

Detecting High-Dimensional Geometric Structure in Random Graphs

Henry Smith

Yale University

December 13, 2021

Abstract

The paper *Testing for High-Dimensional Geometry in Random Graphs* by Bubeck and colleagues considers the hypothesis testing problem of whether or not a random graph G on n vertices has some underlying high-dimensional structure. Under the “no structure” regime, the graph is distributed according to the Erdős–Rényi standard model, whereas under the alternative its edges are determined by n independent random vectors that are uniformly distributed on \mathbb{S}^{d-1} . The authors explore two test statistics, the number of triangles and signed triangles in G , and show that the latter is better suited for higher-dimensional problems. In our report, we present a measure-theoretic examination of two key results from the paper; these results study the distribution of the two aforementioned statistics under H_0 and H_1 .

1 Introduction

Prior to digging into the results we will explore, we present an overview of the problem considered in [1]. Bubeck and colleagues motivate their work by contemporary ambitions in the fields of computational biology and social network theory to understand structure in high-dimensional graphs. Particularly, they consider the question of whether latent geometrical structure exists altogether in such graphs. In the context of graph theory, the term ‘structure’ refers to the edges connecting vertices in a given graph G . Under a regime in which a graph has no geometric structure, the presence or absence of edges between vertices would be purely random. Alternatively, a graph with underlying structure would have some pattern to its edges. This paradigm of structure versus no structure lends itself to a hypothesis testing problem. In the remainder of this section, we discuss the hypotheses and test statistics presented in [1] that are relevant to our paper.

1.1 Hypothesis Testing

Bubeck and colleagues test the null hypothesis that a random graph G defined on n vertices has been generated according to the *Erdős–Rényi standard model*. That is, we have

$$H_0 : G \sim G(n, p). \quad (1)$$

Under this model, each pair of distinct vertices $i, j \in [n]$ is connected by an edge independently with probability p . This null hypothesis represents the aforementioned “no structure” paradigm since the presence or absence of each edge in G has no dependence on the remaining edges in G .

As for the alternative hypothesis, the authors consider random graph G on n vertices distributed as

$$H_1 : G \sim G(n, p, d), \quad (2)$$

where $G(n, p, d)$ is defined as follows. Let $X_1, \dots, X_n \in \mathbb{R}^d$ be independent random vectors that are uniformly distributed on \mathbb{S}^{d-1} . Then distinct vertices $i, j \in [n]$ are connected by an edge if and only if $\langle X_i, X_j \rangle \geq t_{p,d}$, where $\langle \cdot, \cdot \rangle$ denotes the standard inner product in \mathbb{R}^d . Here $t_{p,d} \in [-1, 1]$ is a constant in $p \in (0, 1)$ and $d \in \mathbb{N}$ satisfying $\mathbb{P}(\langle X_i, X_j \rangle \geq t_{p,d}) = p$. Note that $\langle X_i, X_j \rangle = \cos(\theta)$, where θ is the angle between X_i, X_j in \mathbb{R}^d . Thus, we can say that vertices i and j are connected by an edge if and only if the angle between random vectors X_i and X_j is sufficiently small.

1.2 The Uniform Distribution on \mathbb{S}^{d-1}

Our previous statement that $X_1, \dots, X_n \in \mathbb{R}^d$ are independent random vectors, uniformly-distributed on \mathbb{S}^{d-1} is admittedly lacking in mathematical rigor. More specifically, each X_i is a random variable that maps from latent probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}), Q)$. Here, Q denotes the image measure of \mathbb{P} under X_i and is equal to the Lebesgue measure on \mathbb{S}^{d-1} . For the remainder of our paper, we refer to Q as the “spherical measure” under consideration.

For a more formal definition of the spherical measure Q , [2] proves that there exists a unique Borel measure σ_{d-1} on \mathbb{S}^{d-1} such that for each $f \in \mathcal{M}^+(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, then $\mu(f) = (\rho_d \times \sigma_{d-1})(f)$.¹ Here, μ denotes the Lebesgue measure on \mathbb{R}^d and ρ_d is the measure on $((0, \infty), \mathcal{B}(0, \infty))$ having density $\Delta(r) = r^{d-1}$ with respect to Lebesgue measure. That is, the d -dimensional Lebesgue measure may be expressed as the product measure of ρ_n defined on measurable space $((0, \infty), \mathcal{B}(0, \infty))$ and σ_{d-1} defined on measurable space $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$. With this Borel measure on \mathbb{S}^{d-1} , we can equivalently define $Q = \frac{\sigma}{\sigma(\mathbb{S}^{d-1})} = \frac{\Gamma(n/2)}{2\pi^{d/2}} \sigma$ to be our desired spherical measure [2].²

¹This result is subsequently extended to $f \in \mathcal{L}^1(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d), \mu)$ by splitting f into f^+ and f^- .

²For further details about measures on \mathbb{S}^{d-1} , one should reference section 2.7, “Integration in Polar Coordination”, of [2].

1.3 Test Statistics

Having established the distribution of G under both H_0 and H_1 , we present two statistics to test H_0 against H_1 . Define the *adjacency matrix* $A \in \mathbb{R}^{n \times n}$ of graph G such that

$$A_{i,j} = \begin{cases} 1 & \text{there exists an edge between vertices } i \text{ and } j \\ 0 & \text{otherwise} \end{cases}$$

As Bubeck and collaborators note in their work, the number of triangles present in random graph G can be used to uncover its latent geometric structure.³ Accordingly, the authors consider the *triangle statistic*

$$T(G) := \sum_{\{i,j,k\} \in \binom{[n]}{3}} A_{i,j} A_{i,k} A_{j,k} \quad (3)$$

as well as the *signed triangle statistic*

$$\tau(G) := \sum_{\{i,j,k\} \in \binom{[n]}{3}} (A_{i,j} - p)(A_{i,k} - p)(A_{j,k} - p) \quad (4)$$

for testing the hypotheses. Clearly, $T(G)$ is doing no more than counting the number of triangles present in G . The signed triangle statistic, on the other hand, aims to reduce the variance of T under the null hypothesis H_0 . In particular, a key result from [1] is that $\text{Var}(\tau(G(n,p)))$ is on the order of n^3 , whereas $\text{Var}(T(G(n,p)))$ is on the order of n^4 ; we prove the former statement in section 3.1. This fact is then used in Theorem 2 to prove that the signed triangle statistic is asymptotically powerful so long as $d \ll n^3$. The meaning of “asymptotically powerful” is discussed in section 3 of our report.

2 Expected Number of Triangles in $G(n, p, d)$

The first result in [1] that we consider is Lemma 1, which establishes a lower bound on $\mathbb{E}T(G(n, p, d))$, the expected value of the triangle statistic for graph G under H_1 .

In order to prove this result, the authors define the event

$$E := \left\{ \langle X_1, X_2 \rangle \geq t_{p,d}, \langle X_1, X_3 \rangle \geq t_{p,d}, \langle X_2, X_3 \rangle \geq t_{p,d} \right\}$$

for independent random vectors X_1, X_2, X_3 uniformly distributed on \mathbb{S}^{d-1} .

Note that the map $f(\omega) := \langle X_i(\omega), X_j(\omega) \rangle$ is, in fact, $\mathcal{F} \setminus \mathcal{B}([-1, 1])$ -measurable. To understand why this is the case, recall that each of $X_i, X_j : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}))$ is measurable, and so $g : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}) \otimes \mathcal{B}(\mathbb{S}^{d-1}))$ defined $g(\omega) = (X_i(\omega), X_j(\omega))$ is measurable. Further, the inner product mapping $h : (\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1})) \rightarrow ([-1, 1], \mathcal{B}([-1, 1]))$ defined $h(x, y) = \langle x, y \rangle$ is continuous. All together, this gives us that $f = h \circ g$, the composition of a continuous map with a measurable map is indeed $\mathcal{F} \setminus \mathcal{B}([-1, 1])$ -measurable. This is important because it gives us that $\{\langle X_i, X_j \rangle \geq t_{p,d}\} \in \mathcal{F}$, and so $E \in \mathcal{F}$.

To understand why we define this event E , recall from (3) that

$$T(G(n, p, d)) = \sum_{\{i,j,k\} \in \binom{[n]}{3}} A_{i,j} A_{i,k} A_{j,k},$$

³See p. 5 for a more rigorous justification of this statement. Loosely-speaking, if under $G \sim G(n, p, d)$ there exists edges between vertices i and j as well as i and k then, by the triangle inequality, there is more likely to be an edge between j and k .

and so

$$\begin{aligned}
\mathbb{E}T(G(n, p, d)) &= \mathbb{E} \left(\sum_{\{i,j,k\} \in \binom{[n]}{3}} A_{i,j} A_{i,k} A_{j,k} \right) \\
&= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) && \text{linearity} \\
&= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(\mathbb{I}_E) \\
&= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{P}(E) && \mathbb{I}_E \text{ a simple function} \\
&= \binom{n}{3} \mathbb{P}(E).
\end{aligned}$$

Note that the third equality holds because the X_i 's are identically distributed, and so random variables X_1, X_2, X_3 are chosen for convenience in our definition of E . The above derivation tells us that in order to bound $\mathbb{E}T(G(n, p, d))$, we must simply bound $\mathbb{P}(E)$. This task is the subject of the following lemma:

Lemma 1. *There exists a universal constant $C > 0$ such that whenever $p < \frac{1}{4}$ we have that*

$$\mathbb{P}(E) \geq C^{-1} p^3 \frac{(\log \frac{1}{p})^{3/2}}{\sqrt{d}}. \quad (5)$$

Moreover, for every fixed $0 < p < 1$, there exists a constant $C_p > 0$ such that for all $d \geq C_p^{-1}$

$$\mathbb{P}(E) \geq p^3 \left(1 + \frac{C_p}{\sqrt{d}} \right). \quad (6)$$

As a preliminary result, we establish the one-dimensional marginal distribution of a random vector with the uniform distribution on \mathbb{S}^{d-1} . Particularly, section 2 of [9] calculates that this one-dimensional marginal has density

$$f_d(x) = \frac{\Gamma(d/2)}{\Gamma((d-1)/2)\sqrt{\pi}} (1-x^2)^{(d-3)/2}, \quad x \in [-1, 1]$$

with respect to Lebesgue measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$. For convenience, we let

$$\int f(x) dx$$

denote the Lebesgue integral of $f \in \mathcal{L}^1(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mu)$, where μ is the one-dimensional Lebesgue measure.

Proof of Lemma 1. The first case that Bubeck and colleagues treat is that of $d \leq \frac{1}{4} \log(1/p)$.

The authors begin by letting $\angle(\cdot, \cdot)$ denote the geodesic distance between two vectors on \mathbb{S}^{d-1} . We note that the geodesic distance on \mathbb{S}^{d-1} is the same as the great circle distance; that is, the shortest distance between any two vectors on \mathbb{S}^{d-1} is along a great circle. From [3], we know that the great circle distance between $x, y \in \mathbb{S}^{d-1}$ is $\arccos(\langle x, y \rangle)$.

We then define $g(\theta) := \mathbb{P}(\angle(X_1, X_2) < \theta)$. Just as in our previous argument for E , we have $\{\angle(X_1, X_2) < \theta\} \in \mathcal{F}$ since $h(z) = \arccos(z)$ is continuous on $z \in [-1, 1]$, and so $\arccos(\langle X_i, X_j \rangle)$ is measurable as it is the composition of a continuous function and a measurable function.

