$$

Relationship 3: if (i_1, j_1, k_1) and (i_2, j_2, k_2) have 2 shared vertices (we denote such a pair as $\triangle_2$), then say $i_1 = i_2$ and $j_1 = j_2$, we have

$$\mathbb{E}(\mathbf{1}\{Y_{i_1} = Y_{j_1} = Y_{k_1} = \text{red}\} \cdot \mathbf{1}\{Y_{i_2} = Y_{j_2} = Y_{k_2} = \text{red}\}) = \frac{1}{c^4}$$

Now, notice that for a triangle in G_n, there are at most $\frac{3D(D-1)}{2}$ triangles that share one vertex with it, and at most $3(D-2)$ triangles that share one edge with it. Therefore, we have

$$\begin{aligned}\frac{|T(G_n)|^2}{c^6} \leq \mathbb{E}(N(G_n)^2|G_n) \leq \frac{|T(G_n)|}{c^3} + 2\cdot[\frac{3(D-2)(D-3)}{2}\cdot\frac{|T(G_n)|}{c^5} + 3(D-2)\cdot\frac{|T(G_n)|}{c^4} \\ + \frac{|T(G_n)|^2 - |T(G_n)| - \frac{3(D-2)(D-3)}{2}\cdot|T(G_n)| - 3(D-2)\cdot|T(G_n)|}{c^6}].\end{aligned}$$

Then we have

$$\frac{\mathbb{E}(N(G_n)^2|G_n)}{\mathbb{E}(N(G_n)|G_n)^2} \geq \frac{\frac{|T(G_n)|^2}{c^6}}{\frac{|T(G_n)|^2}{c^6}} = 1$$

and

$$\frac{\mathbb{E}(N(G_n)^2|G_n)}{\mathbb{E}(N(G_n)|G_n)^2}$$
$$\leq \frac{\frac{|T(G_n)|}{c^3} + 2\cdot[\frac{3(D-2)(D-3)}{2}\cdot\frac{|T(G_n)|}{c^5} + 3(D-2)\cdot\frac{|T(G_n)|}{c^4} + \frac{|T(G_n)|^2 - |T(G_n)| - \frac{3(D-2)(D-3)}{2}\cdot|T(G_n)| - 3(D-2)\cdot|T(G_n)|}{c^6}]}{\frac{|T(G_n)|^2}{c^6}}$$

As $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \infty$, we have

$$\frac{\mathbb{E}(N(G_n)^2|G_n)}{\mathbb{E}(N(G_n)|G_n)^2} \xrightarrow{\mathcal{P}} 1.$$

It follows that $N(G_n) \xrightarrow{\mathcal{D}} \infty$.

Now we investigate the most interesting case.

Case 3: $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \lambda$.

In this case, let $M(G_n) = \sum_{(i,j,k)\in T(G_n)} Z_{(i,j,k)}$, where $Z_{(i,j,k)} \sim Ber(\frac{1}{c^3})$. We first show that for any $t \geq 1$, we have $|\mathbb{E}(N(G_n)^t|G_n) - \mathbb{E}(M(G_n)^t|G_n)| \xrightarrow{\mathcal{P}} 0$.

Since $N(G_n) = \sum_{(i,j,k)\in T(G_n)} \mathbf{1}\{Y_i = Y_j = Y_k = \text{red}\}$, by multinomial expansion,

$$\mathbb{E}(N(G_n)^t|G_n) = \sum_{(i_i,j_1,k_1)} \sum_{(i_2,j_2,k_2)} \dots \sum_{(i_t,j_t,k_t)} \mathbb{E}(\prod_{m=1}^{t} \mathbf{1}\{Y_{i_m} = Y_{j_m} = Y_{k_m} = \text{red}\}).$$

We also have

$$\mathbb{E}(M(G_n)^t|G_n) = \sum_{(i_i,j_1,k_1)} \sum_{(i_2,j_2,k_2)} \dots \sum_{(i_t,j_t,k_t)} \mathbb{E}(\prod_{m=1}^{t} Z_{(i_m,j_m,k_m)}).$$

We start by considering the case where $t = 1$. Here we have

$$\mathbb{E}(N(G_n)|G_n) = \sum_{(i_i,j_1,k_1)} \mathbb{E}(\mathbf{1}\{Y_{i_1} = Y_{j_1} = Y_{k_1} = 1\}) = \frac{1}{c^3} \cdot |T(G_n)|,$$

and then

$$\mathbb{E}(M(G_n)|G_n) = \sum_{(i_i,j_1,k_1)} \mathbb{E}(Z_{(i_1,j_1,k_1)}) = \frac{1}{c^3} \cdot |T(G_n)| = \mathbb{E}(N(G_n)|G_n),$$

as desired.

