Reappraising the distribution of the number of edge crossings of graphs on a sphere

Lluís Alemany-Puig[*,1], **Mercè Mora**[2] **& Ramon Ferrer-i-Cancho**[1]

[1]Complexity & Quantitative Linguistics Lab,
Departament de Ciències de la Computació,
Laboratory for Relational Algorithmics, Complexity and Learning (LARCA),
Universitat Politècnica de Catalunya,
Barcelona, Catalonia, Spain.
[2] Group on Discrete, Combinatorial and Computational Geometry,
Departament de Matemàtiques,
Universitat Politècnica de Catalunya,
Barcelona, Catalonia, Spain.
* Author to whom any correspondence should be addressed.

E-mail: `lalemany@cs.upc.edu`, `merce.mora@upc.edu`,
`rferrericancho@cs.upc.edu`

Abstract. Many real transportation and mobility networks have their vertices placed on the surface of the Earth. In such embeddings, the edges laid on that surface may cross. In his pioneering research, Moon analyzed the distribution of the number of crossings on complete graphs and complete bipartite graphs whose vertices are located uniformly at random on the surface of a sphere assuming that vertex placements are independent from each other. Here we revise his derivation of that variance in the light of recent theoretical developments on the variance of crossings and computer simulations. We show that Moon's formulae are inaccurate in predicting the true variance and provide exact formulae.

Keywords: crossings in spherical arrangements, variance of crossings.

Contents

1. Introduction

The shape of our planet can be approximated by a sphere with a radius of about $3.9 \cdot 10^3$ miles. Many real transportation and mobility networks have vertices located on the surface of that sphere. These are examples of spatial networks, networks whose vertices are embedded in a space [3]. In many transportation and mobility networks, the surface of the sphere is simplified as a projection on a plane [3] while in some other cases, e.g., air transportation networks [13, 10, 11], such an approximation is often not possible due to the long distances involved.

    When vertices are embedded in some space, edges may cross. While crossings are exceptional in many spatial networks to the point of being neglectable [3], crossings can also be scarce but not neglectable in one-dimensional layouts of certain networks: syntactic dependency and RNA secondary structures, where vertices are arranged linearly (distributed along a line) [8, 5]. The former are networks whose vertices represent words of a sentence and the edges represent syntactic dependencies between

them. These have become the *de facto* standard to represent the syntactic structure of sentences in computational linguistics [12] and the fuel of many quantitative studies [14, 19]. In RNA secondary structures, vertices are nucleotides A, G, U, and C, and edges are Watson-Crick (A-U, G-C) and (U-G) base pairs [5]. In these one-dimensional networks, two edges cross whenever the endpoints' positions are interleaved in the sequence.

Statistical properties of $\mathbf{C}$, the number of edge crossings of a graph G, have been studied in generic embeddings, denoted as $*$, that meet three mathematical conditions [2]: (1) only independent edges can cross (edges that do not share vertices), (2) two independent edges can cross in at most one point, and (3) if several edges of the graph, say e edges, cross at exactly the same point then the amount of crossings equals $\binom{e}{2} = e(e-1)/2$. In our view, generic embeddings are two-fold: a space and a statistical distribution of the vertices in such space. In [16], the space is the surface of a sphere while the distribution of the vertices on that surface is uniformly random. Compact formulae for the expectation and the variance have been obtained [2]. Here we apply such a framework to revise the problem of calculating the distribution of $\mathbf{C}$ in arrangements of vertices on the surface of a sphere. We use $\mathbb{E}_* [\mathbf{C}] = \mathbb{E}_* [\mathbf{C}(G)]$ to denote the expectation of the number of crossings $\mathbf{C}$, and $\mathbb{V}_* [\mathbf{C}] = \mathbb{V}_* [\mathbf{C}(G)]$ to denote the variance of $\mathbf{C}$ in a generic layout $*$.

In his pioneering research [16], J. W. Moon analyzed the properties of the distribution of $\mathbf{C}$ in uniformly random spherical arrangements (rsa), where vertices are arranged on the surface of a sphere uniformly at random and independently from each other, and edges become geodesics on the sphere's surface. Specifically, Moon studied $\mathbb{E}_{rsa} [\mathbf{C}]$ and $\mathbb{V}_{rsa} [\mathbf{C}]$, the expectation and variance of $\mathbf{C}$ in the random spherical layout, for two kinds of graphs: complete graphs of n vertices, $\mathcal{K}_n$, and complete bipartite graphs, $\mathcal{K}_{n_1,n_2}$, with n_1 vertices in one partition and n_2 vertices in the other. His derivations of $\mathbb{E}_{rsa} [\mathbf{C}]$ are straightforward. Borrowing the notation in [2], Moon obtained that the expectation of $\mathbf{C}$ is

$$\mathbb{E}_{rsa} [\mathbf{C}] = q\delta_{rsa},$$

where q is the number of pairs of independent edges [18] and δ_{rsa} is the probability that two independent edges cross. Indeed, q is a handle for the size of the set Q, consisting of the pairs of independent edges of a graph G [18, 2]. Thanks to

$$\delta_{rsa} = 1/8, \tag{1}$$

$$|Q(\mathcal{K}_n)| = \frac{1}{2}\binom{n}{2}\binom{n-2}{2}$$

and

$$|Q(\mathcal{K}_{n_1,n_2})| = 2\binom{n_1}{2}\binom{n_2}{2},$$

Moon obtained

$$
\begin{aligned}
\mathbb{E}_{rsa}\left[\mathbf{C}(\mathcal{K}_n)\right] &= \frac{1}{16}\binom{n}{2}\binom{n-2}{2}, \\
\mathbb{E}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right] &= \frac{1}{4}\binom{n_1}{2}\binom{n_2}{2}.
\end{aligned}
$$

Moon also derived formulae for the variance of $\mathbf{C}$ for these two kinds of graphs, i.e.

$$
\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_n)\right] = \frac{1}{2}\binom{n}{2}\binom{n-2}{2}\left[1\cdot\frac{7}{64} + 2\binom{n-4}{2}\cdot\frac{\pi^2-8}{64\pi^2} + 4(n-4)\cdot\frac{\pi^2-8}{64\pi^2} + 2\cdot\frac{-1}{64}\right],
$$

which simplifies to

$$
\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_n)\right] = 3\binom{n}{4}\left[\frac{5}{64} + \frac{\pi^2-8}{64\pi^2}(n-4)(n-1)\right] \tag{2}
$$

and also

$$
\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right] = \frac{1}{16\pi^2}\binom{n_1}{2}\binom{n_2}{2}[(n_1-1)(n_2-1)(\pi^2-8) + 2(\pi^2+4)], \tag{3}
$$

where the superscript (M) is used to distinguish Moon's work from our own derivations. Here we revise equations 2 and 3 in light of computer simulations and a recently introduced theoretical framework to investigate $\mathbb{V}_*\left[\mathbf{C}\right]$ [2].