Since the coordinate system in \mathbb{R}^d is unspecified and the spherical measure Q is invariant with respect to rotation, then, without loss of generality, we may assume $X_1 = e_1$, where e_1 denotes the first standard basis vector in \mathbb{R}^d . Moreover, let φ denote the angle between $X_1 = e_1$ and X_2 in \mathbb{S}^{d-1} . Then we may rewrite $g(\theta)$ in terms of this angle φ :

$$\begin{aligned} g(\theta) &= \mathbb{P}(\angle(X_1, X_2) < \theta) \\ &= \mathbb{P}(\arccos(\langle X_1, X_2 \rangle) < \theta) \\ &= \mathbb{P}(\arccos(\cos(\varphi)) < \theta) \\ &= \mathbb{P}(\varphi < \theta). \end{aligned} \quad \text{for } \varphi \in [0, \pi/2]$$

The event that the angle φ between $X_1 = e_1$ and X_2 is less than θ defines a *hyperspherical cap* of \mathbb{S}^{d-1} . Notice that we assume (without loss of generality) $\varphi \in [0, \pi/2]$ because otherwise the complementary spherical cap $\{\varphi < \theta \leq \pi\}$ has angle less than $\pi/2$, in which case we would consider this cap instead. Thus, $g(\theta)$ is no more than the spherical measure of this hyperspherical cap. Equivalently, $g(\theta)$ is the surface area of the hyperspherical cap normalized by the total surface area of \mathbb{S}^{d-1} . And so from [8], which computes the area of a hyperspherical cap of \mathbb{S}^{d-1} , we deduce that

$$g(\theta) = \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\theta \sin(x)^{d-2} dx.$$

With a straightforward change of variables $u = 2x$, we compute

$$\begin{aligned} g(\theta/2) &= \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^{\theta/2} \sin(x)^{d-2} dx \\ &= \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\theta \frac{1}{2} \sin(u/2)^{d-2} du. \end{aligned}$$

And since $\sin(u/2) \geq \sin(u)$ for every $u \in [0, \pi]$, then we get that for all $\theta \in [0, \pi]$:

$$\begin{aligned} g(\theta/2) &\geq \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\theta \frac{1}{2} \left(\frac{\sin(u)}{2}\right)^{d-2} du && \text{monotonicity} \\ &= \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\theta 2^{1-d} \sin(u)^{d-2} du \\ &\geq \frac{(d-1)\pi^{(d-1)/2}}{\Gamma\left(\frac{d+1}{2}\right)} \int_0^\theta 2^{-d} \sin(u)^{d-2} du \\ &= 2^{-d} g(\theta). \end{aligned} \tag{7}$$

Having given this inequality for $g(\theta/2)$, let us momentarily consider the event E . First, we note that since $h(z) = \arccos(z)$ is monotonically decreasing on $[-1, 1]$, then

$$\begin{aligned} E &= \{\arccos(\langle X_1, X_2 \rangle) \leq \arccos(t_{p,d}), \arccos(\langle X_1, X_3 \rangle) \leq \arccos(t_{p,d}), \arccos(\langle X_2, X_3 \rangle) \leq \arccos(t_{p,d})\} \\ &= \{\angle(X_1, X_2) \leq \arccos(t_{p,d}), \angle(X_1, X_3) \leq \arccos(t_{p,d}), \angle(X_2, X_3) \leq \arccos(t_{p,d})\}. \end{aligned}$$

And because the geodesic distance on \mathbb{S}^{d-1} defines a proper metric, then we may invoke the triangle inequality to say

$$\begin{aligned} \angle(X_1, X_2) &\leq \frac{1}{2} \arccos(t_{p,d}), \angle(X_1, X_3) \leq \frac{1}{2} \arccos(t_{p,d}) \\ \implies \angle(X_2, X_3) &\leq \angle(X_2, X_1) + \angle(X_1, X_3) \leq \frac{1}{2} \arccos(t_{p,d}) + \frac{1}{2} \arccos(t_{p,d}) = \arccos(t_{p,d}). \end{aligned}$$

Putting together the two previous statements, we deduce

$$E \supseteq \{\angle(X_1, X_2) \leq \frac{1}{2} \arccos(t_{p,d}), \angle(X_1, X_3) \leq \frac{1}{2} \arccos(t_{p,d})\}.$$

Since X_2 and X_3 are independent random variables, then any functions of these random variables $f(X_2)$ and $g(X_3)$ are independent [7]. Particularly, we have $\angle(X_1, X_2) \perp \angle(X_1, X_3)$. Wrapping up this portion of the proof, we have

$$\begin{aligned} \mathbb{P}(E) &\geq \mathbb{P}\left\{\angle(X_1, X_2) < \frac{1}{2} \arccos(t_{p,d}), \angle(X_1, X_3) < \frac{1}{2} \arccos(t_{p,d})\right\} && \text{monotonicity} \\ &= \mathbb{P}\left\{\angle(X_1, X_2) < \frac{1}{2} \arccos(t_{p,d})\right\}^2 && \angle(X_1, X_2) \perp \angle(X_1, X_3) \\ &= g\left(\frac{1}{2} \arccos(t_{p,d})\right)^2 \\ &\geq 2^{-2d} g(\arccos(t_{p,d})) && \text{from (7)} \\ &= p^2 2^{-2d}. && \mathbb{P}(\langle X_i, X_j \rangle \geq t_{p,d}) = p \end{aligned}$$

Because we assumed that $d \leq \frac{1}{4} \log(1/p)$, then

$$p^2 2^{-2d} \geq p^2 2^{\frac{1}{2} \log(1/p)} \geq p^3 \left(1 + c (\log(1/p))^{3/2}\right)$$

for some constant $c > 0$. And since $\mathbb{P}(E) \geq p^3 \left(1 + c (\log(1/p))^{3/2}\right) \geq c p^3 (\log(1/p))^{3/2}$, taking $C = c$ satisfies (5), the first inequality in Lemma 1. Similarly, $\mathbb{P}(E) \geq p^3 \left(1 + c (\log(1/p))^{3/2}\right)$ means that $C_p = c \log(1/p)^{3/2}$ satisfies (6). And so we have proven that Lemma 1 holds for the case of $d \leq \frac{1}{4} \log(1/p)$.

Now for the remainder of the proof, we assume $d \geq \frac{1}{4} \log(1/p)$. We begin with the case of $p < 1/2$.

Define the events $E_{i,j} = \{\langle X_i, X_j \rangle \geq t_{p,d}\} \in \mathcal{F}$ as well as $E_{i,j}(x) = \{\langle X_i, X_j \rangle = x\} \in \mathcal{F}$.

We consider the mapping $T : (\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}) \otimes \mathcal{B}(\mathbb{S}^{d-1}), Q \times Q) \rightarrow ([-1, 1], \mathcal{B}([-1, 1]), V)$ defined $T(x, y) := \langle x, y \rangle$, where V is the image measure of $Q \times Q$ under T . We claim that there exists a conditional probability of $Q \times Q$ under T , which is unique $Q \times Q$ -almost surely. To understand why this is the case, recall from [7] that the following are sufficient conditions for the existence of a conditional probability:

1. $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ is a complete, separable metric space (a *Polish space*)
2. The sigma-field from which we are mapping is $\mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1})$
3. V is the image measure of $Q \times Q$ under map T
4. $g_r(T)$ is $\mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1])$ -measurable

Notice that the third condition is immediately true as a consequence of how we defined our mapping T . Further, $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ endowed with the Euclidean metric $\|\cdot\|_2$ is clearly a metric space. $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ is separable because it is a subset of separable metric space $(\mathbb{R}^d \times \mathbb{R}^d, \|\cdot\|_2)$, and it is complete because it is a closed subset of complete metric space $(\mathbb{R}^d \times \mathbb{R}^d, \|\cdot\|_2)$. Thus, $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$ is indeed a Polish space.

For the second condition, we note that the Borel sigma-field on \mathbb{R} satisfies:

$$\underbrace{\mathcal{B}(\mathbb{R}) \otimes \dots \otimes \mathcal{B}(\mathbb{R})}_{k\text{-times}} = \mathcal{B}(\mathbb{R}^k).$$

Using this property, we have $\mathcal{B}(\mathbb{S}^{d-1}) \otimes \mathcal{B}(\mathbb{S}^{d-1}) = \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1})$, and so T indeed maps from the Borel sigma-field on $\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}$.

To complete our argument, we must show that the graph of T , $g_r(T)(x, y) := \{((x, y), T(x, y)) : (x, y) \in \mathbb{S}^{d-1} \times \mathbb{S}^{d-1}\}$ is $\mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1])$ -measurable. As we have previously discussed, since $h(x, y) = \langle x, y \rangle$ is continuous for $x, y \in \mathbb{S}^{d-1}$, then $T : (\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}), Q \times Q) \rightarrow ([-1, 1], \mathcal{B}([-1, 1]))$ is measurable. Now, we define the map $\tilde{T} : \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1]) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ by $\tilde{T}((x, y), t) = T(x, y) - t$. Notice that we can write $\tilde{T} := \tilde{T}_1 - \tilde{T}_2$, where $\tilde{T}_1((x, y), t) = T(x, y)$ and $\tilde{T}_2((x, y), t) = t$. Clearly, each of \tilde{T}_1, \tilde{T}_2 is measurable by a generating class argument: $\tilde{T}_1^{-1}((-\infty, a]) = \{((x, y), t) : T(x, y) \leq a\} = \{x, y \in \mathbb{S}^{d-1} : T(x, y) \leq a\} \times [-1, 1] \in \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1])$ as well as $\tilde{T}_2^{-1}((-\infty, a]) = \{((x, y), t) : t \leq a\} = (\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \times \{t \leq a\} \in \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1])$. Note that the measurability of $\tilde{T}_1^{-1}((-\infty, a])$ requires the measurability of T . This, in turn, implies that $\tilde{T} = \tilde{T}_1 - \tilde{T}_2$ is $\mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1])$ -measurable. Therefore, we deduce that

$$\begin{aligned} \tilde{T}^{-1}(\{0\}) &= \{((x, y), t) : T(x, y) - t = 0\} \\ &= \{((x, y), t) : T(x, y) = t\} \in \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1]) \\ &\Rightarrow g_r(T)(x, y) \in \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}) \otimes \mathcal{B}([-1, 1]). \end{aligned}$$

Having established each of the sufficient conditions, we deduce that a conditional probability $\{(Q \times Q)_t\}$ mapping from $([-1, 1], \mathcal{B}([-1, 1]), V)$ to $(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}) \otimes \mathcal{B}(\mathbb{S}^{d-1}), Q \times Q)$ indeed exists. While our proof of existence may appear overly pendantic, the conditional probability of $Q \times Q$ under T will play the single most important role throughout the remainder of our paper.

Note that, as part of their argument, Bubeck and colleagues provide a specific definition for the conditional probability. In particular, they claim that for $F \in \mathcal{F}$, the conditional probability of $Q \times Q$ given T is

$$\mathbb{P}(F|T(X_i, X_j) = x) = \lim_{\varepsilon \rightarrow 0} \frac{\mathbb{P}(F \cap \{\langle X_i, X_j \rangle \in [x - \varepsilon, x + \varepsilon]\})}{\mathbb{P}(\langle X_i, X_j \rangle \in [x - \varepsilon, x + \varepsilon])} \quad x \in [-1, 1].$$

Here, we have measure \mathbb{P} rather than $Q \times Q$ because, although the conditional probability was defined on the probability measure $Q \times Q$ under map T , Q is defined to be the image measure of \mathbb{P} under each random variable X_i . While one could establish that this is, in fact, the conditional probability distribution of $Q \times Q$ given T , the remainder of the argument does not rely on this particular definition. Accordingly, we prefer our approach of establishing the existence of the conditional probability without providing superfluous details.

As a consequence of the existence of the conditional probability of $Q \times Q$ under T , we consider

$$\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3}).$$

We rewrite the second term in the above expression as

$$\begin{aligned} \mathbb{P}(E_{1,3}, E_{2,3}) &= \mathbb{P}^x(\mathbb{P}(E_{1,3}, E_{2,3}|\langle X_1, X_2 \rangle = x)) \\ &= \frac{1}{2}\mathbb{P}(E_{1,3}, E_{2,3}|\langle X_1, X_2 \rangle \geq 0) + \frac{1}{2}\mathbb{P}(E_{1,3}, E_{2,3}|\langle X_1, X_2 \rangle < 0). \end{aligned}$$

Note that we are simply appealing to the definition of the conditional probability. In an elementary probability course, this is referred to as the law of total expectation.

We now make what seems like, at first glance, an abstruse claim. Particularly, we suggest that $f(x) := \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(x))$ is a monotonically-increasing function in $x \in [-1, 1]$. To understand why this is the case, we consider $\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(x))$ for some fixed x . Once again, since the coordinate system in \mathbb{R}^d is unspecified and Q is rotation invariant, then, without loss of generality, we may let $X_1 = e_1$ and $X_2 = xe_1 + \sqrt{1-x^2}e_2$ so that $\langle X_1, X_2 \rangle = x$. Now, conditioning on $E_{1,2}(x)$, each of $\{\langle X_1, X_3 \rangle \geq t_{p,d}\} = \{X_{3,1} \geq t_{p,d}\}$ and $\{\langle X_2, X_3 \rangle \geq t_{p,d}\} = \{xX_{3,1} + \sqrt{1-x^2}X_{3,2} \geq t_{p,d}\}$ defines a hyperspherical cap of \mathbb{S}^{d-1} . In particular, $\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(x))$ is equal to the spherical measure of the intersection of these two hyperspherical caps.