Now we choose t triangles, and denote the resulting subgraph by H. We could without loss of generality view H as a simple graph, because if a triangle (i_m, j_m, k_m) is chosen more than once, its contribution to the two expressions below is equivalent to only being chosen once. Then we have

$$\mathbb{E}(\prod_{m=1}^{t} \mathbf{1}\{Y_{i_m} = Y_{j_m} = Y_{k_m} = \text{red}\}) = \frac{1}{c^{|V(H)|}},$$

and

$$\mathbb{E}(\prod_{m=1}^{t} Z_{(i_m,j_m,k_m)}) = \frac{1}{c^{3t}}.$$

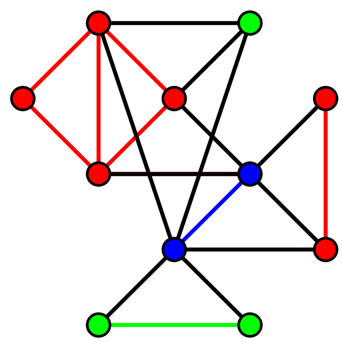
Notice $|V(H)| \leq 3t$ and $|V(H)| = 3t$ if and only if none of the triangles share any vertex.

Therefore,

$$|\mathbb{E}(N(G_n)^t|G_n) - \mathbb{E}(M(G_n)^t|G_n)| \lesssim_t \sum_{H \in H_t} N(H, G_n) \cdot (\frac{1}{c^{|V(H)|}} - \frac{1}{c^{3t}}),$$

where H_t is the set of subgraphs of G_n with at most t triangles, and $N(H, G_n)$ is the number of copies of H in G_n.

Let us denote the number of triangle-connected components of the graph H by $Tc(H)$. For a given triangle, triangles in the same triangle-connected component as itself is the largest set of triangles in G_n such that the graph formed by them together with the given triangle (denoted G) satisfies $V(G) = T(G) + 2$. For example, the graph below has one connected component but three triangle-connected components. In this graph, there are two red triangles.



Notice that $N(H, G_n) \lesssim_D |T(G_n)|^{Tc(H)}$, because within each triangle-connected component, once we fix one triangle, the total number of ways of choosing triangles in that triangle-connected component is merely a function of D.

We now consider the expression $|T(G_n)|^{Tc(H)} \cdot (\frac{1}{c^{|V(H)|}} - \frac{1}{c^{3t}})$. Since each triangle-connected component must have at least one triangle and thus three vertices, $3Tc(H) \leq |V(H)|$.

If any of the triangles exhibit relationship 2 or 3 as described above, we have $3Tc(H) < |V(H)|$. Then $3Tc(H) = |V(H)|$ if and only if all t triangles have relationship as in case 1.

Therefore, if all t triangles have relationship 1, then $|V(H)| = 3t$, and so $\frac{1}{c^{3t-|V(H)|}} - 1 = 0$, which means $\frac{|T(G_n)|^{Tc(H)}}{c^{3t}} \cdot (\frac{1}{c^{|V(H)|-3t}} - 1) = 0$. If any pair of triangles has relationship 2 or 3, then $3Tc(H) < |V(H)| < 3t$, and we conclude that $|\frac{|T(G_n)|^{Tc(H)}}{c^{3t}} \cdot (\frac{1}{c^{|V(H)|-3t}} - 1)| \xrightarrow{\mathcal{P}} 0$. Then

$$|\mathbb{E}(N(G_n)^t|G_n) - \mathbb{E}(M(G_n)^t|G_n)| \xrightarrow{\mathcal{P}} 0,$$

as desired.