This article is organized as follows. Section 2 summarizes the general mathematical framework for the calculation of $\mathbb{V}_*\left[\mathbf{C}\right]$ [2], and section 3 adapts it to the case of random spherical arrangements. In section 4, we review Moon's calculations, and compare them with our own with the help of numerical estimates of $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$ in complete and complete bipartite graphs in sections 4.1 and 4.2 respectively. These numerical estimates confirm the correctness of our derivations and show that Moon's (2)-(3) are inaccurate. Section 5 discusses our findings and attempts to shed light on the origins of the inaccuracy of $\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_n)\right]$ and $\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$. Section 6 details all the numerical methods involved in the numerical calculation of $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$. This section is placed after the discussion to make the presentation of the main arguments more streamlined.

2. The variance of C in generic layouts

In [2], $\mathbf{C}$ was defined as a summation of pairwise crossings between independent edges, i.e.

$$
\mathbf{C} = \sum_{\{e_1,e_2\}\in Q}\alpha(e_1,e_2), \tag{4}
$$

where $\alpha(e_1,e_2)$ is an indicator random variable that equals 1 whenever the independent edges e_1 and e_2 cross in the given layout. This definition was used to derive the expectation of $\mathbf{C}$ as

$$
\mathbb{E}_*\left[\mathbf{C}\right] = q\delta_*, \tag{5}
$$

Table 1: The classification of the types of products $\alpha(e_1, e_2)\alpha(e_3, e_4)$ abstracting from the order of the elements of the pair $(\{e_1, e_2\}, \{e_3, e_4\}) \in Q \times Q$. The type ω is the result of concatenating of τ and ϕ, $|v|$ is the number of different vertices of the type, $\tau = |\{e_1, e_2\} \cap \{e_3, e_4\}|$ and $\phi = |(e_1 \cup e_2) \cap (e_3 \cup e_4)|$. The form of every type of product is illustrated using $s, t, ..., y, z$ to indicate distinct vertices (st indicates the edge formed by vertices s and t; the same for uv, wx,...). Types fully acknowledged by Moon are marked in bold.

ω	$(\{e_1, e_2\}, \{e_3, e_4\})$	$\|v\|$	τ	ϕ
00	$(\{st, uv\}, \{wx, yz\})$	8	0	0
01	$(\{st, uv\}, \{sw, xy\})$	7	0	1
021	$(\{st, uv\}, \{su, wx\})$	6	0	2
022	$(\{st, uv\}, \{sw, ux\})$	6	0	2
03	$(\{st, uv\}, \{su, vw\})$	5	0	3
04	$(\{st, uv\}, \{su, tv\})$	4	0	4
12	$(\{st, uv\}, \{st, wx\})$	6	1	2
13	$(\{st, uv\}, \{st, uw\})$	5	1	3
24	$(\{st, uv\}, \{st, uv\})$	4	2	4

where

$$\delta_* = \mathbb{E}_* \left[\alpha(e_1, e_2) \right] \tag{6}$$

for two independent edges e_1 and e_2 embedded in the layout. Hereafter, an edge is a set of two vertices, denoted as $e = \{s, t\} = st$. Since α is an indicator random variable, δ_* is the probability that two independent edges cross in the given layout $*$. For example, in Moon's random spherical arrangement $\delta_{rsa} = 1/8$ [16]. Therefore, using (5) we obtain

$$\mathbb{E}_{rsa}[\mathbf{C}] = \frac{1}{8}q. \tag{7}$$

Moreover, in uniformly random linear arrangements (rla), where the vertices of a graph are placed along a linear sequence uniformly at random, $\delta_{rla} = 1/3$ [7], and then

$$\mathbb{E}_{rla}[\mathbf{C}] = \frac{1}{3}q.$$

The same definition of $\mathbf{C}$ was used again in [2] to study the variance of $\mathbf{C}$ in a general layout, similarly to the way Moon did for the particular case of random spherical arrangements [16]. In [2], the variance of $\mathbf{C}$ was expressed compactly as a summation over products between graph-dependent terms, the f_ω's, and layout-dependent terms, the $\mathbb{E}_* [\gamma_\omega]$'s, i.e.

$$\mathbb{V}_* [\mathbf{C}] = \sum_{\omega \in \Omega} f_\omega \mathbb{E}_* [\gamma_\omega]. \tag{8}$$

Formally, the type of a product is obtained applying a function $\mathcal{T}$ on a pair $(\{e_1, e_2\}, \{e_3, e_4\}) \in Q \times Q$ and then [2],

$$f_\omega = \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q \\ \mathcal{T}(q_1, q_2) = \omega}} 1.$$

Table 2: a_ω and F_ω as function of type ω. $\mathcal{L}_n$ and $\mathcal{C}_n$ denote linear trees (or path graphs) and cycle graphs of n vertices, respectively, and the operator $\oplus$ indicates the disjoint union of graphs.

ω	a_ω	F_ω
00	6	$\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$
01	4	$\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$
021	2	$\mathcal{L}_4 \oplus \mathcal{L}_2$
022	4	$\mathcal{L}_3 \oplus \mathcal{L}_3$
03	2	$\mathcal{L}_5$
04	2	$\mathcal{C}_4$
12	6	$\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$
13	2	$\mathcal{L}_3 \oplus \mathcal{L}_2$
24	1	$\mathcal{L}_2 \oplus \mathcal{L}_2$

The crux to understand (8) are the layout-dependent terms, $\mathbb{E}_*[\gamma_\omega]$, actually a shorthand for

$$\mathbb{E}_*[\gamma_\omega] = \mathbb{E}_*[\alpha(e_1, e_2)\alpha(e_3, e_4) \mid \mathcal{T}(\{e_1, e_2\}, \{e_3, e_4\}) = \omega] - \delta_*^2, \tag{9}$$

where $\omega \in \Omega$ is the type of the product $\alpha(e_1, e_2)\alpha(e_3, e_4)$ for $(\{e_1, e_2\}, \{e_3, e_4\}) \in Q \times Q$. The type of product is determined by the vertices forming the edges of $\{e_1, e_2\}, \{e_3, e_4\}$ as explained in detail in table 1. The set of all distinct types of products is

$$\Omega = \{00, 01, 021, 022, 03, 04, 12, 13, 24\} \tag{10}$$

following the encoding of each type in table 1.