Notice that $\langle X_1, X_2 \rangle = x \Rightarrow \cos(\varphi) = x$, where φ is the angle formed between X_1 and X_2 . Therefore, as x increases from -1 to 1 , the angle φ between X_1 and X_2 decreases from π to 0 , and so the region of \mathbb{S}^{d-1} on which the hyperspherical caps intersect has greater measure. In particular, when $x = 1$, then $\cos(\varphi) = 1 \Rightarrow \varphi = 0 \Rightarrow X_1 = X_2$, in which case we have $\{\langle X_1, X_3 \rangle \geq t_{p,d}\} = \{X_{3,1} \geq t_{p,d}\} = \{\langle X_2, X_3 \rangle \geq t_{p,d}\}$, meaning that the two hyperspherical caps coincide exactly.⁴

Having established the increasing property of $f(x) = \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(x))$ on $x \in [-1, 1]$, we are now ready to bound $\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3})$:

$$\begin{aligned} & \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3}) \\ &= \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | \langle X_1, X_2 \rangle \geq 0) - \frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | \langle X_1, X_2 \rangle < 0) \\ &= \left(\frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | \langle X_1, X_2 \rangle \geq 0) \right) + \left(\frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2} < 0) \right). \end{aligned}$$

Notice that since we assumed $p < 1/2$, then it must be the case that $t_{p,d} > 0$ by definition of $\mathbb{P}(\langle X_i, X_j \rangle \geq t_{p,d}) = p$. Accordingly, we see that the term $\frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \frac{1}{2} \mathbb{P}(E_{1,3}, E_{2,3} | \langle X_1, X_2 \rangle \geq 0) \geq 0$ by the increasing property of f . This allows us to simplify the above expression as

$$\begin{aligned} & \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3}) \\ & \geq \frac{1}{2} \left(\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2} < 0) \right) \\ & \geq \frac{1}{2} \left(\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2} < 0) \right) \quad \text{increasing property of } f \\ & \geq \frac{1}{2} \left(\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(0)) \right) \quad \text{increasing property of } f \end{aligned} \quad (8)$$

Consider the random variable $Z_1 = \langle X_1, X_3 \rangle$. By a similar trick to as above, we make use of the rotational invariance of Q to define the coordinate axes of \mathbb{R}^d in terms of X_1, X_3 . In particular, without loss of generality, we may take $X_1 = e_1, X_3 = z_1 e_1 + \sqrt{1 - z_1^2} e_2$ so that $Z_1 = \langle X_1, X_3 \rangle = z_1$, the first coordinate of the vector X_3 . Thus, from the preliminaries to Lemma 1, we know that Z_1 has density f_d with respect to Lebesgue measure. And so we invoke Bayes' rule to rewrite the integral in (8) as an integral over the one-dimensional marginal distribution Z_1 :

$$\begin{aligned} & \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(0)) \\ &= \int_{z_1 \geq t_{p,d}} \{ \mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(0)) \} f_d(z_1) dz_1. \end{aligned} \quad (9)$$

That is, we disintegrate the joint measure $\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(x))$ as the integral of the conditional probability $\mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(x))$ times the marginal density $f_d(z_1)$ over the region $z_1 \geq t_{p,d}$ [7].

Now, conditioning on $Z_1 = z_1$, we claim that $Z_2 := \left\langle X_3, \frac{\text{Proj}_{X_1^\perp} X_2}{|\text{Proj}_{X_1^\perp} X_2|} \right\rangle$ is distributed as $\sqrt{1 - z_1^2} Z'$, where Z' has density f_{d-1} with respect to Lebesgue measure. Here, $\text{Proj}_{X_1^\perp} X_2$ denotes the orthogonal projection of X_2 onto the subspace X_1^\perp . To see why this statement is true, recall from above that we took $X_1 = e_1$ and $X_3 = z_1 e_1 + \sqrt{1 - z_1^2} e_2$. And so we have that

$$\begin{aligned} Z_2 &= \left\langle X_3, \frac{\text{Proj}_{X_1^\perp} X_2}{|\text{Proj}_{X_1^\perp} X_2|} \right\rangle = \left\langle X_3, \frac{X_2 - \langle X_1, X_2 \rangle X_1}{|\text{Proj}_{X_1^\perp} X_2|} \right\rangle \\ &= \frac{1}{|\text{Proj}_{X_1^\perp} X_2|} (\langle X_2, X_3 \rangle - \langle X_1, X_2 \rangle \langle X_1, X_3 \rangle) \end{aligned}$$

⁴For more about the measure of the intersection of hyperspherical caps, see [5].

$$\begin{aligned}
&= \frac{1}{|\text{Proj}_{X_1^\perp} X_2|} \left(\left(z_1 X_{2,1} + \sqrt{1 - z_1^2} X_{2,2} \right) - z_1 X_{2,1} \right) \\
&= \frac{1}{|\text{Proj}_{X_1^\perp} X_2|} \sqrt{1 - z_1^2} X_{2,2} \\
&= \frac{1}{\sqrt{1 - \langle X_1, X_2 \rangle^2}} \sqrt{1 - z_1^2} X_{2,2} \\
&= \frac{1}{\sqrt{1 - X_{2,1}^2}} \sqrt{1 - z_1^2} X_{2,2}.
\end{aligned}$$

Note that $\frac{X_{2,2}}{\sqrt{1 - X_{2,1}^2}} \sim f_{d-1}$, and so Z_2 indeed has the same distribution as $\sqrt{1 - z_1^2} Z'$.

Using this result, we rewrite $E_{2,3}$ conditioned on $Z_1 = z_1$ as $\left\{ \omega : \left| \text{Proj}_{X_1^\perp} X_2 \right| Z_2 + \langle X_3, \text{Proj}_{X_1} X_2 \rangle \geq t_{p,d} \right\}$:

$$\begin{aligned}
&\mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(t_{p,d})) \\
&= \mathbb{P} \left(\left(\sqrt{1 - \langle X_1, X_2 \rangle^2} \right) \sqrt{1 - z_1^2} Z' + \langle X_1, X_2 \rangle z_1 \geq t_{p,d} | E_{1,2}(t_{p,d}) \right).
\end{aligned}$$

And since $E_{1,2}(t_{p,d}) = \{ \langle X_1, X_2 \rangle \geq t_{p,d} \}$, then we have

$$\begin{aligned}
&\mathbb{P} \left(\left(\sqrt{1 - \langle X_1, X_2 \rangle^2} \right) \sqrt{1 - z_1^2} Z' + \langle X_1, X_2 \rangle z_1 \geq t_{p,d} | E_{1,2}(t_{p,d}) \right) \\
&= \mathbb{P} \left(\sqrt{1 - t_{p,d}^2} \sqrt{1 - z_1^2} Z' + t_{p,d} z_1 \geq t_{p,d} \right) \\
&= \mathbb{P} \left(\sqrt{1 - t_{p,d}^2} Z' \geq \sqrt{\frac{1 - z_1}{1 + z_1}} t_{p,d} \right). \tag{10}
\end{aligned}$$

Notice that $f(z_1) = \sqrt{\frac{1 - z_1}{1 + z_1}}$ is a decreasing function in $z_1 \in (-1, 1]$, and so we get that $\mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(t_{p,d}))$, and thus the right-hand side of (8), is an increasing function in z_1 . Moreover, conditioning on $E_{1,2}(0)$, we see that $\langle X_2, X_3 \rangle = \langle X_3, \text{Proj}_{X_1^\perp} X_2 \rangle = Z_2$ since $\text{Proj}_{X_1} X_2 = \langle X_1, X_2 \rangle X_1 = 0$. That is, conditioning on $Z_1 = z_1$ and $E_{1,2}(0)$, $E_{2,3}$ is distributed according to $\sqrt{1 - z_1^2} Z'$. Recall from above that Z' has density f_{d-1} with respect to Lebesgue measure. This means that $\mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(0)) = \mathbb{P}(\sqrt{1 - z_1^2} Z' \geq t_{p,d})$ is the right tail probability of $\sqrt{1 - z_1^2} Z'$ and so is a decreasing function in z_1 .

Putting everything together, we may bound the integral (9) below by each of the two conditional probabilities at $Z_1 = t_{p,d}$. That is, we have

$$\begin{aligned}
&\mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3} | E_{1,2}(0)) \\
&= \int_{z_1 \geq t_{p,d}} \{ \mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3} | Z_1 = z_1, E_{1,2}(0)) \} f_d(z_1) dz_1 \\
&\geq \int_{z_1 \geq t_{p,d}} \{ \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(0)) \} f_d(z_1) dz_1 \quad \text{monotonicity} \\
&= \{ \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(0)) \} \left(\int_{z_1 \geq t_{p,d}} f_d(z_1) dz_1 \right).
\end{aligned}$$

Since $\mathbb{P}(Z_1 \geq t_{p,d}) = \mathbb{P}(\langle X_1, X_3 \rangle \geq t_{p,d}) = p$, then

$$\{ \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3} | Z_1 = t_{p,d}, E_{1,2}(0)) \} \left(\int_{z_1 \geq t_{p,d}} f_d(z_1) dz_1 \right)$$

$$\begin{aligned}
&= p \{ \mathbb{P}(E_{2,3}|Z_1 = t_{p,d}, E_{1,2}(t_{p,d})) - \mathbb{P}(E_{2,3}|Z_1 = t_{p,d}, E_{1,2}(0)) \} \\
&= p \left\{ \mathbb{P} \left(\sqrt{1-t_{p,d}^2} Z' \geq \sqrt{\frac{1-t_{p,d}}{1+t_{p,d}}} t_{p,d} \right) - \mathbb{P} \left(\sqrt{1-t_{p,d}^2} Z' \geq t_{p,d} \right) \right\} \quad \text{from (10)} \\
&= p \left\{ \mathbb{P} \left(Z' \geq \frac{t_{p,d}}{1+t_{p,d}} \right) - \mathbb{P} \left(Z' \geq \frac{t_{p,d}}{\sqrt{1-t_{p,d}^2}} \right) \right\} \\
&= p \mathbb{P} \left(\frac{t_{p,d}}{1+t_{p,d}} \leq Z' \leq \frac{t_{p,d}}{\sqrt{1-t_{p,d}^2}} \right).
\end{aligned}$$

And since Z' has density f_{d-1} with respect to Lebesgue measure, then

$$\begin{aligned}
&\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(0)) \\
&\geq p \int_{\frac{t_{p,d}}{1+t_{p,d}}}^{\frac{t_{p,d}}{\sqrt{1-t_{p,d}^2}}} f_{d-1}(x) dx \\
&\geq p \int_{\frac{t_{p,d}}{1+t_{p,d}}}^{t_{p,d}} f_{d-1}(x) dx. \quad \text{monotonicity}
\end{aligned}$$

Now, by a previous technical lemma proven in [1], we have

$$f_{d-1}(z) \geq c' d p t_{p,d} \quad \text{for all } 0 \leq z \leq t_{p,d},$$

which serves as a bound on f_{d-1} for some absolute constant $c' > 0$.⁵ And so again by monotonicity of the [Lebesgue] integral,

$$\begin{aligned}
&\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(0)) \\
&\geq p \int_{\frac{t_{p,d}}{1+t_{p,d}}}^{t_{p,d}} c' d p t_{p,d} dx \\
&= \left(c' \left(t_{p,d} - \frac{t_{p,d}}{1+t_{p,d}} \right) \right) d p^2 t_{p,d} \\
&= \left(\frac{c'}{1+t_{p,d}} \right) d p^2 t_{p,d}^3.
\end{aligned}$$

From the same technical lemma in [1], we take $c = \frac{c'}{1+t_{p,d}} > 0$ to be an absolute constant so that

$$c d p^2 t_{p,d}^3 \geq c' \left(\frac{1}{2} - p^3 \right) p^2 \left(\frac{(\log(1/p))^{3/2}}{\sqrt{d}} \wedge d \right).$$

Moreover, because we assumed that $d \geq \frac{1}{4} \log(1/p) \Rightarrow d^{3/2} \geq (\frac{1}{4})^{3/2} \log(1/p)^{3/2} \Rightarrow d \geq \frac{(\frac{1}{4})^{3/2} \log(1/p)^{3/2}}{\sqrt{d}}$, then we achieve the bound

$$\begin{aligned}
\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(0)) &\geq c' \left(\frac{1}{2} - p^3 \right) p^2 \left(\frac{(\log(1/p))^{3/2}}{\sqrt{d}} \wedge d \right) \\
&\geq \frac{1}{16} c' \left(\frac{1}{2} - p^3 \right) p^2 \frac{\log(1/p)^{3/2}}{\sqrt{d}}. \quad (11)
\end{aligned}$$

To wrap up this case of the proof, we return to $\mathbb{P}(E) = \mathbb{P}(E_{1,2}, E_{1,3}, E_{2,3})$. Since $p \in (0, 1)$, then the event $E_{1,2}$ has strictly positive measure with respect to \mathbb{P} . As a result, we have that $\mathbb{P}(F|E_{1,2})$, $F \in \mathcal{F}$ respects

⁵For a formal derivation, one should reference Lemma 2 on pp. 509-510 in [1].