It is clear that $M(G_n) \xrightarrow{\mathcal{D}} \text{Poisson}(\lambda)$ as $c \to \infty$. Now we introduce a lemma that would enable us to use the method of moments to arrive at the desired conclusion.

Lemma 1: (Bhattacharya, Diaconis, and Mukherjee (2017)) Take a sequence of probability spaces $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$ together with a sequence $\mathcal{G}_n \subset \mathcal{F}_n$ and a sequence of random variables (X_n, Y_n) on that sequence of probability spaces. If for all $k > 0$,

$$\limsup_{n\to\infty} \mathbb{P}(|\mathbb{E}(N(G_n)^k|G_n) - \mathbb{E}(M(G_n)^k|G_n)| > \varepsilon) = 0,$$

and there exists a $m > 0$ such that

$$\limsup_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(|\frac{m^k}{k!}\mathbb{E}(M(G_n)^k|G_n)| > \varepsilon) = 0,$$

then for any $t \in \mathbb{R}$, we have

$$\mathbb{E}(e^{itX_n}) - \mathbb{E}(e^{itY_n}) \to 0.$$

Now we check the property mentioned in the lemma above. We want to show that for any $\varepsilon > 0$ and $t \in \mathbb{R}$, we have

$$\limsup_{k\to\infty} \limsup_{n\to\infty} \mathbb{P}(|\frac{t^k}{k!}\mathbb{E}(M(G_n)^k|G_n)| > \varepsilon) = 0.$$

To see this, notice that

$$\mathbb{E}(M(G_n)^k|G_n) = \sum_{i=0}^{k} \begin{Bmatrix} \text{k} \\ \text{t} \end{Bmatrix} \cdot \binom{|T(G_n)|}{i} \cdot c^{-3i},$$

where

$$\begin{Bmatrix} \text{k} \\ \text{t} \end{Bmatrix} := \frac{1}{t!} \cdot \sum_{m=0}^{t} (-1)^m \binom{t}{m} \cdot (t-m)^k$$

is the Stirling number of the second kind. Since $\frac{t^k}{k!} \sum_{i=0}^{k} \begin{Bmatrix} \text{k} \\ \text{t} \end{Bmatrix} \cdot \lambda^i$ converges to 0 almost surely as $k \to \infty$, the desired conclusion follows from the fact that

$$\limsup_{k\to\infty}\limsup_{n\to\infty}\mathbb{P}(|\frac{t^k}{k!}\mathbb{E}(M(G_n)^k|G_n)|>\varepsilon)$$
$$\leq \limsup_{k\to\infty}\limsup_{n\to\infty}\frac{1}{\varepsilon}\cdot\frac{t^k}{k!}\mathbb{E}(M_n^k|G_n)$$

by Markov's inequality. ■

Alternatively, the case where $\frac{|T(G_n)|}{c^3}\xrightarrow{\mathcal{P}}\lambda$ could be proved using the method of dependency graphs.

Definition 1: Given $\{X_i|i\in V\}$ as a collection of random variables, $G=(V,E)$ is called a dependency graph for $\{X_i|i\in V\}$ if for any two subsets of vertices $S_1,S_2\subseteq V$ such that there exists no edge from any vertex in S_1 to any vertex in S_2, the collections $\{X_i|i\in S_1\}$ and $\{X_i|i\in S_2\}$ are independent.

Now, given a graph G_n, we label its triangles as $1,2,...,|T(G_n)|$, with the associated vertices of the i^{th} triangle being v_{i_1},v_{i_2},v_{i_3}, and then define

$$X_i:=\mathbf{1}\{Y_{i_1}=Y_{i_2}=Y_{i_3}=\text{red}\}.$$

Let us associate each triangle in G_n with a vertex, and two vertices are connected by an edge if and only if their respective associated triangles share at least one vertex. It is easy to verify that the resulting graph G_d is a dependency graph.

Definition 2: Given two probability measures μ and v, the total variation distance between them is

$$TV(\mu,v):=\sup_{A\in\Omega}|\mu(A)-v(A)|,$$

where Ω is a measurable space.