The amount of products $\alpha(e_1, e_2)\alpha(e_3, e_4)$ of type ω in the given graph G satisfies [2]

$$f_\omega = a_\omega n_G(F_\omega),$$

where a_ω is a positive integer constant that depends only on ω, and $n_G(F_\omega)$ is the number of subgraphs of G isomorphic to the graph F_ω defined by the edges involved in the product $\alpha(e_1, e_2)\alpha(e_3, e_4)$ of type ω. Figure 1 depicts all these graphs. Table 2 shows the values of the a_ω and the formal definition of each F_ω.

For the sake of brevity, we use the shorthand

$$p_{*, \omega} = \mathbb{E}_*[\alpha(e_1, e_2)\alpha(e_3, e_4) \mid \mathcal{T}(\{e_1, e_2\}, \{e_3, e_4\}) = \omega]. \tag{11}$$

Since α is an indicator random variable, $p_{*, \omega}$ is the probability that both pairs of independent edges $\{e_1, e_2\}, \{e_3, e_4\} \in Q$ cross in a generic embedding *.

Therefore, as it was concluded in [2], in order to calculate the variance of $\mathbf{C}$ of a graph G in a given layout * one only needs to know the values of the f_ω's in G (which are independent of the layout) and the values of the $\mathbb{E}_*[\gamma_\omega]$ in the given layout (which are independent of the graph). The values of the f_ω's in complete and complete bipartite graphs, shown in table 3, have allowed to derive expressions for the variance of $\mathbf{C}$ that
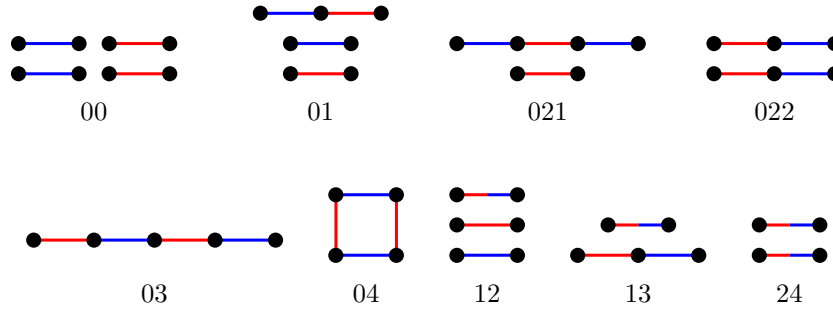
Figure 1:   The subgraphs corresponding to each type of product of the form $\alpha(e_1, e_2)\alpha(e_3, e_4)$.   The type $\omega \in \Omega$ (10) is indicated below them.   In the product $\alpha(e_1, e_2)\alpha(e_3, e_4)$, e_1 and e_2 share the same color and e_3 and e_4 share another color. Equally-colored edges are to cross when calculating $\alpha(e_1, e_2)\alpha(e_3, e_4)$ and, by definition, belong to the same element of Q.   Bi-colored edges (as in types 12, 13 and 24) belong to both elements of Q.

are valid for a generic embedding * [2]. The substitution of the these values in (8) yields

$$\mathbb{V}_* \left[\mathbf{C}(\mathcal{K}_n) \right] = 3 \binom{n}{4} ((n - 4)(n - 5)(\mathbb{E}_* \left[\gamma_{12} \right] + 4(\mathbb{E}_* \left[\gamma_{021} \right] + \mathbb{E}_* \left[\gamma_{022} \right]))$$
$$+ 4(n - 4)(\mathbb{E}_* \left[\gamma_{13} \right] + 2\mathbb{E}_* \left[\gamma_{03} \right])$$
$$+ 2\mathbb{E}_* \left[\gamma_{04} \right] + \mathbb{E}_* \left[\gamma_{24} \right]) \tag{12}$$

and, likewise,

$$\mathbb{V}_* \left[\mathbf{C}(\mathcal{K}_{n_1, n_2}) \right] = 2(\mathbb{E}_* \left[\gamma_{24} \right] + \mathbb{E}_* \left[\gamma_{04} \right]) \binom{n_1}{2} \binom{n_2}{2}$$
$$+ 12(\mathbb{E}_* \left[\gamma_{03} \right] + \mathbb{E}_* \left[\gamma_{13} \right]) \left[\binom{n_1}{3} \binom{n_2}{2} + \binom{n_1}{2} \binom{n_2}{3} \right]$$
$$+ 36(\mathbb{E}_* \left[\gamma_{12} \right] + \mathbb{E}_* \left[\gamma_{022} \right] + 2\mathbb{E}_* \left[\gamma_{021} \right]) \binom{n_1}{3} \binom{n_2}{3}$$
$$+ 24\mathbb{E}_* \left[\gamma_{022} \right] \left[\binom{n_1}{4} \binom{n_2}{2} + \binom{n_1}{2} \binom{n_2}{4} \right]. \tag{13}$$

A step further consists of instantiating the equations above replacing * by a uniformly random linear arrangements (*rla*).   After calculating the values of the $\mathbb{E}_{rla} \left[\gamma_\omega \right]$ and substituting them into (12), one obtains [2]

$$\mathbb{V}_{rla} \left[\mathbf{C}(\mathcal{K}_n) \right] = 0$$

as expected. Interestingly, the same approach on (13) produces

$$\mathbb{V}_{rla} \left[\mathbf{C}(\mathcal{K}_{n_1, n_2}) \right] = \frac{1}{90} \binom{n_1}{2} \binom{n_2}{2} ((n_1 + n_2)^2 + n_1 + n_2).$$

Next section shows how equivalent results can be obtained when replacing * by a random spherical arrangement, which turns out to be a more complex case.

Table 3: Summary of the values of the f_ω's in complete graphs, $\mathcal{K}_n$, and in complete bipartite graphs, $\mathcal{K}_{n_1,n_2}$. Values extracted from [2, Table 5].