the traditional notion of conditional probability from an elementary statistics course: $\mathbb{P}(F|E_{1,2}) = \frac{\mathbb{P}(F \cap E_{1,2})}{\mathbb{P}(E_{1,2})}$. From a measure-theoretic perspective, one could show that the proposed $\mathbb{P}(F|E_{1,2})$ satisfies the definition of a conditional probability and thereafter cite the fact that a conditional probability is unique $Q \times Q$ almost-surely [7].⁶ Specifically, this allows us to express

$$\begin{aligned} \mathbb{P}(E) &= \mathbb{P}(E_{1,2}, E_{1,3}, E_{2,3}) \\ &= \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2})\mathbb{P}(E_{1,2}) \\ &= \{\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}) - \mathbb{P}(E_{1,3}, E_{2,3})\}\mathbb{P}(E_{1,2}) + \mathbb{P}(E_{1,3}, E_{2,3})\mathbb{P}(E_{1,2}) \\ &\geq \frac{1}{2}\{\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(0))\}\mathbb{P}(E_{1,2}) + \mathbb{P}(E_{1,3}, E_{2,3})\mathbb{P}(E_{1,2}). \quad \text{from (8)} \end{aligned}$$

With our previous lower bound (11) on $\mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}, E_{2,3}|E_{1,2}(0))$,

$$\begin{aligned} \mathbb{P}(E) &\geq \left\{ \frac{1}{32}c' \left(\frac{1}{2} - p^3 \right) p^2 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \right\} \mathbb{P}(E_{1,2}) + \mathbb{P}(E_{1,3}, E_{2,3})\mathbb{P}(E_{1,2}) \\ &= \left\{ \frac{1}{32}c' \left(\frac{1}{2} - p^3 \right) p^2 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \right\} p + \mathbb{P}(E_{1,3}, E_{2,3})p. \end{aligned}$$

Since $\langle X_1, X_3 \rangle, \langle X_2, X_3 \rangle$ are functions of independent random variables X_1, X_2 , then they are also independent. Finally, we conclude

$$\begin{aligned} \mathbb{P}(E) &\geq \left\{ \frac{1}{32}c' \left(\frac{1}{2} - p^3 \right) p^2 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \right\} p + \mathbb{P}(E_{1,3})\mathbb{P}(E_{2,3})p \\ &= \left\{ \frac{1}{32}c' \left(\frac{1}{2} - p^3 \right) p^2 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \right\} p + p^3 \\ &= p^3 \left(1 + \frac{1}{32}c' \left(\frac{1}{2} - p \right)^3 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \right). \quad (12) \end{aligned}$$

This, in turn, implies that $\mathbb{P}(E) \geq p^3 \frac{1}{32}c' \left(\frac{1}{2} - p \right)^3 \frac{\log(1/p)^{3/2}}{\sqrt{d}} \geq \left(\frac{c'}{256} \right) p^3 \frac{\log(1/p)^{3/2}}{\sqrt{d}}$ for $p < \frac{1}{4}$. And so taking $C = \frac{256}{c'} > 0$, we have that (5) holds. This concludes the proof of (5), as we have shown that the desired inequality holds for all cases in which $p < \frac{1}{4}$.

From (12), we also see that $C_p = \frac{1}{32}c' \left(\frac{1}{2} - p \right)^3 \log(1/p)^{3/2}$ is a positive constant for all $p < \frac{1}{2}$ such that $\mathbb{P}(E) \geq p^3 \left(1 + \frac{C_p}{\sqrt{d}} \right)$ whenever $d \geq C_p^{-1}$. Thus, this value of C_p satisfies inequality (6) for every $p < \frac{1}{2}$.

It remains to verify inequality (6) for $p \geq \frac{1}{2}$. The particular instance of $p = \frac{1}{2}$ is proven in Lemma 5 of [1], and so we focus our attention on $p > \frac{1}{2}$. This case follows from an analogous argument to that for $p < \frac{1}{2}$. Thus, for the sake of brevity, we omit these repeated details.

To begin, let us consider $\tilde{E} := E_{1,2}^c \cap E_{1,3}^c \cap E_{2,3}^c \in \mathcal{F}$, the event that there exists no edge between each of vertices 1, 2, and 3 in $G \sim G(n, p, d)$. Once again, we know that $\tilde{E} \in \mathcal{F}$ since $f(\omega) = \langle X_i(\omega), X_j(\omega) \rangle$ is $\mathcal{F} \setminus \mathcal{B}([-1, 1])$ -measurable. Just as before, $X_2 \perp X_3$ implies $\langle X_1, X_2 \rangle \perp \langle X_1, X_3 \rangle$, and so we have $\mathbb{P}(E_{1,2} \cap E_{1,3}) = \mathbb{P}(E_{1,2})^2 = p^2$. Since $E \cup (E_{1,2}^c \cup E_{1,3}^c \cup E_{2,3}^c) = \Omega$, then we may write

$$\begin{aligned} \mathbb{P}(E) &= 1 - \mathbb{P}(E_{1,2}^c \cup E_{1,3}^c \cup E_{2,3}^c) \\ &= 1 - \left(\binom{3}{1} \mathbb{P}(E_{1,2}^c) - \binom{3}{2} \mathbb{P}(E_{1,2}^c, E_{1,3}^c) + \mathbb{P}(E_{1,2}^c, E_{1,3}^c, E_{2,3}^c) \right) \\ &= 1 - (3(1-p) - 3(1-p)^2 + \mathbb{P}(E_{1,2}^c, E_{1,3}^c, E_{2,3}^c)) \quad \langle X_1, X_2 \rangle \perp \langle X_1, X_3 \rangle \\ &= 1 - 3p + 3p^2 - \mathbb{P}(\tilde{E}). \quad (13) \end{aligned}$$

⁶That is, there is no ambiguity in the definition of the conditional probability $\mathbb{P}(\cdot|F)$ for sets $F \in \mathcal{F}$ of positive measure.

Notice that we used the principle of inclusion-exclusion to decompose $\mathbb{P}(\tilde{E}) = \mathbb{P}(\mathbb{I}_{\tilde{E}})$ as

$$\mathbb{I}_{\tilde{E}} = \mathbb{I}_{E_{1,2}^c} + \mathbb{I}_{E_{1,3}^c} + \mathbb{I}_{E_{2,3}^c} - \mathbb{I}_{E_{1,2}^c \cap E_{1,3}^c} - \mathbb{I}_{E_{1,2}^c \cap E_{2,3}^c} - \mathbb{I}_{E_{1,3}^c \cap E_{2,3}^c} + \mathbb{I}_{E_{1,2}^c \cap E_{1,3}^c \cap E_{2,3}^c}.$$

From (13), we see that in order to bound $\mathbb{P}(E)$ below, it suffices to find an appropriate upper bound for $\mathbb{P}(\tilde{E})$. Just as in the previous case, because $\mathbb{P}(E_{1,2}^c) > 0$ we may express $\mathbb{P}(\tilde{E})$ using the elementary notion of conditional probability (of $Q \times Q$ under T)

$$\begin{aligned} \mathbb{P}(\tilde{E}) &= \mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}^c) \mathbb{P}(E_{1,2}^c) \\ &= \{\mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}^c) - \mathbb{P}(E_{1,3}^c, E_{2,3}^c)\} \mathbb{P}(E_{1,2}^c) + \mathbb{P}(E_{1,3}^c, E_{2,3}^c) \mathbb{P}(E_{1,2}^c) \\ &= \{\mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}^c) - \mathbb{P}(E_{1,3}^c, E_{2,3}^c)\} (1-p) + (1-p)^3. \end{aligned} \quad (14)$$

Accordingly, it only remains to find an appropriate upper bound on $\mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}^c) - \mathbb{P}(E_{1,3}^c, E_{2,3}^c)$, which would then give us the desired upper bound on $\mathbb{P}(\tilde{E})$.

By a nearly identical argument to the case of $p < \frac{1}{2}$, one may derive

$$\mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}^c) - \mathbb{P}(E_{1,3}^c, E_{2,3}^c) \leq \frac{1}{2} \{\mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}(t_{p,d})) - \mathbb{P}(E_{1,3}^c, E_{2,3}^c | E_{1,2}(0))\} \leq -c_p / \sqrt{d}$$

for c_p a positive constant in $p > \frac{1}{2}$.⁷ Plugging this inequality into (14) yields $\mathbb{P}(\tilde{E}) \leq -(1-p)c_p / \sqrt{d} + (1-p)^3$. Thus, $\mathbb{P}(E) \geq 1 - 3p + 3p^2 + (1-p)c_p / \sqrt{d} - (1-p)^3 = p^3 + (1-p)c_p / \sqrt{d}$.

We rewrite the above inequality as $\mathbb{P}(E) \geq p^3 \left(1 + \frac{(1-p)c_p}{p^3 \sqrt{d}}\right) = p^3 \left(1 + \frac{C_p}{\sqrt{d}}\right)$, where $C_p := \frac{(1-p)c_p}{p^3}$ is a positive constant for every $p > \frac{1}{2}$. And so we conclude that by choosing C_p as above, then inequality (6) holds for all $d \geq C_p^{-1}$.

Having considered all possible cases, $d \leq \frac{1}{4} \log(1/p)$ and $d > \frac{1}{4} \log(1/p)$ for each of $p < \frac{1}{2}$, $p = \frac{1}{2}$, and $p > \frac{1}{2}$, then we have shown that $\mathbb{P}(E)$ satisfies the two specified lower bounds (5) and (6). Accordingly, for each fixed $p \in (0, 1)$, we have given a lower bound on the expected value of the triangle statistic $T(G)$ under $G \sim G(n, p, d)$, the alternative hypothesis H_1 . \square

3 Total Variation Distance Between $\tau(G(n, p))$ and $\tau(G(n, p, d))$

Having examined the triangle statistic T , we now turn our attention to the signed triangle statistic τ . As discussed in section 1.3, we want to say something about the power of the statistic τ for testing the null hypothesis H_0 against H_1 . In particular, we consider what it means for statistic τ to be “asymptotically powerful” as long as $d \ll n^3$. This statement is explicated in Theorem 2:

Theorem 2. *Let $p \in (0, 1)$ be fixed and assume that $d/n^3 \rightarrow 0$. Then*

$$TV(\tau(G(n, p)), \tau(G(n, p, d))) \rightarrow 1.$$

Note that $\tau(G(n, p)) : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q_0)$ and $\tau(G(n, p, d)) : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}), Q_1)$, where Q_0 and Q_1 denote the image measures of \mathbb{P} under $\tau(G(n, p))$ and $\tau(G(n, p, d))$, respectively. Therefore, the *total variation distance*, denoted TV , between the distribution of $\tau(G(n, p))$ and that of $\tau(G(n, p, d))$ is given by

$$\begin{aligned} TV(\tau(G(n, p)), \tau(G(n, p, d))) &:= \sup_{B \in \mathcal{B}(\mathbb{R})} |Q_0(B) - Q_1(B)| \\ &= \sup_{B \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(\{\omega : \tau(G(n, p)) \in B\}) - \mathbb{P}(\{\omega : \tau(G(n, p, d)) \in B\})|. \end{aligned}$$

⁷See pp. 512-513 of [1] for a detailed derivation of this bound.

Proof of Theorem 2. For convenience, let us denote $\bar{A}_{i,j} := A_{i,j} - \mathbb{E}(A_{i,j})$ as well as $\tau_G(i, j, k) = \bar{A}_{i,j} \bar{A}_{i,k} \bar{A}_{j,k}$. With this notation, the signed triangle statistic (4) of graph G on n vertices can be expressed simply as

$$\tau(G) = \sum_{\{i,j,k\} \in \binom{[n]}{3}} \tau_G(i, j, k).$$

Intuitively, Theorem 2 tells us that the distribution of the signed triangle statistic under H_0 is very different from its distribution under H_1 so long as d is much smaller than n^3 [1]. In order to prove this statement, Bubeck and colleagues show that the difference between $\mathbb{E}(\tau(G(n, p)))$ and $\mathbb{E}(\tau(G(n, p, d)))$ exceeds the individual standard deviations of these statistics.⁸ We follow this same structure in our proof, breaking down the problem into four parts.

3.1 Calculations for $\mathbb{E}(\tau(G(n, p)))$ & $\text{Var}(\tau(G(n, p)))$

Computing the expectation and variance of the signed triangle statistic is simple for G distributed according to the Erdős–Rényi model. We begin by calculating $\mathbb{E}(\tau(G(n, p)))$:

$$\begin{aligned} \mathbb{E}(\tau(G(n, p))) &= \mathbb{E} \left(\sum_{\{i,j,k\} \in \binom{[n]}{3}} \tau_{G(n,p)}(i, j, k) \right) \\ &= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(\tau_{G(n,p)}(i, j, k)) \\ &= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(\bar{A}_{i,j} \bar{A}_{i,k} \bar{A}_{j,k}) \\ &= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(\bar{A}_{i,j}) \mathbb{E}(\bar{A}_{i,k}) \mathbb{E}(\bar{A}_{j,k}) \\ &= 0. \end{aligned}$$

The fourth equality holds because for G generated according to the Erdős–Rényi standard model, each random variable $A_{i,j}$, $i, j \in [n]$ is independent, and so each $\bar{A}_{i,j} = A_{i,j} - p$ is independent. As for the final equality, we use $\mathbb{E}(\bar{A}_{i,j}) = \mathbb{E}(A_{i,j}) - p = 0$ since $(\Omega, \mathcal{F}, \mathbb{P})$ a probability space as well as $\mathbb{E}(A_{i,j}) = p$.