Lemma 2: (Chatterjee, Diaconis, and Meckes (2005)) Given a dependency graph $G_d=(V(G_d),E(G_d))$, for a vertex $i\in V(G_d)$, Let $N_i:=\{j\in V(G_d)|(i,j)\in E(G_d)\}\cup\{i\}$. Then

$$TV(N(G_n)-\text{Poisson}(\lambda))\leq\min\{1,\frac{1}{\lambda}\}\cdot(\sum_{i\in V(G_d)}\sum_{j\in N_i\setminus\{i\}}p_{ij}+\sum_{i\in V(G_d)}\sum_{j\in N_i}p_ip_j),$$

where $p_i=\mathbb{P}(X_i=1)$, and $p_{ij}=\mathbb{P}(X_i=1,X_j=1)$. In this case, $p_i=\frac{1}{c^3}$ for all $i\in V(G_d)$, $p_{ij}=\frac{1}{c^5}$ if the triangles associated with vertices i and j are in relationship 2, and $\frac{1}{c^4}$ if the triangles are in relationship 3.

As discussed above, if we take a triangle in G_n, there are at most $\frac{3D(D-1)}{2}$ triangles that share one vertex with it, and at most $3(D-2)$ triangles that share one edge with it. Therefore, we have

$$\sum_{i\in V(G_d)}\sum_{j\in N_i\setminus\{i\}} p_{ij} \leq \frac{3D(D-1)}{2}\cdot\frac{|T(G_n|}{c^5} + 3(D-2)\cdot\frac{|T(G_n)|}{c^4},$$

and

$$\sum_{i\in V(G_d)}\sum_{j\in N_i} p_i p_j \leq (\frac{3D(D-1)}{2} + 3(D-2))\cdot\frac{|T(G_n)|}{c^6}.$$

Then

$$\begin{aligned} TV(N(G_n)-\text{Poisson}(\lambda)) &\leq \min\{1,\frac{1}{\lambda}\}\cdot(\sum_{i\in V(G_d)}\sum_{j\in N_i\setminus\{i\}} p_{ij} + \sum_{i\in V(G_d)}\sum_{j\in N_i} p_i p_j) \\ \leq \frac{3D(D-1)}{2}\cdot\frac{|T(G_n)|}{c^5} &+ 3(D-2)\cdot\frac{|T(G_n)|}{c^4} + (\frac{3D(D-1)}{2} + 3(D-2))\cdot\frac{|T(G_n)|}{c^6}. \end{aligned}$$

Notice that as $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \lambda$ and $c\to\infty$, we have

$$\frac{|T(G_n)|}{c^4} \xrightarrow{\mathcal{P}} \frac{\lambda}{c} \xrightarrow{\mathcal{P}} 0,$$
$$\frac{|T(G_n)|}{c^5} \xrightarrow{\mathcal{P}} \frac{\lambda}{c^2} \xrightarrow{\mathcal{P}} 0,$$

and

$$\frac{|T(G_n)|}{c^6} \xrightarrow{\mathcal{P}} \frac{\lambda}{c^3} \xrightarrow{\mathcal{P}} 0,$$

So we have

$$TV(N(G_n)-\text{Poisson}(\lambda)) \xrightarrow{\mathcal{P}} 0.$$

This implies the weak convergence of $N(G_n)$ to Poisson(λ).

Example 1: Generalized Birthday Problems in the US Population

The classic birthday problem is the following: in a group of n people, what is the probability that some of them share the same birthday? The classic result shows that in a group of 23 people, there is about a 50% chance that at least two of them would share the same birthday.

There has been many generalizations of this problem. Bhattacharya et al. (2016) consider the following generalization: What is the probability that two people in the US who know each other have the same birthday, their fathers have the same birthday, their grandfathers have the same birthday, and their great-grandfathers have the same birthday? Using the Poisson approximation for monochromatic edges while modeling people in the US with an Erdös-Renyi graph, they find that the chance is about 99.8%.

Now we give another generalization of the birthday problem that could be analyzed with the Poisson approximation for triangles with a specific color. Let us consider the following question: What is the chance that there are three people in the US who would meet each other in their lives and are all born on January 1^{st}, 2000 (i.e., the first day of the 21^{st} century)?