ω	$f_\omega(\mathcal{K}_n)$	$f_\omega(\mathcal{K}_{n_1,n_2})$
00	$630\binom{n}{8}$	$144\binom{n_1}{4}\binom{n_2}{4}$
01	$1260\binom{n}{7}$	$144\binom{n_1}{4}\binom{n_2}{3} + 144\binom{n_1}{3}\binom{n_2}{4}$
021	$360\binom{n}{6}$	$72\binom{n_1}{3}\binom{n_2}{3}$
022	$360\binom{n}{6}$	$24\binom{n_1}{2}\binom{n_2}{4} + 24\binom{n_1}{4}\binom{n_2}{2} + 36\binom{n_1}{3}\binom{n_2}{3}$
03	$120\binom{n}{5}$	$12\binom{n_1}{3}\binom{n_2}{2} + 12\binom{n_1}{2}\binom{n_2}{3}$
04	$6\binom{n}{4}$	$2\binom{n_1}{2}\binom{n_2}{2}$
12	$90\binom{n}{6}$	$36\binom{n_1}{3}\binom{n_2}{3}$
13	$60\binom{n}{5}$	$12\binom{n_1}{3}\binom{n_2}{2} + 12\binom{n_1}{2}\binom{n_2}{3}$
24	$3\binom{n}{4}$	$2\binom{n_1}{2}\binom{n_2}{2}$

3. The variance of C in spherical random arrangements

Here we aim to calculate the values $\mathbb{E}_{rsa}[\gamma_\omega]$ so as to establish the foundations to derive an arithmetic expression for the variance of $\mathbf{C}$ in uniformly random spherical arrangements of complete and complete bipartite graphs. Recall that, in this layout, vertices are distributed on the surface of a sphere uniformly at random, and edges become geodesics on that surface, i.e., great arcs (see section 6.2 for a detailed account of what we consider valid arc-arc intersections). We use the $\mathbb{E}_{rsa}[\gamma_\omega]$'s to instantiate equations 12 and 13 and then obtain arithmetic expressions for $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$ and $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ (section 3.4). Each $\mathbb{E}_{rsa}[\gamma_\omega]$ is calculated via (9): once $p_{rsa,\omega}$ has been derived from (11), δ_{rsa}^2 (6) is subtracted.

We first calculate the $p_{rsa,\omega}$ for the simplest cases. Thanks to [2] we have that $p_{*,24} = \delta_*$ and $p_{*,00} = p_{*,01} = \delta_*^2$. Since $\delta_{rsa} = 1/8$ [16],

$$p_{rsa,24} = 1/8$$

$$p_{rsa,00} = p_{rsa,01} = 1/64$$

and, following (9), we obtain

$$\mathbb{E}_{rsa}[\gamma_{24}] = 7/64 \tag{14}$$

$$\mathbb{E}_{rsa}[\gamma_{00}] = \mathbb{E}_{rsa}[\gamma_{01}] = 0. \tag{15}$$

Furthermore [2], $p_{rsa,04} = 0$ and then

$$\mathbb{E}_{rsa}[\gamma_{04}] = -1/64. \tag{16}$$

Notice that, given a pair of edges $\{st, uv\} \in Q$, we can form a type 04 combining $\{st, uv\}$ with $\{sv, tu\}$ or $\{su, tv\}$, which gives two configurations of type 04, $(\{st, uv\}, \{sv, tu\})$ and $(\{st, uv\}, \{su, tv\})$. For $p_{rsa,04} > 0$, we need both indicator variables α to be 1 for each pair of edges. However, if $\alpha(st, uv) = 1$ then it is impossible that $\alpha(su, tv) = 1$ or that $\alpha(sv, tu) = 1$.

So far we have calculated $p_{rsa,\omega}$ and $\mathbb{E}_{rsa}[\gamma_\omega]$ for $\omega \in \{00, 01, 04, 24\}$ using arguments that can be applied to most layouts. Now we move on to $p_{rsa,\omega}$ for
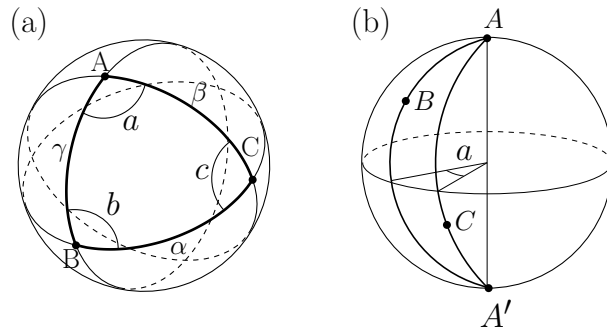
Figure 2: (a) A spherical triangle with vertices A, B and C; sides α, β and γ, and angles a, b and c. (b) The lune $w(A; B, C)$ of angle a.

$\omega \in \{021, 022, 03, 12, 13\}$ using an *ad hoc* approach for the spherical layout that is based on spherical trigonometry and integration over surfaces. Such surfaces are delimited by the edges that make up type $\omega \in \Omega$ (table 1).

We proceed gradually towards such aim. First, we introduce the relevant background from spherical trigonometry (section 3.1). Second, we propose a derivation for $\delta_{rsa} = 1/8$ that is more detailed than that of Moon [16] (section 3.2). Finally, we proceed with the remainder of types of products namely 12, 021, 13, 03, 022 (section 3.3).

3.1. Spherical trigonometry

We first recall some definitions and properties of spherical trigonometry [9]. Henceforth, we assume a unit-radius sphere. Let A, B and C be three points on a sphere of center O, such that A, B and C do not lie all together on a plane containing O.

Points A, B, C define a *spherical triangle*, denoted by $tr(A, B, C)$, whose vertices are A, B and C, and whose sides are the geodesics joining A with B, B with C, and C with A. Let α, β and γ denote the length of the sides BC, AC and AB, respectively (figure 2(a)).

For every point P, let P' denote its antipodal vertex, that is, P and P' are diametrically opposite. Thus, the segment $\overline{PP'}$ goes through the center O. The semicircle on the sphere containing A, A' and B, and the semicircle on the sphere containing A, A' and C delimit two disjoint regions on the sphere. The *lune* $w(A; B, C)$ is the one with smaller area. The angle of $w(A; B, C)$ is the angle in $(0, \pi)$ defined by the planes containing the semicircles delimiting the lune (figure 2(b)). The angles at vertices A, B and C of a spherical triangle $tr(A, B, C)$ are the angles a, b and c in $(0, \pi)$ of the lunes $w(A; B, C)$, $w(B; A, C)$ and $w(C; A, B)$, respectively (figure 2(a)).