Now for the variance, we make a few observations about $\tau_{G(n,p)}$. Namely, for each of $i, j, k, l \in [n]$ distinct,

$$\begin{aligned} \mathbb{E}(\tau_{G(n,p)}(i, j, k) \tau_{G(n,p)}(i, j, l)) &= \mathbb{E}(\bar{A}_{i,j}^2 \bar{A}_{i,k} \bar{A}_{j,k} \bar{A}_{i,l} \bar{A}_{j,l}) \\ &= \mathbb{E}(\bar{A}_{i,j}^2 \bar{A}_{i,k} \bar{A}_{j,k} \bar{A}_{i,l}) \mathbb{E}(\bar{A}_{j,l}) && \text{independence of the } \bar{A}_{i,j} \text{'s} \\ &= 0. && \text{from above, } \mathbb{E}(\bar{A}_{i,j}) = 0 \end{aligned}$$

Additionally, note that for distinct $i, j \in [n]$

$$\begin{aligned} \mathbb{E}(\bar{A}_{i,j}^2) &= \mathbb{E}(A_{i,j} - p)^2 \\ &= \mathbb{E}(A_{i,j}^2 - 2pA_{i,j} + p^2) \\ &= \mathbb{E}(A_{i,j}^2) - 2p\mathbb{E}(A_{i,j}) + p^2 \\ &= p - 2p^2 + p^2 = p(1 - p). \end{aligned}$$

⁸As always, the “expectation” of $\tau(G)$, denoted $\mathbb{E}(\tau(G))$, refers to the integral of random variable $\tau(G) : (\Omega, \mathcal{F}, \mathbb{P}) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$. That is, $\mathbb{E}(\tau(G))$ is just an equivalent expression for $\mathbb{P}(\tau(G))$. We prefer the former notation since it provides some intuition for what this quantity represents.

The fourth equality holds because $A_{i,j}(\omega) \in \{0, 1\}$, $\forall \omega \in \Omega$, and so $A_{i,j}^2 = A_{i,j}$ (recall the definition of $A_{i,j}$ from section 1.3). Putting together the two previous results, then for $i, j, k \in [n]$ unique and $i', j', k' \in [n]$ unique,

$$\mathbb{E}(\tau_{G(n,p)}(i, j, k)\tau_{G(n,p)}(i', j', k')) = \begin{cases} 0 & \text{if } \{i, j, k\} \neq \{i', j', k'\} \\ p^3(1-p)^3 & \text{if } \{i, j, k\} = \{i', j', k'\} \end{cases} \quad (15)$$

Therefore, the variance of $\tau_{G(n,p)}$ may be expressed as

$$\begin{aligned} \text{Var}(\tau(G(n,p))) &= \mathbb{E}(\tau(G(n,p))^2) & \mathbb{E}(\tau(G(n,p))) &= 0 \\ &= \mathbb{E}\left(\sum_{\{i,j,k\} \in \binom{[n]}{3}} \tau_{G(n,p)}(i, j, k)\right)^2 \\ &= \mathbb{E}\left(\sum_{\{i,j,k\} \in \binom{[n]}{3}} \tau_{G(n,p)}(i, j, k)^2\right) \\ &\quad + \mathbb{E}\left(\sum_{\substack{\{i,j,k\} \in \binom{[n]}{3}, \{i',j',k'\} \in \binom{[n]}{3} \\ \{i,j,k\} \neq \{i',j',k'\}}} 2\tau_{G(n,p)}(i, j, k)\tau_{G(n,p)}(i', j', k')\right) & \text{multinomial theorem} \\ &= \sum_{\{i,j,k\} \in \binom{[n]}{3}} \mathbb{E}(\tau_{G(n,p)}(i, j, k)^2) & \text{linearity and (15)} \\ &= \binom{n}{3} p^3(1-p)^3. & \text{from (15)} \end{aligned}$$

And so since $p \in (0, 1)$, then $\text{Var}(\tau(G(n,p))) = \binom{n}{3} p^3(1-p)^3 \leq \binom{n}{3} \leq n^3$.

Thus we have shown

$$\mathbb{E}(\tau(G(n,p))) = 0, \quad \text{Var}(\tau(G(n,p))) \leq \binom{n}{3}. \quad (16)$$

Notice that this agrees with our assertion from section 1.3 that the variance of the signed triangle statistic is on the order of n^3 for $G \sim G(n, p)$.

3.2 A lower bound on $\mathbb{E}(\tau(G(n, p, d)))$

Since computing the exact expectation of the signed triangle statistic under $G \sim G(n, p, d)$ is challenging, we instead settle for a lower bound. For fixed $p \in (0, 1)$, this bound is a function of n and d , as summarized in the following lemma:

Lemma 3. *For every $0 < p < 1$, there exists a constant C_p (depending only on p) such that that for all n and d we have*

$$\mathbb{E}(\tau(G(n, p, d))) \geq \binom{n}{3} \frac{C_p}{\sqrt{d}}.$$

Proof. Let $i, j, k \in [n]$ be distinct vertices in the random graph G . We note that it is sufficient to consider $\mathbb{E}(\tau_{G(n,p)}(i, j, k))$ since, by linearity of the integral,

$$\begin{aligned}\mathbb{E}(\tau(G(n, p, d))) &= \mathbb{E}\left(\sum_{\{i, j, k\} \in \binom{[n]}{3}} \tau_{G(n, p, d)}(i, j, k)\right) \\ &= \sum_{\{i, j, k\} \in \binom{[n]}{3}} \mathbb{E}(\tau_{G(n, p, d)}(i, j, k)) \\ &= \binom{n}{3} \mathbb{E}(\tau_{G(n, p, d)}(i, j, k)).\end{aligned}$$

The final equality holds since by definition of $G \sim G(n, p, d)$, the random variables $\tau_{G(n, p, d)}(i, j, k) = \bar{A}_{i,j} \bar{A}_{i,k} \bar{A}_{j,k}$ are identically-distributed for every $\{i, j, k\} \in \binom{[n]}{3}$.

With this in mind, we rewrite $\mathbb{E}(\tau_{G(n,p)}(i, j, k))$ as follows:

$$\begin{aligned}\mathbb{E}(\tau_{G(n,p,d)}(i, j, k)) &= \mathbb{E}(\bar{A}_{i,j} \bar{A}_{i,k} \bar{A}_{j,k}) \\ &= \mathbb{E}(A_{i,j} - p)(A_{i,k} - p)(A_{j,k} - p) \\ &= \mathbb{E}(A_{i,j} A_{i,k} A_{j,k} - p(A_{i,j} A_{i,k} + A_{i,j} A_{j,k} + A_{i,k} A_{j,k}) \\ &\quad + p^2(A_{i,j} + A_{i,k} + A_{j,k}) - p^3) \\ &= \mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) - 3p\mathbb{E}(A_{i,j} A_{i,k}) + 3p^2\mathbb{E}(A_{i,j}) - p^3 && \text{linearity} \\ &= \mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) - 3p\mathbb{E}(A_{i,j})^2 + 3p^2\mathbb{E}(A_{i,j}) - p^3 && A_{i,j} \perp A_{i,k} \\ &= \mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) - p^3 && P(A_{i,j}) = p\end{aligned}$$

Recall that for $G \sim G(n, p, d)$, then $A_{i,j} = \mathbb{I}_{\{\langle X_i, X_j \rangle \geq t_{p,d}\}}$. And so

$$\begin{aligned}\mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) &= \mathbb{E}(\mathbb{I}_{\{\langle X_i, X_j \rangle \geq t_{p,d}\}} \mathbb{I}_{\{\langle X_i, X_k \rangle \geq t_{p,d}\}} \mathbb{I}_{\{\langle X_j, X_k \rangle \geq t_{p,d}\}}) \\ &= \mathbb{E}(\mathbb{I}_{\{\langle X_i, X_j \rangle \geq t_{p,d}, \langle X_i, X_k \rangle \geq t_{p,d}, \langle X_j, X_k \rangle \geq t_{p,d}\}}) \\ &= \mathbb{P}(\{\langle X_i, X_j \rangle \geq t_{p,d}, \langle X_i, X_k \rangle \geq t_{p,d}, \langle X_j, X_k \rangle \geq t_{p,d}\})\end{aligned}$$

since the set in the second line is \mathcal{F} -measurable. In section 2 we defined this set to be $E = \{\langle X_i, X_j \rangle \geq t_{p,d}, \langle X_i, X_k \rangle \geq t_{p,d}, \langle X_j, X_k \rangle \geq t_{p,d}\}$ and found an upper bound on its probability for each fixed $0 < p < 1$ in Lemma 1. That is, we have expressed the expectation of $\tau(G(n, p, d))$ in terms of that of $T(G(n, p, d))$, the triangle statistic for $G \sim G(n, p, d)$.

Now exploiting the result from Lemma 1, we have that

$$\begin{aligned}\mathbb{E}(\tau_{G(n,p,d)}(i, j, k)) &= \mathbb{E}(A_{i,j} A_{i,k} A_{j,k}) - p^3 \\ &= \mathbb{P}(E) - p^3 \\ &\geq p^3 \left(1 + \frac{C_p}{\sqrt{d}}\right) - p^3 \\ &= \frac{C_p}{\sqrt{d}}\end{aligned}$$

for some constant $C_p > 0$ that depends only on $0 < p < 1$.

Thus, we deduce

$$\mathbb{E}(\tau(G(n, p, d))) = \binom{n}{3} \mathbb{E}(\tau_{G(n,p,d)}(i, j, k)) \geq \binom{n}{3} \frac{C_p}{\sqrt{d}}. \quad (17)$$

□

3.3 An upper bound on $\text{Var}(\tau(G(n, p, d)))$

The most challenging part in our proof of Theorem 2 is establishing an upper bound on the variance of the signed triangle statistic for $G \sim G(n, p, d)$. Prior to considering $\text{Var}(\tau(G(n, p, d)))$, Bubeck and colleagues find an upper bound on $\mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)\tau_{G(n, p, d)}(1, 2, 4))$. Notice that $\{1, 2, 3, 4\}$ are chosen arbitrarily among $\{i, j, k, l\} \in \binom{[n]}{4}$ since each $\tau_{G(n, p, d)}(i, j, k)\tau_{G(n, p, d)}(i, j, l)$ is identically-distributed. We state the result from [1] in the following lemma, and subsequently give a proof.

Lemma 4. *For every $p \in [0, 1]$ we have that*

$$\mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)\tau_{G(n, p, d)}(1, 2, 4)) \leq \pi^2/d.$$

Proof. Numerous pieces of our proof mirror those present in Lemma 1, such as rotating the coordinate system in \mathbb{R}^d by the rotational invariance of Q and finding the measure of a hyperspherical cap of \mathbb{S}^{d-1} .

Recall from our proof of Lemma 1 that for map T from the product space $(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1} \times \mathbb{S}^{d-1}), Q \times Q)$ to $([-1, 1], \mathcal{B}([-1, 1]))$ defined by $T(x, y) := \langle x, y \rangle$, there exists a conditional probability of $Q \times Q$ under T that is unique $Q \times Q$ -almost surely. Further, because Q is defined to be the image measure of \mathbb{P} under each random variable X_i , then there exists a conditional probability of \mathbb{P} under $\tilde{T}(\omega) = \langle X_i(\omega), X_j(\omega) \rangle$.

Using the existence of the conditional expectation, we define

$$g(x) := \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)\tau_{G(n, p, d)}(1, 2, 4) | \langle X_1, X_2 \rangle = x), \quad x \in [-1, 1]. \quad (18)$$

Just as before, the expectation denotes the integral of $\tau_{G(n, p, d)}(1, 2, 3)\tau_{G(n, p, d)}(1, 2, 4)$ with respect to the conditional probability measure $\mathbb{P}_{\langle X_1, X_2 \rangle = x}$.

Since the coordinate system in \mathbb{R}^d is unspecified and spherical measure Q is rotationally invariant, then we can assume, without loss of generality, that $X_1 = e_1$ and $X_2 = xe_1 + \sqrt{1-x^2}e_2$. This is equivalent to stating that for some fixed coordinate system, we may rotate the axes so that $X_1 = e_1$ and $X_2 = xe_1 + \sqrt{1-x^2}e_2$. Thus,

$$\begin{aligned} g(x) &= \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)\tau_{G(n, p, d)}(1, 2, 4) | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2) \\ &= \mathbb{E}(\bar{A}_{1,2}^2 \bar{A}_{2,3} \bar{A}_{1,3} \bar{A}_{1,4} \bar{A}_{2,4} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2). \end{aligned}$$

Since $X_3 \perp X_4$ as well as X_3, X_4 identically-distributed, then we have

$$\begin{aligned} g(x) &= \mathbb{E}(\bar{A}_{1,2} \bar{A}_{2,3} \bar{A}_{1,3} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2)^2 \\ &= \mathbb{E}\left((p\mathbb{I}_{x < t_{p,d}} + (1-p)\mathbb{I}_{x \geq t_{p,d}}) \bar{A}_{2,3} \bar{A}_{1,3} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right)^2 \quad \text{def of } \bar{A}_{1,2} \\ &= (p^2\mathbb{I}_{x < t_{p,d}} + (1-p)^2\mathbb{I}_{x \geq t_{p,d}}) \mathbb{E}(\bar{A}_{2,3} \bar{A}_{1,3} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2)^2. \quad \text{linearity} \end{aligned} \quad (19)$$

At this point, Bubeck and colleagues introduce the normalized Haar measure on the sphere, σ . However, as discussed in [6], we note that the normalized Haar measure on \mathbb{S}^{d-1} coincides *exactly* with our spherical measure Q , which we defined using a result from [2]. This is because, as discussed just above, our spherical measure is uniform on \mathbb{S}^{d-1} , and so it is clearly invariant with respect to rotation.