We could approximate this probability with Theorem 1. There are currently about 330 million people in the US, and according to a popular estimate, a person typically would meet 80,000 people in his/her life. Therefore, when we model the scenario with an Erdös-Renyi graph, we have

$$p = \frac{80000}{330 \cdot 10^6},$$

and

$$\mathbb{E}(|T(G_n)|) = \sum_{Y_i, Y_j, Y_k \in V(G_n)} \mathbb{E}[\mathbf{1}\{(i,j,k) \in T(G_n)\}] = \binom{330 \cdot 10^6}{3} \cdot (\frac{80000}{330 \cdot 10^6})^3.$$

There are $c = 80 \cdot 365$ "colors," and

$$\lambda = \frac{\mathbb{E}(|T(G_n)|)}{c^3} = \frac{\binom{330 \cdot 10^6}{3} \cdot (\frac{80000}{330 \cdot 10^6})^3}{(80 \cdot 365)^3} \approx 3.427.$$

Then

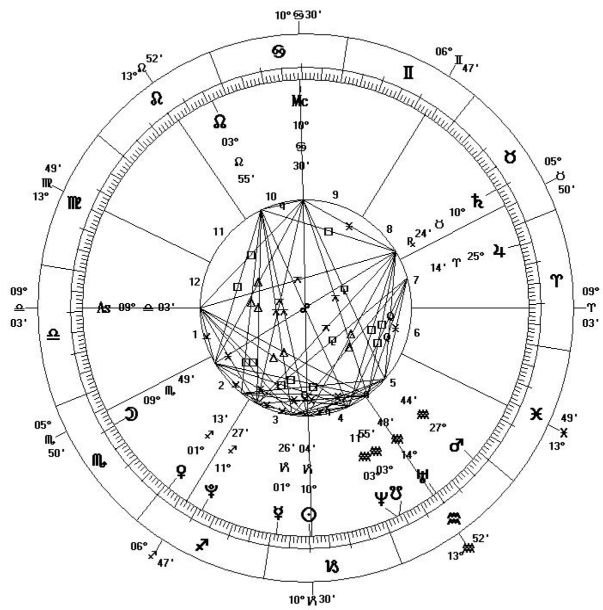
$$\mathbb{P}(N(G_n) > 0) \approx 1 - e^{-3.427} \approx 96.75\%.$$

Therefore, this seemingly exotic event is almost surely bound to take place.

Example 2: Coincidences in Horoscopes

Let us consider another problem. Horoscopes are charts used in classic Western astrology constructed using the time and location of one's birth. It contains information such as the positions of the Sun, the Moon, as well as other planets when the person is born. There are twelve signs and twelve houses involved. The twelve signs are Aries, Taurus, Gemini, Cancer, Leo, Virgo, Libra, Scorpio, Sagittarius, Capricorn, Aquarius, and Pisces. The twelve houses are of self, value, sharing, home, pleasure, health, balance, transformation, purpose, enterprise, blessings, and sacrifice respectively.

In astrological theory, the Sun is the most important influence in a horoscope, and it determines one's personality; the Moon is the second most important, and it manifests one's emotions. There are $(12^2)\cdot(12^2) = 20736$ different Sun-Moon combinations considering the respective houses. Below is a picture of a horoscope for a person who is born on January 1^{st}, 2000 at 12:00:01 am in New York. This hypothetical person would have the Sun in Capricorn and in the third house, and the Moon in the Scorpio and in the second house.



Source: Wikipedia

We now consider the following question: What is the probability that there are three people in New York city who would meet each other in their lives, and share the same Sun-Moon combination in their horoscopes as the person described above?

Here p and $\mathbb{E}(|T(G_n)|)$ are same as above, while $c = 20736$ and

$$\lambda = \frac{\mathbb{E}(|T(G_n)|)}{c^3} = \frac{\binom{8.85\cdot 10^6}{3}\cdot\left(\frac{80000}{8.85\cdot 10^6}\right)^3}{(20736)^3} \approx 9.57.$$

As a result,

$$\mathbb{P}(N(G_n) > 0) \approx 1 - e^{-9.57} \approx 1.$$

Therefore, if every New Yorker is keen on visiting astrologists, then the event that three of them who would meet each other in their lives and be deemed as having almost identical personalities and emotional states as the person born on the first second of the new century is also almost bound to take place.