With this notation, the following formula relates the lengths α, β of two sides with the angles b and c of the spherical triangle $tr(A, B, C)$

$$\cot(\beta)\sin(\alpha) = \cos(\alpha)\cos(c) + \sin(c)\cot(b). \tag{17}$$

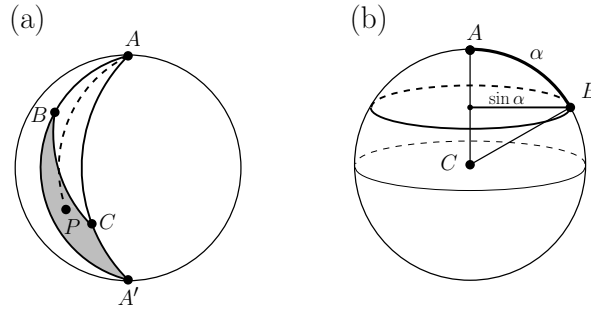Using this equation, from the length of two sides and the angle at the shared vertex of

Figure 3: (a) Given three points A, B and C, and a random point P, the probability that the geodesic AP crosses the side BC is proportional to the area of the spherical triangle $tr(A', B, C)$. (b) Given two random points A and B, the probability that the geodesic AB has length α is proportional to the length of the circle that goes through B and lying on the plane perpendicular to the radius CA; notice that this circle has radius $\sin \alpha$ if the radius of the sphere is 1.

a spherical triangle, the remaining angles can be calculated. Concretely,

$$\cot(b) = \frac{\cot(\beta) \sin(\alpha) - \cos(\alpha) \cos(c)}{\sin(c)}$$

and, analogously,

$$\cot(a) = \frac{\cot(\alpha) \sin(\beta) - \cos(\beta) \cos(c)}{\sin(c)}.$$

Since this relation is used often in our calculations, let us define a function g such that for any real numbers $x, y, z \in (0, \pi)$,

$$g(x, y, z) = \operatorname{arccot} \left(\frac{\cot(x) \sin(y) - \cos(y) \cos(z)}{\sin(z)} \right). \tag{18}$$

Let $S(R)$ denote the area of a region R of the sphere of radius 1. It is well-known that the area of a lune is $S(w(A; B, C)) = 2a$, where a is the angle of the lune, and the area of a spherical triangle is $S(tr(A, B, C)) = a + b + c - \pi$, where a, b and c are the angles at vertices A, B and C, respectively.

3.2. The probability that two edges cross

Let A, B and C be three fixed points on the sphere and let P be a random point. The geodesic AP crosses BC if and only if P lies in the spherical triangle $tr(A', B, C)$ (figure 3(a)). Hence, the probability of this occurring is the area of the spherical triangle $tr(A', B, C)$ divided by the area of the sphere's surface,

$$\Pr\left[AP \text{ and } BC \text{ cross}\right] = \frac{S(tr(A', B, C))}{4\pi} = \frac{S(w(A; B, C)) - S(tr(A, B, C))}{4\pi}$$

$$= \frac{2a - (a + b + c - \pi)}{4\pi} = \frac{a - b - c + \pi}{4\pi}. \tag{19}$$

Let α denote the length of the geodesic defined by two random points on the sphere. The density function of α is

$$f(\alpha) = \frac{1}{2}\sin(\alpha). \tag{20}$$

Indeed, we know that $\int_0^\pi f(\alpha) = 1$ and that $f(\alpha)$ must be proportional to the length of the circle obtained by intersecting the plane containing one of the points and perpendicular to the line that goes through the other point and the center of the sphere (figure 3(b)). Taking into account these facts, we conclude that $f(\alpha)$ satisfies (20).

Let s_1 be a geodesic of length α. The probability that two random points lie on the different hemispheres determined by s_1 is $1/2$, and the probability that the geodesic determined by two random points on different hemispheres cross a given arc of length α is $\alpha/2\pi$. Hence, the probability that a geodesic s_2 defined by two random points crosses a geodesic s_1 of length α is

$$\mathbb{P}r\left[s_2 \text{ crosses } s_1 \middle| \text{ length of } s_1 \text{ equal to } \alpha\right] = \alpha/4\pi. \tag{21}$$

Hence, the probability that two random edges on the sphere cross is, as already derived by Moon (1),

$$\delta_{rsa} = \int_0^\pi \frac{\alpha}{4\pi} f(\alpha)d\alpha = \int_0^\pi \frac{\alpha}{8\pi}\sin(\alpha)d\alpha = 1/8. \tag{22}$$

3.3. The $p_{rsa,\omega}$'s and the $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$'s

Let $(\{e_1, e_2\}, \{e_3, e_4\}) \in Q \times Q$ be an element of type $\omega \in \Omega$ as described in table 1 (see also figure 1). Below we calculate $p_{rsa,\omega}$ for every $\omega \in \{12, 021, 13, 03, 022\}$. Table 4 summarizes the results on $p_{rsa,\omega}$ that are derived next analytically and confirms the accuracy of the results with the help of computer simulations. For a better understanding of the explanations below, we refer the reader to table 1 (where we describe the types of products following [2]), and figure 1 (that illustrates the graphs characterizing each type).

Case $\omega = 12$. Let $e = e_1 = e_3$. By (20) and (21), the probability that two random edges e_2 and e_4 cross the edge e of length α is

$$p_{rsa,\,12} = \int_0^\pi \left(\frac{\alpha}{4\pi}\right)^2 f(\alpha)\,d\alpha = \frac{1}{32\pi^2}\int_0^\pi \alpha^2 \sin(\alpha)\,d\alpha = \frac{\pi^2 - 4}{32\pi^2},$$

and then

$$\mathbb{E}_{rsa}\left[\gamma_{12}\right] = \frac{\pi^2 - 8}{64\pi^2}. \tag{23}$$

Case $\omega = 021$. Recall that $e_1 = st$, $e_2 = uv$, $e_3 = su$ and $e_4 = wx$, with s, t, u, v, w, x pairwise distinct. The relative position of s, t and u can be given by three independent parameters: the length $\alpha \in (0, \pi)$ of the geodesic st, the length $\beta \in (0, \pi)$ of the geodesic su and the angle $c \in (-\pi, \pi)$ of the lune $w(s; t, u)$ (figure 4 (a)).