Now, we define the hyperspherical cap

$$B_x := \{y \in \mathbb{S}^{d-1} : \langle y, X_2 \rangle \geq t_{p,d}\} = \left\{y \in \mathbb{S}^{d-1} : \langle y, xe_1 + \sqrt{1-x^2}e_2 \rangle \geq t_{p,d}\right\} \quad (20)$$

and let $S_x := B_1 \cap B_x$ be the intersection of two such caps. Recall from our definition of $G \sim G(n, p, d)$ that for each X_i , $i \in [n]$, then $\mathbb{P}(\{\omega : \langle X_i(\omega), X_2(\omega) \rangle \geq t_{p,d}\}) = p$. And so because $\sigma = Q$ is the image measure of \mathbb{P} under X_i (that is, Q is the distribution of random variable X_i), then

$$\begin{aligned}\sigma(B_x) &= Q\left(\left\{y \in \mathbb{S}^{d-1} : \langle y, xe_1 + \sqrt{1-x^2}e_2 \rangle \geq t_{p,d}, \right\}\right) \\ &= \mathbb{P}\left(\left\{\omega \in \Omega : \langle X_i(\omega), xe_1 + \sqrt{1-x^2}e_2 \rangle \geq t_{p,d}, \right\}\right) \\ &= p.\end{aligned}$$

We now use these B_x to simplify our integral in (19):

$$\begin{aligned}&\mathbb{E}\left(\overline{A}_{2,3}\overline{A}_{1,3}|X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &= \mathbb{E}\left((A_{2,3} - p)(A_{1,3} - p)|X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &= \mathbb{E}\left((\mathbb{I}_{\{\langle X_2, X_3 \rangle \geq t_{p,d}\}} - p)(\mathbb{I}_{\{\langle X_1, X_3 \rangle \geq t_{p,d}\}} - p) | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right).\end{aligned}$$

By linearity of the conditional expectation,

$$\begin{aligned}&= (1-p)^2\mathbb{P}\left(\langle X_2, X_3 \rangle \geq t_{p,d}, \langle X_1, X_3 \rangle \geq t_{p,d} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &- p(1-p)\mathbb{P}\left(\langle X_2, X_3 \rangle \geq t_{p,d}, \langle X_1, X_3 \rangle < t_{p,d} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &- p(1-p)\mathbb{P}\left(\langle X_2, X_3 \rangle < t_{p,d}, \langle X_1, X_3 \rangle \geq t_{p,d} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &+ p^2\mathbb{P}\left(\langle X_2, X_3 \rangle < t_{p,d}, \langle X_1, X_3 \rangle < t_{p,d} | X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right) \\ &= (1-p)^2\sigma(S_x) - p(1-p)\sigma(B_x \setminus S_x) - p(1-p)\sigma(B_1 \setminus S_x) + p^2\sigma(\mathbb{S}^{d-1} \setminus (B_1 \cup B_x)).\end{aligned}$$

Here we use the fact that from (20), $\sigma(B_1) = \sigma(\{y \in \mathbb{S}^{d-1} : \langle y, e_1 \rangle \geq t_{p,d}\}) = \sigma(\{y \in \mathbb{S}^{d-1} : \langle y, X_1 \rangle \geq t_{p,d}\}) = \mathbb{P}(\{\omega : \langle X_3(\omega), X_1(\omega) \rangle \geq t_{p,d}\})$. And from the finite additivity of σ ,

$$\begin{aligned}&= (1-p)^2\sigma(S_x) - p(1-p)(p - \sigma(S_x)) - p(1-p)(p - \sigma(S_x)) + p^2\sigma(\mathbb{S}^{d-1} \setminus (B_1 \cup B_x)) \\ &= (1-p)^2\sigma(S_x) - 2p(1-p)(p - \sigma(S_x)) + p^2(1 - 2p + \sigma(S_x)) = \sigma(S_x) - p^2.\end{aligned}$$

Observe that the penultimate equality holds by finite additivity and inclusion-exclusion:

$$\begin{aligned}\sigma(\mathbb{S}^{d-1} \setminus (B_1 \cup B_x)) &= \sigma(\mathbb{S}^{d-1}) - \sigma(B_1 \cup B_x) \\ &= \sigma(\mathbb{S}^{d-1}) - (\sigma(B_1) + \sigma(B_x) - \sigma(B_1 \cap B_x)) \\ &= 1 - (2p - \sigma(S_x)).\end{aligned}$$

Putting this all together with (19), we get

$$\begin{aligned}g(x) &= (p^2\mathbb{I}_{x < t_{p,d}} + (1-p)^2\mathbb{I}_{x \geq t_{p,d}}) \mathbb{E}\left(\overline{A}_{2,3}\overline{A}_{1,3}|X_1 = xe_1, X_2 = xe_1 + \sqrt{1-x^2}e_2\right)^2 \\ &= (p^2\mathbb{I}_{x < t_{p,d}} + (1-p)^2\mathbb{I}_{x \geq t_{p,d}}) (\sigma(S_x) - p^2)^2 \leq (\sigma(S_x) - p^2)^2.\end{aligned}\tag{21}$$

Using our definition for g as well as the conditional probability of $Q \times Q$ under T , then we can bound $\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3)\tau_{G(n,p,d)}(1, 2, 4))$ by

$$\begin{aligned}&\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3)\tau_{G(n,p,d)}(1, 2, 4)) \\ &= \mathbb{E}^x(\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3)\tau_{G(n,p,d)}(1, 2, 4)|\langle X_1, X_2 \rangle = x)) \quad \text{def of conditional probability} \\ &= \mathbb{E}^x(g(x))\end{aligned}$$

$$\leq \mathbb{E}^x ((\sigma(S_x) - p^2)^2). \quad \text{monotonicity} \quad (22)$$

Next, we claim that

$$\mathbb{E}^x(\sigma(S_x)) = p^2,$$

which will help us simplify our bound in (22).

To see why this is true, recall that

$$\begin{aligned} \mathbb{E}^x(\sigma(S_x)) &= \mathbb{E}^x(\sigma(B_1 \cap B_x)) \\ &= \mathbb{E}^x \left(\sigma \left(\left\{ y \in \mathbb{S}^{d-1} : \langle y, e_1 \rangle, \langle y, x e_1 + \sqrt{1-x^2} e_2 \rangle \geq t_{p,d} \right\} \right) \right). \end{aligned}$$

Moreover, remember that for X_1, X_2 independent random vectors, uniformly-distributed on \mathbb{S}^{d-1} , we rotated the coordinate axes of \mathbb{R}^d so that $X_1 = e_1, X_2 = x e_1 + \sqrt{1-x^2} e_2$. Therefore, letting Y and Z be independent, uniformly-distributed vectors on \mathbb{S}^{d-1} , then $\sigma(S_x)$ has the same distribution as $\sigma(\{y \in \mathbb{S}^{d-1} : \langle y, Y \rangle, \langle y, Z \rangle \geq t_{p,d}\})$. That is, we rewrite the previous expression as

$$\begin{aligned} &\mathbb{E}^x(\sigma(S_x)) \\ &= \mathbb{E}^x \left(\sigma \left(\left\{ y \in \mathbb{S}^{d-1} : \langle y, e_1 \rangle, \langle y, x e_1 + \sqrt{1-x^2} e_2 \rangle \geq t_{p,d} \right\} \right) \right) \\ &= \mathbb{E}^x \left(\sigma(\{y \in \mathbb{S}^{d-1} : \langle y, Y \rangle, \langle y, Z \rangle \geq t_{p,d}\}) \right). \end{aligned}$$

By defining W to be a random vector with the uniform distribution on \mathbb{S}^{d-1} , independent of Y and Z , we further simplify this expression as

$$\begin{aligned} &= \mathbb{E}^x \left(Q(\{y \in \mathbb{S}^{d-1} : \langle y, Y \rangle, \langle y, Z \rangle \geq t_{p,d}\}) \right) \\ &= \mathbb{E}^x \left(\mathbb{P}(\{\omega \in \Omega : \langle W, Y \rangle, \langle W, Z \rangle \geq t_{p,d}\}) \right) \quad \text{def of measure } Q \\ &= \mathbb{P}(\{\omega \in \Omega : \langle W, Y \rangle, \langle W, Z \rangle \geq t_{p,d}\}). \end{aligned}$$

Additionally, because Y and Z are independent random vectors, then so are $\langle W, Y \rangle, \langle W, Z \rangle : (\Omega \times \Omega, \mathcal{F} \otimes \mathcal{F}, \mathbb{P} \times \mathbb{P}) \rightarrow ([-1, 1], \mathcal{B}([-1, 1]))$. Thus, we deduce

$$\begin{aligned} \mathbb{E}^x(\sigma(S_x)) &= \mathbb{P}(\{\omega \in \Omega : \langle W, Y \rangle \geq t_{p,d}\})^2 \\ &= p^2. \quad \text{def of } t_{p,d} \end{aligned}$$

With this result, we may reexpress the bound in (22) as

$$\begin{aligned} \mathbb{E} \left(\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 2, 4) \right) &\leq \mathbb{E}^x ((\sigma(S_x) - p^2)^2) \\ &= \mathbb{E}^x ((\sigma(S_x) - \mathbb{E}(\sigma(S_x)))^2) \\ &= \text{Var}(\sigma(S_x)). \end{aligned} \quad (23)$$

Recall that since $X_1 = e_1, X_2 = x e_1 + \sqrt{1-x^2} e_2$, then $\langle X_1, X_2 \rangle = x = X_{2,1}$. That is, $\langle X_1, X_2 \rangle$ has the same distribution as the one-dimensional marginal of a uniform random point on \mathbb{S}^{d-1} . Let X be a random variable with the same distribution as $\langle X_1, X_2 \rangle$. Then from (23) we can write $\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 2, 4)) \leq \text{Var}(\sigma(S_X))$. Further, let us define $h(x) := \sigma(S_x)$. That is, $h(x)$ is the spherical measure of the intersection of hyperspherical caps B_1 and B_x for $X = x$.

As a short digression, we once again let Y denote a random vector with the uniform distribution on \mathbb{S}^{d-1} . We know that $\mathbb{P}(\{\omega : \|Y(\omega)\|^2 = 1\}) = 1$ by definition of $Y \in \mathbb{S}^{d-1}$, and so

$$1 = \mathbb{E}(\|Y\|^2)$$

$$\begin{aligned}
&= \mathbb{E} \left(\sum_{i=1}^n Y_i^2 \right) && \text{def of } \|\cdot\| \\
&= \sum_{i=1}^n \mathbb{E}(Y_i^2) && \text{linearity} \\
&= d \times \mathbb{E}(X^2).
\end{aligned} \tag{24}$$

Note that the fourth equality holds since each of the one-dimensional marginals of random vector Y is identically-distributed with the same distribution as X . Therefore we have that $\mathbb{E}(X^2) = \frac{1}{d}$.

From here, Bubeck and colleagues claim that proving the lemma reduces to showing that for all $x, y \in [-1, 1]$,

$$|h(x) - h(y)| \leq 2 |\arcsin(x) - \arcsin(y)|. \tag{25}$$

This is because $f(x) = \arcsin(x)$ is convex on $x \in [0, 1]$, meaning that we have

$$\frac{\arcsin(x) - \arcsin(0)}{x - 0} \leq \frac{\arcsin(1) - \arcsin(0)}{1 - 0} \quad \text{for } x \in (0, 1].$$

Further, since $f(x)$ is an odd function, then we can use the above inequality to bound $f(x)$ for all $x \in [-1, 1]$:

$$\begin{aligned}
\frac{|\arcsin(x)|}{|x|} &\leq \arcsin(1) \\
\Rightarrow |\arcsin(x)| &\leq |x| \arcsin(1) \leq \frac{\pi}{2} |x|.
\end{aligned} \tag{26}$$

Pulling the above statements together, we have

$$\begin{aligned}
\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 2, 4)) &\leq \text{Var}(h(X)) && \text{from (23)} \\
&\leq \mathbb{E}(h(X) - h(0))^2 \\
&\leq 4 \mathbb{E}(\arcsin(X))^2 && \text{from (25)} \\
&\leq \pi^2 \mathbb{E}(X^2) && \text{from (26)} \\
&= \frac{\pi^2}{d}. && \text{from (24)}
\end{aligned}$$

Note that the final three inequalities invoke monotonicity of the integral. Looking at the last line, we see this is indeed the upper bound on $\mathbb{E}(\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 2, 4))$ that we wished to derive. And so, in order to conclude our proof of Lemma 4, it only remains to substantiate (25).