The two examples above could be viewed as applications of the law of truly large numbers described by Diaconis and Mosteller (1989), which says that almost any outrageous and exotic event is bound to occur when the interactions among an enormous number of people accumulate over time.

Now we give the second main theorem of the paper.

Theorem 2: As $c = c(n) \to \infty$ and $\frac{|T(G_n)|}{c^3} \to \infty$, we have $\frac{N(G_n)-\frac{|T(G_n)|}{c^3}}{\sqrt{\frac{|T(G_n)|}{c^3}}} \xrightarrow{\mathcal{D}} N(0,1)$.

Proof. We prove this result using the method of dependency graphs. Firstly, we construct the dependency graph in the same way as described in the alternative proof of theorem 1, while modifying the definition of X_i into $\tilde{X}_i := \mathbf{1}\{Y_{i_1} = Y_{i_2} = Y_{i_3} = \text{red} - \frac{1}{c^3}\}$. Now we introduce another distance between probability measures.

Definition 3: The Wasserstein distance between two probability measures μ and v on R is

$$\text{Wass}(\mu, v) := \sup\{|\int f\, d\mu - \int f\, dv| \mid f \text{ is 1-Lipschitz}\}.$$

We now give an important lemma for the proof of Theorem 2.

Lemma 3: (Chatterjee 2007) Given a dependency graph $G_d = (V(G_d), E(G_d))$, if $\mathbb{E}(\tilde{X}_i) = 0, W = \frac{\sum_i X_i}{\sqrt{\text{Var}(\sum_i \tilde{X}_i)}}$, and $Z \sim N(0,1)$, then we have

$$\text{Wass}(W, Z) \le \frac{4}{\sqrt{\pi} \cdot \text{Var}(\sum_i \tilde{X}_i)} \cdot \sqrt{\bar{D}^3 \sum_i \mathbb{E}|\tilde{X}_i|^4} + \frac{\bar{D}^2}{(\text{Var}(\sum_i \tilde{X}_i))^{\frac{3}{2}}} \cdot \sum_i \mathbb{E}|\tilde{X}_i|^3,$$

where $\bar{D} = 1 + \max_{v \in V(G_d)} \deg(v)$.

As mentioned above, given a triangle in G_n, there are at most $\frac{3D(D-1)}{2}$ triangles that share one vertex with it, and at most $3(D-2)$ triangles that share one edge with it. So we have

$$\bar{D} \le 1 + \frac{3(D-2)(D-3)}{2} + 3(D-2) < \infty.$$

Now, notice

$$\begin{aligned}\mathrm{Var}(\sum_i \tilde{X}_i) &= \mathrm{Var}(\sum_{(i,j,k)\in T(G_n)} \{\mathbf{1}\{Y_i = Y_j = Y_k = \text{red}\} - \frac{1}{c^3}\}) \\ &= \mathrm{Var}(\sum_{(i,j,k)\in T(G_n)} \mathbf{1}\{Y_i = Y_j = Y_k = \text{red}\}) \\ &= \mathrm{Var}(N(G_n)).\end{aligned}$$

We have

$$\mathbb{E}(N(G_n)^2|G_n) = \frac{|T(G_n)|}{c^3} + 2\cdot[\frac{N(\triangle_0, G_n)}{c^6} + \frac{N(\triangle_1, G_n)}{c^5} + \frac{N(\triangle_2, G_n)}{c^4}]$$

and

$$\mathbb{E}(N(G_n)|G_n)^2 = \frac{|T(G_n)|^2}{c^6}.$$

Then Here $N(\triangle_0, G_n)$ is the number of unordered pairs of triangles in relationship 1, and $N(\triangle_1, G_n)$ and $N(\triangle_2, G_n)$ are defined similarly. Then

$$\begin{aligned}\mathrm{Var}(N(G_n)) &= \mathbb{E}(N(G_n)^2|G_n) - \mathbb{E}(N(G_n)|G_n)^2 \\ &= \frac{|T(G_n)|}{c^3} + 2\cdot[\frac{N(\triangle_0, G_n)}{c^6} + \frac{N(\triangle_1, G_n)}{c^5} + \frac{N(\triangle_2, G_n)}{c^4}] - \frac{|T(G_n)|^2}{c^6} \\ &\geq \frac{|T(G_n)|}{c^3}\cdot(1 - \frac{1}{c^3}).\end{aligned}$$