On the one hand, given a random point v, the probability that the edge $e_2 = uv$ crosses the edge $e_1 = st$ whenever $c \in (0, \pi)$ can be derived using (19) for the triangle $tr(A, B, C)$ when $A = u$, $B = t$ and $C = s$. Besides, this probability is the same for the opposite angle $-c$. On the other hand, by (21), the probability that the edge e_4 crosses the edge e_3 of length β is $\beta/(4\pi)$. Therefore,

$$
\begin{aligned}
p_{rsa,\,021} &= 2 \iiint_0^\pi \frac{a - b - c + \pi}{4\pi} \cdot \frac{\beta}{4\pi} f(\alpha) f(\beta) \frac{1}{2\pi} \, d\alpha \, d\beta \, dc \\
&= 2 \iiint_0^\pi \frac{g(\alpha, \beta, c) - g(\beta, \alpha, c) - c + \pi}{4\pi} \cdot \frac{\beta}{32\pi^2} \sin(\alpha) \sin(\beta) \, d\alpha \, d\beta \, dc \\
&\approx 0.013
\end{aligned}
$$

and thus

$$
\mathbb{E}_{rsa}\left[\gamma_{021}\right] = p_{rsa,\,021} - \frac{1}{64} \approx -0.003. \tag{24}
$$

Case $\omega = 13$. Recall that $e_1 = e_3 = st$, $e_2 = uv$ and $e_4 = uw$, with s, t, u, v, w pairwise distinct. As in the preceding case, the relative position of s, t and u can be given by three independent parameters: the length $\alpha \in (0, \pi)$ of the geodesic st, the length $\beta \in (0, \pi)$ of the geodesic su and the angle $c \in (-\pi, \pi)$ of the lune $w(s; t, u)$ (figure 4 (a)). Moreover, given a random point v (resp. w), the probability that the edge $e_2 = uv$ (resp. $e_4 = uw$) crosses the edge $e_1 = st$ whenever $c \in (0, \pi)$ can be derived using (19) for the triangle $tr(A, B, C)$ when $A = u$, $B = t$ and $C = s$. Besides, the probability of crossing is the same for the opposite angle $-c$. Therefore,

$$
\begin{aligned}
p_{rsa,\,13} &= 2 \iiint_0^\pi \left(\frac{a - b - c + \pi}{4\pi}\right)^2 f(\alpha) f(\beta) \frac{1}{2\pi} \, d\alpha \, d\beta \, dc \\
&= 2 \iiint_0^\pi \left(\frac{g(\alpha, \beta, c) - g(\beta, \alpha, c) - c + \pi}{4\pi}\right)^2 \frac{1}{8\pi} \sin(\alpha) \sin(\beta) \, d\alpha \, d\beta \, dc \\
&\approx 0.031
\end{aligned}
$$

and thus

$$
\mathbb{E}_{rsa}\left[\gamma_{13}\right] = p_{rsa,\,13} - \frac{1}{64} \approx 0.016. \tag{25}
$$

Case $\omega = 03$. Assume that $e_1 = st$, $e_2 = uv$, $e_3 = su$ and $e_4 = vw$, with s, t, u, v, w pairwise distinct. Similarly as in the preceding case, the relative position of u, v and s can be given by three independent parameters: the length α of the geodesic uv, the length β of the geodesic us and the angle c of the lune $w(u; v, s)$ (figure 4(b)). Moreover, given a random point t, the probability that the edge $e_1 = st$ crosses $e_2 = uv$ whenever $c \in (0, \pi)$ can be derived using (19) for the triangle $tr(A, B, C)$ when $A = s$, $B = v$ and $C = u$. Analogously, given a random point w, the probability that the edge $e_4 = vw$
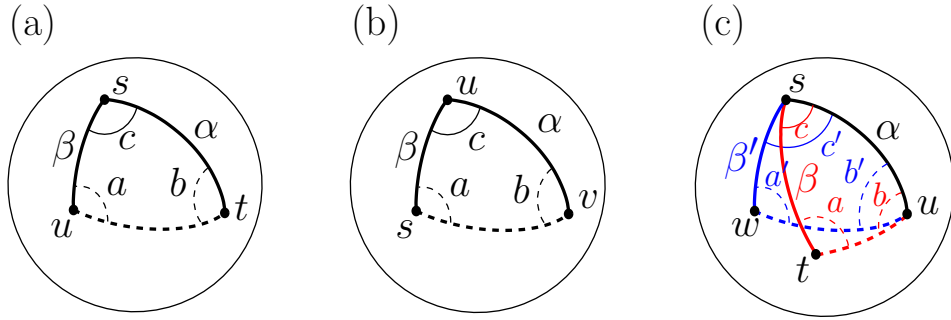
(a)   (b)   (c)

Figure 4: (a) Types 021 and 13: the relative position of points s, t and u is determined by parameters α, β and c. (b) Type 03: the relative position of points s, u and v is determined by parameters α, β and c. (c) Type 022: the relative position of points s, t, u and w is determined by parameters α, β, β', c and c'. The spherical triangles $tr(s,u,w)$ and $tr(s,u,t)$ are drawn in black-blue and in black-red, respectively.

crosses $e_3 = su$ can be derived using (19) for the triangle $tr(v,s,u)$. Therefore,

$$
\begin{aligned}
p_{rsa,\,03} &= 2\iiint_0^\pi \left(\frac{a-b-c+\pi}{4\pi}\right)\left(\frac{b-a-c+\pi}{4\pi}\right) f(\alpha)f(\beta)\frac{1}{2\pi}\, d\alpha\, d\beta\, dc \\
&= 2\iiint_0^\pi \left(\frac{g(\alpha,\beta,c)-g(\beta,\alpha,c)-c+\pi}{4\pi}\right)\left(\frac{g(\beta,\alpha,c)-g(\alpha,\beta,c)-c+\pi}{4\pi}\right) \\
&\quad \cdot \frac{1}{8\pi}\sin(\alpha)\sin(\beta)\, d\alpha\, d\beta \\
&\approx 0.01.
\end{aligned}
$$

Thus

$$
\mathbb{E}_{rsa}\left[\gamma_{03}\right] = p_{rsa,\,03} - \frac{1}{64} \approx -0.0052. \tag{26}
$$

Case $\omega = 022$. Recall that $e_1 = st$, $e_2 = uv$, $e_3 = sw$ and $e_4 = ux$, with s,t,u,v,w,x pairwise distinct. The relative position of points s, t, u and w can be given by 5 independent parameters, α, β, β', c, and c', where α, β and β' are the lengths of the geodesics su, st and sw, respectively; c is the angle of the lune $w(s;t,u)$; and c' is the angle of the lune $w(s;u,w)$, with $c,c' \in (-\pi,\pi)$ (figure 4(c)).