Proving the bound on $|h(x) - h(y)|$ for $x, y \in [-1, 1]$ is admittedly unintuitive. In their paper, Bubeck and colleagues use the approach of foliation from differential geometry. That is, they split \mathbb{S}^{d-1} a $d-1$ manifold into disjoint, connected subsets called leaves that satisfy certain properties [4]. The particular properties of this foliation are not of key importance for understanding the proof. Rather, we need the specific foliation that the authors choose. This foliation is

$$\mathbb{S}^{d-1} = \bigcup_{z \in \mathbf{B}^{d-2}} W_z, \tag{27}$$

where \mathbf{B}^{d-2} is the $(d-2)$ -dimensional Euclidean unit ball and $W_z = \{(x, y, z) : x^2 + y^2 = 1 - |z|^2\}$. Then for each $\theta \in \mathbb{S}^{d-1}$ we have a corresponding decomposition $\theta \mapsto (x_\theta, y_\theta, z_\theta)$.

Let μ denote the measure on \mathbf{B}^{d-2} defined to be the image measure of σ under the coordinate map $\theta \mapsto z_\theta$. Moreover, let σ_z denote the uniform measure on W_z . And so we may express our original probability space $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}), \sigma)$ as the product space $(\mathbf{B}^{d-2} \times W_z, \mathcal{B}(\mathbf{B}^{d-2}) \otimes \mathcal{B}(W_z), \mu \times \sigma_z)$. In other words, we use the

foliation from (27) to “deconstruct” our probability space as a product space. And so because μ, σ_z are finite measures, then we may invoke Fubini’s theorem to write

$$\sigma(f(\theta)) = \mu^z(\sigma_z(f(x, y, z)))$$

for every $f \in \mathcal{L}^1(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}), \sigma)$ [7].

Clearly $\mathbb{I}_{S_x} = \mathbb{I}_{B_1 \cap B_x}$ is integrable for every $x \in [-1, 1]$ since each $B_1, B_x \in \mathcal{B}(\mathbb{S}^{d-1})$, and thus \mathbb{I}_{S_x} is a simple function. Accordingly, for all $x, y \in (0, 1)$, we use above expression to write

$$\begin{aligned} & \sigma(S_x) - \sigma(S_y) \\ &= \sigma(\mathbb{I}_{S_x} - \mathbb{I}_{S_y}) \\ &= \mu^z(\sigma_z(\mathbb{I}_{S_x} - \mathbb{I}_{S_y})) \\ &= \mu^z(\sigma_z(\mathbb{I}_{S_x}) - \sigma_z(\mathbb{I}_{S_y})) && \text{linearity of } \sigma_z \\ &= \mu^z(\sigma_z(S_x \cap W_z) - \sigma_z(S_y \cap W_z)). \end{aligned} \tag{28}$$

Now, let us take $A \Delta B = (A \setminus B) \cup (B \setminus A)$ to be the symmetric difference between sets A and B . Then by the absolute value properties of the integral, we have

$$\begin{aligned} |\sigma_z(S_x \cap W_z) - \sigma_z(S_y \cap W_z)| &= |\sigma_z(\mathbb{I}_{S_x \cap W_z} - \mathbb{I}_{S_y \cap W_z})| \\ &\leq \sigma_z(|\mathbb{I}_{S_x \cap W_z} - \mathbb{I}_{S_y \cap W_z}|) \\ &= \sigma_z((S_x \Delta S_y) \cap W_z) \\ &= \sigma_z(((B_1 \cap B_x) \Delta (B_1 \cap B_y)) \cap W_z) && \text{def of } S_x, S_y \\ &\leq \sigma_z((B_x \Delta B_y) \cap W_z). && \text{monotonicity} \end{aligned} \tag{29}$$

Taking a step back, we notice that for fixed $z \in \mathbf{B}^{d-2}$, then $W_z = \{(l, q, z) \in \mathbb{S}^{d-1} : l^2 + q^2 = 1 - |z|^2\}$ defines a circle with radius $\sqrt{1 - |z|^2}$. If this radius is strictly less than $t_{p,d}$, then for every $x \in [-1, 1]$ and for every $y \in B_x = \{y \in \mathbb{S}^{d-1} : \langle y, x e_1 + \sqrt{1 - x^2} e_2 \rangle \geq t_{p,d}\}$, we have $\langle y, x e_1 + \sqrt{1 - x^2} e_2 \rangle \geq t_{p,d} \Rightarrow y_1 x + y_2 \sqrt{1 - x^2} \geq t_{p,d} \Rightarrow y_1^2 + y_2^2 \geq t_{p,d}^2$. That is, B_x intersects nowhere with the circle W_z . On the other hand, if the radius $\sqrt{1 - |z|^2} \geq t_{p,d}$, then

$$B_x \cap W_z = \left\{ y \in \mathbb{S}^{d-1} : y_1^2 + y_2^2 = 1 - |z|^2, y_1 x + y_2 \sqrt{1 - x^2} \geq t_{p,d} \right\}$$

defines an arc of the circle W_z for each $x \in [-1, 1]$. And so we have that $B_x \cap W_z, B_y \cap W_z$ are two arcs of the circle W_z .

To complete their argument, Bubeck and colleagues claim that these two arcs can be transformed into one another by a rotation of angle $|\arcsin(x) - \arcsin(y)|$ [1]. Therefore, we have

$$\sigma_z((B_x \Delta B_y) \cap W_z) \leq 2 |\arcsin(x) - \arcsin(y)| \sigma_z(W_z). \tag{30}$$

In other words, the spherical measure of the set difference between B_x and B_y is at most the measure of the region by which one must rotate to transform B_x into B_y .

To wrap things up, we combine the above results as follows for all $x, y \in [-1, 1]$:

$$\begin{aligned} |\sigma(S_x) - \sigma(S_y)| &= |\mu^z(\sigma_z(S_x \cap W_z) - \sigma_z(S_y \cap W_z))| && \text{from (28)} \\ &\leq \mu^z(|\sigma_z(S_x \cap W_z) - \sigma_z(S_y \cap W_z)|) \\ &\leq \mu^z(\sigma_z((B_x \Delta B_y) \cap W_z)) && \text{(29) and monotonicity} \\ &\leq \mu^z(2 |\arcsin(x) - \arcsin(y)| \sigma_z(W_z)) && \text{(30) and monotonicity} \end{aligned}$$

$$\begin{aligned}
&= 2 |\arcsin(x) - \arcsin(y)| \mu^z(\sigma_z(W_z)) \\
&= 2 |\arcsin(x) - \arcsin(y)| \sigma(\mathbb{S}^{d-1}) && \text{from (27)} \\
&= 2 |\arcsin(x) - \arcsin(y)|. && \text{def of spherical measure}
\end{aligned}$$

Notice that this is exactly the statement (25) that remained to be proven. And so, by our previous argument, we have concluded our proof of Lemma 4. \square

Now equipped with the result from Lemma 4, we are ready to give an upper bound on the variance of the signed triangle statistic for $G \sim G(n, p, d)$. In particular, we use linearity to rewrite

$$\begin{aligned}
\text{Var}(\tau(G(n, p, d))) &= \text{Var} \left(\sum_{\{i, j, k\} \in \binom{[n]}{3}} \tau_{G(n, p, d)}(i, j, k) \right) \\
&= \mathbb{E} \left(\sum_{\{i, j, k\} \in \binom{[n]}{3}} \sum_{\{i', j', k'\} \in \binom{[n]}{3}} \tau_{G(n, p, d)}(i, j, k) \tau_{G(n, p, d)}(i', j', k') \right) \\
&\quad - \mathbb{E} \left(\sum_{\{i, j, k\} \in \binom{[n]}{3}} \tau_{G(n, p, d)}(i, j, k) \right)^2 \\
&= \sum_{\{i, j, k\} \in \binom{[n]}{3}} \sum_{\{i', j', k'\} \in \binom{[n]}{3}} \left(\mathbb{E}(\tau_{G(n, p, d)}(i, j, k) \tau_{G(n, p, d)}(i', j', k')) \right. \\
&\quad \left. - \mathbb{E}(\tau_{G(n, p, d)}(i, j, k)) \mathbb{E}(\tau_{G(n, p, d)}(i', j', k')) \right). \tag{31}
\end{aligned}$$

Recall that $\tau_{G(n, p, d)}(i, j, k) = \bar{A}_{i,j} \bar{A}_{i,k} \bar{A}_{j,k}$, where $\bar{A}_{i,j} = A_{i,j} - p$. And since $G \sim G(n, p, d)$, then we have $A_{i,j} = \mathbb{I}_{\{\langle X_i, X_j \rangle \geq t_{p,d}\}}$. Therefore, because X_1, \dots, X_n are independent and identically-distributed random vectors on \mathbb{S}^{d-1} , then it suffices to consider only four terms in our calculation (31) for the variance.

$$\begin{aligned}
W_1 &:= \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3) \tau_{G(n, p, d)}(1, 2, 3)) - \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)) \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)), \\
W_2 &:= \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3) \tau_{G(n, p, d)}(1, 2, 4)) - \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)) \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 4)), \\
W_3 &:= \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3) \tau_{G(n, p, d)}(1, 4, 5)) - \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)) \mathbb{E}(\tau_{G(n, p, d)}(1, 4, 5)), \\
W_4 &:= \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3) \tau_{G(n, p, d)}(4, 5, 6)) - \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)) \mathbb{E}(\tau_{G(n, p, d)}(4, 5, 6)).
\end{aligned}$$

That is, each of the terms in (31) corresponds to one of W_1, W_2, W_3, W_4 . We proceed by bounding each term individually.

For W_1 , we make the observation that $\mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)^2) \geq \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3))^2 \geq 0$ by Jensen's inequality. This is no more than the statement that the variance of random variable $\tau_{G(n, p, d)}(1, 2, 3)$ is non-negative. Further, we notice that by the definition of $\tau_{G(n, p, d)}(1, 2, 3)$ from above, then $-1 < \bar{A}_{i,j} < 1 \Rightarrow -1 < \tau_{G(n, p, d)}(1, 2, 3) < 1 \Rightarrow \tau_{G(n, p, d)}(1, 2, 3)^2 < 1$. Thus, by monotonicity of the integral we have $\mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)^2) \leq \mathbb{E}(\mathbb{I}_\Omega) = 1$. Finally, from our first statement we deduce that $0 \leq \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3)^2) - \mathbb{E}(\tau_{G(n, p, d)}(1, 2, 3))^2 \leq 1$, and so $W_1 \leq 1$.

To bound W_2 , we simply invoke the result from Lemma 4 to say that $W_2 \leq \frac{\pi^2}{d}$.

W_3 , on the other hand, requires slightly more work. By a similar vein to the conditional probability whose existence we proved in section 2, one can show that the conditional probability of \mathbb{P} under random variable X_1 indeed exists. In particular, one can consider the identity mapping from the probability space $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}), Q)$ to $(\mathbb{S}^{d-1}, \mathcal{B}(\mathbb{S}^{d-1}), V)$, where V denotes the image measure. Since $(\mathbb{S}^{d-1}, \|\cdot\|_2)$ is a “nice”

metric space and the identity map is $Q \setminus V$ -measurable, then the conditional probability of Q under the identity map exists. Moreover, because Q is defined to be the image measure of \mathbb{P} under X_i , we then deduce that the conditional probability of \mathbb{P} under X_i exists.