Moreover, as $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \to \infty$, we have

$$\sum_{i\in\{1,\ldots,|T(G_n)|\}} \mathbb{E}|\tilde{X}_i|^4 = |T(G_n)|\cdot[(1-\frac{1}{c^3})^4\cdot\frac{1}{c^3} + (\frac{1}{c^3})^4\cdot(1-\frac{1}{c^3})] < 2\cdot\frac{|T(G_n)|}{c^3},$$

and

$$\sum_{i\in\{1,\ldots,|T(G_n)|\}} \mathbb{E}|\tilde{X}_i|^3 = |T(G_n)|\cdot[(1-\frac{1}{c^3})^3\cdot\frac{1}{c^3} + (\frac{1}{c^3})^3\cdot(1-\frac{1}{c^3})] < 2\cdot\frac{|T(G_n)|}{c^3}.$$

Plugging these results into the original expression, we have

$$
\begin{aligned}
\mathrm{Wass}(W,Z) &\le \frac{4}{\sqrt{\pi}\cdot \mathrm{Var}(\sum_i \tilde{X}_i)}\cdot\sqrt{\bar{D}^3\sum_i \mathbb{E}|\tilde{X}_i|^4}+\frac{\bar{D}^2}{(\mathrm{Var}(\sum_i \tilde{X}_i))^{\frac{3}{2}}}\cdot\sum_i \mathbb{E}|\tilde{X}_i|^3\\
&\le \frac{4}{\sqrt{\pi}\cdot\frac{|T(G_n)|}{c^3}\cdot(1-\frac{1}{c^3})}\cdot\sqrt{(1+\frac{3(D-2)(D-3)}{2}+3(D-2))^3\cdot 2\cdot\frac{|T(G_n)|}{c^3}}\\
&\qquad +\frac{(1+\frac{3D(D-1)}{2}+D-2)^2}{(\frac{|T(G_n)|}{c^3}\cdot(1-\frac{1}{c^3}))^{\frac{3}{2}}}\cdot 2\cdot\frac{|T(G_n)|}{c^3}\\
&= \frac{\sqrt{\frac{|T(G_n)|}{c^3}}}{\frac{|T(G_n)|}{c^3}}\cdot\frac{4}{\sqrt{\pi}\cdot(1-\frac{1}{c^3})}\cdot\sqrt{2(1+\frac{3(D-2)(D-3)}{2}+3(D-2))^3}\\
&\qquad +2\frac{\frac{|T(G_n)|}{c^3}}{(\frac{|T(G_n)|}{c^3})^{\frac{3}{2}}}\frac{(1+\frac{3(D-2)(D-3)}{2}+3(D-2))^2}{(1-\frac{1}{c^3})^{\frac{3}{2}}}\xrightarrow{P}0.
\end{aligned}
$$

Lemma 4 (Villani (2009)) addresses the following property:

Given a Polish space (X,d), if (μ_k) is a sequence of measures in $P(X)$ and μ is another measure, then the following two statements are equivalent:

(i) $\mathrm{Wass}(\mu_k,\mu)\to 0$

(ii) μ_k converges weakly to μ in $P(X)$.

Under this lemma, we immediately have that

$$\mathrm{Wass}(W,Z)\xrightarrow{P}0$$

implies

$$W\xrightarrow{\mathcal{D}}N(0,1).$$

Now, if we define

$$\tilde{W}:=\frac{N(G_n)-\frac{|T(G_n)|}{c^3}}{\sqrt{\frac{|T(G_n)|}{c^3}}},$$

We have

$$W-\tilde{W}=(N(G_n)-\frac{|T(G_n)|}{c^3})\cdot(\frac{1}{\sqrt{\frac{|T(G_n)|}{c^3}+2\cdot[\frac{N(\triangle_0,G_n)}{c^6}+\frac{N(\triangle_1,G_n)}{c^5}+\frac{N(\triangle_2,G_n)}{c^4}]-\frac{|T(G_n)|^2}{c^6}}}-\frac{1}{\sqrt{\frac{|T(G_n)|}{c^3}}})$$