Let a, b denote the angles at vertices t and u, respectively, of the spherical triangle $tr(t,u,s)$ and let a', b' denote the angles at vertices w and u, respectively, of the spherical triangle $tr(w,u,s)$. For $c \in (0,\pi)$ and $c' \in (0,\pi)$, given two random points v and x, the probability that $e_2 = uv$ crosses $e_1 = st$ and the probability that $e_4 = ux$ crosses $e_3 = sw$ can be calculated using (19) for the triangles $tr(t,u,s)$ and $tr(w,u,s)$, respectively. Since the probability of crossing is the same for the opposites angles of c and c', we derive

Table 4: $p_{rsa,\,\omega}$ is the theoretical probability that $\alpha(e_1, e_2)\alpha(e_3, e_4) = 1$ for a product of type ω (11). $\hat{p}_{rsa,\,\omega}$ is the corresponding numerical estimate via computer simulation (section 6). The last column points to the equation where the exact value of $p_{rsa,\,\omega}$ is given. $p_{rsa,\,\omega}$ is provided in two forms: exact value (types 00, 01, 04, 12 and 24) or approximate value by solving numerically the integrals in section 3.3 (types 021, 022, 03 and 13).

ω	$p_{rsa,\,\omega}$	$\hat{p}_{rsa,\,\omega}$	Equation
00	$1/64 = 0.015625$	0.015625	15
01	$1/64$	0.015626	15
021	≈ 0.012665	0.012670	24
022	≈ 0.018566	0.018581	27
03	≈ 0.010401	0.010417	26
04	0	0	16
12	$(\pi^2 - 4)/32\pi^2 \approx 0.018585$	0.018581	23
13	≈ 0.031265	0.031251	25
24	$1/8$	0.125001	14

that

$$
p_{rsa,\,022} = 4\int \cdots \int_0^\pi \left(\frac{b-a-c+\pi}{4\pi}\right)\left(\frac{b'-a'-c'+\pi}{4\pi}\right)
$$
$$
f(\alpha)f(\beta)f(\beta')\left(\frac{1}{2\pi}\right)^2 d\alpha\, d\beta\, d\beta'\, dc\, dc'
$$
$$
= 4\int \cdots \int_0^\pi \left(\frac{g(\beta,\alpha,c)-g(\alpha,\beta,c)-c+\pi}{4\pi}\right)\left(\frac{g(\beta',\alpha,c')-g(\alpha,\beta',c')-c'+\pi}{4\pi}\right)
$$
$$
\frac{1}{32\pi^2}\sin(\alpha)\sin(\beta)\sin(\beta')\,d\alpha\, d\beta\, d\beta'\, dc\, dc'
$$
$$
\approx 0.019.
$$

Thus

$$
\mathbb{E}_{rsa}\left[\gamma_{022}\right] = p_{rsa,\,022} - \frac{1}{64} \approx 0.0029. \tag{27}
$$

The values of the integrals have been approximated as explained in section 6.

3.4. The variance of $\boldsymbol{C}$ in complete and complete bipartite graphs

By substituting the values of the $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$'s (table 4) into equations 12 and 13, we obtain, respectively,

$$
\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_n)\right] \approx 3\binom{n}{4}\left[(n-4)(0.0029(n-5)+0.021)+\frac{5}{64}\right] \tag{28}
$$

and

$$\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right] \approx \frac{3}{16}\binom{n_1}{2}\binom{n_2}{2} - 0.00068\binom{n_1}{3}\binom{n_2}{3}$$

$$+ 0.12\left[\binom{n_1}{3}\binom{n_2}{2} + \binom{n_1}{2}\binom{n_2}{3}\right]$$

$$+ 0.07\left[\binom{n_1}{4}\binom{n_2}{2} + \binom{n_1}{2}\binom{n_2}{4}\right]. \tag{29}$$

4. Revision of Moon's work

Here we revise Moon's pioneering work using the derivations above. We first interpret Moon's formula and try to identify Moon's calculations for the values of $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$ (table 4). Then, we compare Moon's results with ours in order to obtain a formalization of the deviation between Moon's formulae from the actual value of the variance (sections 4.1 and 4.2).

Notice that (2) follows the pattern of (8) with some differences. An obvious one is that q has been factored out. Therefore, it is convenient to rewrite the general formula of $\mathbb{V}_*\left[\mathbf{C}\right]$ (8) equivalently as

$$\mathbb{V}_*\left[\mathbf{C}\right] = q\sum_{\omega\in\Omega}\frac{f_\omega}{q}\mathbb{E}_*\left[\gamma_\omega\right].$$

The values of f_ω/q for a complete graph are summarized in table 5 (see Appendix A for a detailed derivation), and allow one to hypothesize that (2) is actually showing

$$\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_n)\right] = q\left[\frac{f_{24}}{q}\mathbb{E}_{rsa}\left[\gamma_{24}\right] + \frac{f_{12}}{q}\mathbb{E}_{rsa}\left[\gamma_{12}\right] + \frac{f_{13}}{q}\mathbb{E}_{rsa}\left[\gamma_{13}\right] + \frac{f_{04}}{q}\mathbb{E}_{rsa}\left[\gamma_{04}\right]\right] \tag{30}$$

with

$$\mathbb{E}_{rsa}^{(M)}\left[\gamma_{24}\right] = \frac{7}{64}$$

$$\mathbb{E}_{rsa}^{(M)}\left[\gamma_{04}\right] = -\frac{1}{64}$$

$$\mathbb{E}_{rsa}^{(M)}\left[\gamma_{12}\right] = \mathbb{E}_{rsa}^{(M)}\left[\gamma_{13}\right] = \frac{\pi^2 - 8}{64\pi^2}.$$

Moon suggested that the product types 00, 01, 021, 022 and 03 do not have any contribution in (8). On the one hand, the absence of types 00 and 01 could be explained by the fact that $\mathbb{E}_*\left[\gamma_{00}\right] = \mathbb{E}_*\left[\gamma_{01}\right] = 0$ [2]. However, as we have shown in section 3, the other types, i.e., 021, 022 and 03, satisfy $\mathbb{E}_{rsa}\left[\gamma_\omega\right] \neq 0$. On the other hand, the value of $\mathbb{E}_{rsa}^{(M)}\left[\gamma_{13}\right]$ differs from ours (table 4). In the next two sections, we show that Moon's values lead to inaccurate arithmetic expressions for $\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_n)\right]$ and $\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$.

In order to validate our hypothesis (30), we can instantiate (12) with the values $\mathbb{E}_{rsa}^{(M)}\left[\gamma_\omega\right]$. Doing so, we obtain $\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_n)\right]$ in (2). It is possible to repeat the same analysis for complete bipartite graphs (3) but it is more complex because Moon gave it in this compact form (without the intermediate results he showed for complete graphs). However, we can still observe that q was factored out in (3).