Now, appealing to the definition of the conditional probability, we write

$$\mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 4, 5)) = \mathbb{E}^x (\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 4, 5) | X_1 = x)).$$

Upon conditioning on $X_1 = x$, we notice that $\tau_{G(n,p,d)}(1, 2, 3)$ is a function of X_2, X_3 , whereas $\tau_{G(n,p,d)}(1, 4, 5)$ is a function of X_4, X_5 . And so since X_2, X_3, X_4, X_5 are independent, then $\tau_{G(n,p,d)}(1, 2, 3) \perp \tau_{G(n,p,d)}(1, 4, 5)$, conditioned on $X_1 = x$. That is, we have

$$\begin{aligned} & \mathbb{E}^x (\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) \tau_{G(n,p,d)}(1, 4, 5) | X_1 = x)) \\ &= \mathbb{E}^x (\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) | X_1 = x) \mathbb{E}_x (\tau_{G(n,p,d)}(1, 4, 5) | X_1 = x)). \end{aligned}$$

Now consider the term $\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) | X_1 = x)$. We make the assertion that $\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) | X_1 = x) = \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3))$. Simply, this tells us that conditioning on X_1 does not provide any additional “information” about the integral of $\tau_{G(n,p,d)}(1, 2, 3)$. This makes intuitive sense, since knowing the value of vector X_1 on \mathbb{S}^{d-1} does not tell us about its inner product with either of X_2 or X_3 . Mathematically, since the spherical measure σ is invariant with respect to rotation, then for any $X_1 = x$ we can rotate our coordinate axes so as to satisfy $X_1 = e_1$ while preserving the distribution of $\tau_{G(n,p,d)}(1, 2, 3)$. Therefore,

$$\begin{aligned} & \mathbb{E}^x (\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3) | X_1 = x) \mathbb{E}_x (\tau_{G(n,p,d)}(1, 4, 5) | X_1 = x)) \\ &= \mathbb{E}^x (\mathbb{E}_x (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E}_x (\tau_{G(n,p,d)}(1, 4, 5))) \\ &= \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E} (\tau_{G(n,p,d)}(1, 4, 5)). \end{aligned}$$

And so $W_3 = \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E}_x (\tau_{G(n,p,d)}(1, 4, 5)) - \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E}_x (\tau_{G(n,p,d)}(1, 4, 5)) = 0$.

Lastly, we show that $W_4 = 0$ with another independence argument. Namely, X_1, \dots, X_6 independent implies $\tau_{G(n,p,d)}(1, 2, 3) \perp \tau_{G(n,p,d)}(4, 5, 6)$. Consequently, we deduce $W_4 = \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E} (\tau_{G(n,p,d)}(4, 5, 6)) - \mathbb{E} (\tau_{G(n,p,d)}(1, 2, 3)) \mathbb{E} (\tau_{G(n,p,d)}(4, 5, 6)) = 0$, as desired.

Having considered each of the four terms in the variance of $\tau(G(n, p, d))$, we turn our attention back to (31). In particular, we notice that there are $\binom{n}{3}$ terms of the form W_1 , since this binomial coefficient counts the number of ways to select 3 unordered vertices among n total. Further, there are $\binom{n}{4} \binom{4}{2}$ terms of the form W_2 because there are $\binom{n}{4}$ ways to choose 4 unordered vertices among n total and then $\binom{4}{2}$ ways to order them as in W_2 . Thus, we conclude that

$$\text{Var}(\tau(G(n, p, d))) \leq \binom{n}{3} W_1 + \binom{n}{4} \binom{4}{2} W_2 \leq \binom{n}{3} + \binom{n}{4} \binom{4}{2} \frac{\pi^2}{d} \leq n^3 + \frac{3n^4}{d}. \quad (32)$$

And so we have found an upper bound on the variance of the signed triangle statistic under $G \sim G(n, p, d)$. While this calculation was admittedly long-winded, it will be of key importance for finding a lower bound on $\text{TV}(\tau(G(n, p)), \tau(G(n, p, d)))$ in the final section of our proof.

3.4 Concluding the Proof of Theorem 2

From (16), (17), and (32), we have

$$\mathbb{E} (\tau(G(n, p))) = 0, \quad \text{Var} (\tau(G(n, p))) \leq \binom{n}{3} \leq n^3$$

as well as

$$\mathbb{E}(\tau(G(n, p, d))) \geq \binom{n}{3} \frac{C_p}{\sqrt{d}}, \quad \text{Var}(\tau(G(n, p, d))) \leq \binom{n}{3} \leq n^3 + \frac{3n^4}{d}$$

for $C_p > 0$ a constant in $p \in (0, 1)$. Thus, as we suggested in the introduction to our proof, the difference $\mathbb{E}(\tau(G(n, p))) - \mathbb{E}(\tau(G(n, p, d)))$ exceeds the individual standard deviations of these statistics.

In particular, we have that

$$\max \left\{ \text{Var}(\tau(G(n, p))), \text{Var}(\tau(G(n, p, d))) \right\} \leq n^3 + \frac{3n^4}{d}. \quad (33)$$

And so we see that the maximum of the individual standard deviations is on the order of n^2 , whereas our lower bound on $\mathbb{E}(\tau(G(n, p, d))) - \mathbb{E}(\tau(G(n, p)))$ is on the order of n^3 .

With this fact, we compute a lower bound on the total variation distance between $\tau(G(n, p, d))$ and $\tau(G(n, p))$. Our goal is to find a set $B \in \mathcal{B}(\mathbb{R})$ such that $\mathbb{P}\{\omega : \tau(G(n, p)) \in B\} \rightarrow 1$ as $\frac{d}{n^3} \rightarrow 0$ whereas $\mathbb{P}\{\omega : \tau(G(n, p, d)) \in B\} \rightarrow 0$ as $\frac{d}{n^3} \rightarrow 0$. This, in turn, would suggest that the total variation distance between the distribution of τ under $G(n, p)$ and $G(n, p, d)$ tends to 1 as $\frac{d}{n^3} \rightarrow 0$. From our above discussion, $(-\infty, \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))) \in \mathcal{B}(\mathbb{R})$ is a natural candidate for this set since $\frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))$ grows asymptotically away from $\mathbb{E}(\tau(G(n, p)))$ and at a faster rate than either of $\text{Var}(\tau(G(n, p)))$ or $\text{Var}(\tau(G(n, p, d)))$.

With this proposed set in mind, we now compute an upper bound on $\mathbb{P}\{\tau(G(n, p, d)) \leq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\}$:

$$\begin{aligned} & \mathbb{P}\left\{\tau(G(n, p, d)) \leq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} \leq \mathbb{P}\left\{|\tau(G(n, p, d)) - \mathbb{E}(\tau(G(n, p, d)))| \geq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} \\ &= \mathbb{P}\left\{|\tau(G(n, p, d)) - \mathbb{E}(\tau(G(n, p, d)))|^2 \geq \frac{1}{4}(\mathbb{E}(\tau(G(n, p, d))))^2\right\} \\ &\leq \frac{\mathbb{E}\left(|\tau(G(n, p, d)) - \mathbb{E}(\tau(G(n, p, d)))|^2\right)}{\frac{1}{4}\left(\mathbb{E}(\tau(G(n, p, d)))\right)^2} \\ &\leq \frac{n^3 + \frac{3n^4}{d}}{\frac{1}{4d}\binom{n}{3}^2 C_p^2} \leq \frac{dn^3 + 3n^4}{\frac{1}{4}\binom{n}{3}^2 C_p^2} \leq 200 \frac{dn^3 + 3n^4}{n^6 C_p^2}. \end{aligned} \quad (34)$$

Note that the equality in the second line follows from the fact that $\mathbb{E}(\tau(G(n, p, d))) \geq 0$. Thereafter, the third line invokes Markov's inequality. Finally, the fourth line utilizes (33) as well as a lower bound on the binomial coefficient $\binom{n}{3}^2$.

By a nearly identical argument, we compute an upper bound on $\mathbb{P}\{\tau(G(n, p)) \geq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\}$:

$$\begin{aligned} & \mathbb{P}\left\{\tau(G(n, p)) \geq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} \\ &= \mathbb{P}\left\{\tau(G(n, p))^2 \geq \frac{1}{4}\mathbb{E}(\tau(G(n, p, d)))^2\right\} \quad \mathbb{E}(\tau(G(n, p, d))) \geq 0 \\ &\leq \frac{\mathbb{E}(\tau(G(n, p))^2)}{\frac{1}{4}\left(\mathbb{E}(\tau(G(n, p, d)))\right)^2} \quad \text{Markov's inequality} \\ &\leq \frac{n^3 + \frac{3n^4}{d}}{\frac{1}{4d}\binom{n}{3}^2 C_p^2} \leq \frac{dn^3 + 3n^4}{\frac{1}{4}\binom{n}{3}^2 C_p^2} \leq 200 \frac{dn^3 + 3n^4}{n^6 C_p^2}. \quad \text{lower bound on } \binom{n}{3}^2 \end{aligned}$$

Considering the complement of the above event, we have that

$$\mathbb{P}\left\{\tau(G(n, p)) \leq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} = 1 - \mathbb{P}\left\{\tau(G(n, p)) \geq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\}$$

$$\geq 1 - 200 \frac{dn^3 + 3n^4}{n^6 C_p^2}. \quad (35)$$

Finally, we give a lower bound on the total variation distance between the distribution of the signed-triangle statistic under the Erdős–Rényi standard model $G(n, p)$ and its distribution under the geometric random graph $G(n, p, d)$. Specifically, we have

$$\begin{aligned} & \text{TV}(\tau(G(n, p)), \tau(G(n, p, d))) \\ &= \sup_{B \in \mathcal{B}(\mathbb{R})} |\mathbb{P}(\{\omega : \tau(G(n, p)) \in B\}) - \mathbb{P}(\{\omega : \tau(G(n, p, d)) \in B\})| \\ &\geq \left| \mathbb{P}\left\{\tau(G(n, p)) \leq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} - \mathbb{P}\left\{\tau(G(n, p, d)) \leq \frac{1}{2}\mathbb{E}(\tau(G(n, p, d)))\right\} \right| \\ &\geq \left| \left(1 - 200 \frac{dn^3 + 3n^4}{n^6 C_p^2}\right) - \left(200 \frac{dn^3 + 3n^4}{n^6 C_p^2}\right) \right| \\ &= 1 - \left(\frac{400}{C_p^2}\right) \frac{d}{n^3} - \left(\frac{400}{C_p^2}\right) \frac{3}{n^2} \\ &\geq 1 - C \frac{d}{n^3} - C \frac{1}{n^2}, \end{aligned}$$

where $C := \frac{400}{C_p^2} > 0$ is a constant for each $p \in (0, 1)$.

Therefore, we have that for fixed $p \in (0, 1)$, then

$$\text{TV}(\tau(G(n, p)), \tau(G(n, p, d))) \geq 1 - C \frac{d}{n^3} - C \frac{1}{n^2} \longrightarrow 1 \quad \text{as} \quad \frac{d}{n^3} \longrightarrow 0.$$

And so as long as d , the dimension of the random vectors $X_i \in \mathbb{S}^{d-1}$ which determine G under H_1 , is less than n^3 , where n the number of vertices of G , then the signed triangle statistic is asymptotically powerful. \square

4 Conclusion

Our report has considered the hypothesis testing problem of determining whether or not a random graph G on n vertices has some high-dimensional geometric structure. To do so, we studied two particular distributions for G : the Erdős–Rényi model $G(n, p)$ in which the presence of edges is determined independently and an alternative $G(n, p, d)$ in which the presence of edges depends on independent random vectors $X_1, \dots, X_n \in \mathbb{S}^{d-1}$. In order to test these hypotheses, we considered the distribution of two distinct test statistics T , the triangle statistic, and τ , the signed triangle statistic, under $G \sim G(n, p)$ and $G \sim G(n, p, d)$. First, we found a lower bound on the expected value of the triangle statistic T under $G \sim G(n, p, d)$. Using this result, we then showed that the signed triangle statistic τ is asymptotically powerful for the hypothesis testing problem so long as $d \ll n^3$.

This result proven by Bubeck and colleagues is significant insofar as it marks an improvement from the triangle statistic, which is only asymptotically powerful for $d \ll n^2$. More important than the specific problem itself, we developed rigorous, measure-theoretic proofs for the two key results on which we focused. From this project, I have gained a deeper understanding of essential probabilistic concepts, such as the conditional expectation, product spaces, and measures for the distance between probability distributions.

References

- [1] S. BUBECK, J. DING, R. ELDAN, AND M. Z. RÁ CZ, *Testing for high-dimensional geometry in random graphs*, Random Structures & Algorithms, 49 (2016), pp. 503–532.
- [2] G. B. FOLLAND, *Real analysis: modern techniques and their applications*, vol. 40, John Wiley & Sons, 1999.
- [3] K. GADE, *A non-singular horizontal position representation*, The journal of navigation, 63 (2010), pp. 395–417.
- [4] H. B. LAWSON JR, *Codimension-one foliations of spheres*, Annals of Mathematics, (1971), pp. 494–503.
- [5] Y. LEE AND W. C. KIM, *Concise formulas for the surface area of the intersection of two hyperspherical caps*, KAIST Technical Report, (2014).
- [6] I. NAVERNI OUK, I. SUTSKEVER, AND H. HATAMI, *Csc2414: Metric embeddings, lecture 5: Dimension reduction*, 2006. <http://www.cs.toronto.edu/~avner/teaching/S6-2414/LN5.pdf>.
- [7] D. POLLARD, *A User’s Guide to Measure Theoretic Probability*, Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press, 2001.
- [8] L. S, *Concise formulas for the area and volume of a hyperspherical cap*, Asian Journal of Mathematics Statistics, 4 (2011).
- [9] S. SODIN, *Tail-sensitive gaussian asymptotics for marginals of concentrated measures in high dimension*, arXiv preprint math/0501382, (2005).