Then

$$W - \tilde{W} \leq (N(G_n) - \frac{|T(G_n)|}{c^3}) \cdot (\frac{1}{\sqrt{\frac{|T(G_n)|}{c^3} \cdot (1 - \frac{1}{c^3})}} - \frac{1}{\sqrt{\frac{|T(G_n)|}{c^3}}})$$
$$= \tilde{W} \cdot (\frac{1}{\sqrt{1 - \frac{1}{c^3}}} - 1) \xrightarrow{\mathcal{P}} 0$$

and

$$W - \tilde{W} \geq (N(G_n) - \frac{|T(G_n)|}{c^3})$$
$$\cdot(\frac{1}{\sqrt{\frac{|T(G_n)|}{c^3} + 2 \cdot [\frac{3(D-2)(D-3)}{2} \cdot \frac{|T(G_n)|}{c^5} + 3(D-2) \cdot \frac{|T(G_n)|}{c^4} - \frac{|T(G_n)| + \frac{3(D-2)(D-3)}{2} \cdot |T(G_n)| + 3(D-2) \cdot |T(G_n)|}{c^6}]}} - \frac{1}{\sqrt{\frac{|T(G_n)|}{c^3}}})$$
$$= \tilde{W} \cdot (\frac{1}{\sqrt{1 + \frac{3(D-2)(D-3)}{c^2} + \frac{3(D-2)}{c} - \frac{1 + \frac{3(D-2)(D-3)}{2} + 3(D-2)}{c^3}}} - 1) \xrightarrow{\mathcal{P}} 0$$

From these two inequalities above, it is clear that as $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \to \infty$, we have $W - \tilde{W} \xrightarrow{P} 0$. Therefore, we conclude that

$$\tilde{W} = \frac{N(G_n) - \frac{|T(G_n)|}{c^3}}{\sqrt{\frac{|T(G_n)|}{c^3}}} \xrightarrow{D} N(0, 1).$$

■

4 Conclusion

In this paper, we gave two main results on the limiting distribution of the number of triangles with a specific color when the degree of each vertex is bounded. The first is that as $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \lambda$, we have $N(G_n) \xrightarrow{\mathcal{D}} \text{Poisson}(\lambda)$. The second is a Central Limit Theorem for $N(G_n)$ when $c \to \infty$ and $\frac{|T(G_n)|}{c^3} \xrightarrow{\mathcal{P}} \infty$.

Future works might want to consider the Central Limit Theorem when the number of colors is fixed, which could be viewed as a variation of the Central Limit Theorem proved by Bhattacharya, Fang, and Yan (2021). It might also be interesting to consider the limiting distribution when the bounded degree assumption is relaxed. Finally, another direction might be to study the limiting distribution of other geometric structures (e.g., quadrilaterals and hexagons).

5 Reference

Bhattacharya, Bhaswar B., and Sumit Mukherjee. "Limit Theorems for Monochromatic 2-Stars and Triangles." 2017.

Bhattacharya, Bhaswar B., Persi Diaconis, and Sumit Mukherjee. "Universal Limit Theorems in Graph Coloring Problems with Connections to Extremal Combinatorics." *The Annals of Applied Probability* 27, no. 1 (2017). doi:10.1214/16-aap1205.

Bhattacharya, Bhaswar B., Xiao Fang, and Han Yan. "Normal Approximation and Fourth Moment Theorems for Monochromatic Triangles." *Random Structures & Algorithms* (2021): 25-53. doi:10.1002/rsa.21017.

Chatterjee, Sourav, Persi Diaconis, and Elizabeth Meckes. "Exchangeable Pairs and Poisson Approximation." *Probability Surveys* (2005). doi:10.1214/154957805100000096.

Chatterjee, Sourav. "Stein's Method and Applications." 2007.

"Horoscope." Wikipedia. April 11, 2022. https://en.wikipedia.org/wiki/Horoscope.

Villani, Cedric. Optimal Transport: Old and New. Berlin: *Springer*, 2009.

Vital, Anna. "Counting the People Your Life Impacts - Infographic." Adioma. August 10, 2017. https://blog.adioma.com/counting-the-people-you-impact-infographic/.