Table 5: Summary of the frequency of every product type in complete graphs. f_ω, the frequency of product type ω, and f_ω/q, the ratio between its frequency and q, the size of the set of independent edge pairs. The values of the f_ω for complete and complete bipartite graphs can be found in table 3.

ω	$f_\omega(\mathcal{K}_n)/\lvert Q(\mathcal{K}_n)\rvert$	$f_\omega(\mathcal{K}_{n_1,n_2})/\lvert Q(\mathcal{K}_{n_1,n_2})\rvert$
00	$3\binom{n-4}{4}$	$2\binom{n_1-2}{2}\binom{n_2-2}{2}$
01	$12\binom{n-4}{3}$	$4(n_1-2)(n_2-2)(n_1+n_2-6)$
021	$4(n-4)(n-5)=8\binom{n-4}{2}$	$4(n_1-2)(n_2-2)$
022	$4(n-4)(n-5)=8\binom{n-4}{2}$	$(n_2+n_1-5)(n_2+n_1-4)$
03	$8(n-4)$	$2(n_1+n_2-4)$
04	2	1
12	$(n-4)(n-5)=2\binom{n-4}{2}$	$2(n_1-2)(n_2-2)$
13	$4(n-4)$	$2(n_1+n_2-4)$
24	1	1

4.1. The variance of $\boldsymbol{C}$ in complete graphs

First, we approximate the deviation of $\mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_n)]$ (2) from the actual value of the variance (28)

$$\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)] - \mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_n)] \approx 3\binom{n}{4}\left(-7\cdot 10^{-5}(n-4)(n-122.48)\right) = O\left(n^6\right). \quad (31)$$

The accuracy of (28) in predicting $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$ is shown in figure 5. Our prediction fits extremely well the numerical estimate of $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$ obtained via computer simulations (section 6) while $\mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_n)]$ deviates from $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$. Upon solving the equation $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)] - \mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_n)] = 0$ for n, we can see that such deviation takes the form of underestimation for $n \lesssim 123$, and overestimation for $n \gtrsim 123$. This is why figure 5 shows only underestimation.

4.2. The variance of $\boldsymbol{C}$ in complete bipartite graphs

First, we approximate the deviation of $\mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ (3) from the actual value of the variance (29)

$$\begin{aligned}
\mathbb{V}_{rsa}&[\mathbf{C}(\mathcal{K}_{n_1,n_2})] - \mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]\\
&\approx \frac{1}{16}\binom{n_1}{2}\binom{n_2}{2}\left[3 - \frac{1}{\pi^2}((n_1-1)(n_2-1)(\pi^2-8)+2(\pi^2-4))\right]\\
&\quad - 0.00068\binom{n_1}{3}\binom{n_2}{3} + 0.12\left[\binom{n_1}{3}\binom{n_2}{2} + \binom{n_1}{2}\binom{n_2}{3}\right]\\
&\quad + 0.071\left[\binom{n_1}{4}\binom{n_2}{2} + \binom{n_1}{2}\binom{n_2}{4}\right]\\
&= O\left(n_1^2 n_2^2(n_1^2 + n_2^2 + n_1 + n_2 - n_1 n_2)\right). \quad (32)
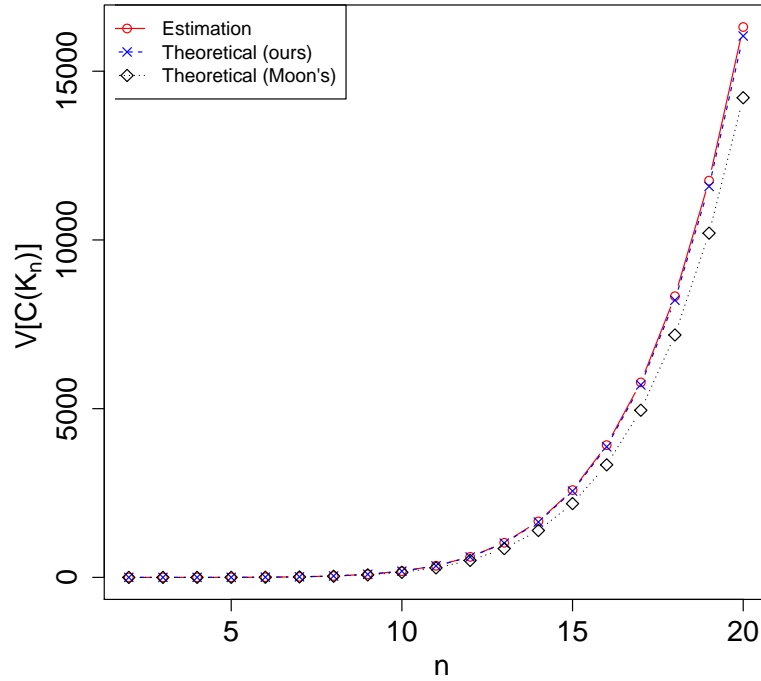\end{aligned}$$

Figure 5: Comparison of $\mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_n)]$ (2, black diamonds), our proposal of $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$ (28, blue crosses), and $\hat{\mathbb{V}}_{rsa}[\mathbf{C}(\mathcal{K}_n)]$, a numerical estimate obtained via computer simulations (red circles) using $N = 10^7$ random arrangements for each n.

The accuracy of (29) in predicting $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ is shown in figure 6. Again, our prediction fits extremely well the numerical estimate of $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ (section 6) while $\mathbb{V}_{rsa}^{(M)}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ deviates from $\mathbb{V}_{rsa}[\mathbf{C}(\mathcal{K}_{n_1,n_2})]$ as expected from (32).

The analysis of 32 shows that Moon's equation underestimates the true variance unless n_1 and n_2 are sufficiently large and not too far from each other; overestimation occurs otherwise (figure 7). Accordingly, figure 6 shows only understimation.

5. Discussion

Back in 1965, Moon studied the variance of edge crossings in spherical arrangements of graphs, $\mathbb{V}_{rsa}[\mathbf{C}]$. In this article, we have revised Moon's work in light of recent discoveries on $\mathbb{V}_*[\mathbf{C}]$, the variance in general layouts [2]. We have applied them to derive an arithmetic expression for $\mathbb{V}_{rsa}[\mathbf{C}]$, which consisted of calculating the values for $\mathbb{E}_{rsa}[\gamma_\omega]$, the expectation of the types of products summarized in table 4, and then deriving expressions for the variance in complete graphs (28) and in complete bipartite graphs (29).

While some of the $\mathbb{E}_{rsa}[\gamma_\omega]$ we have calculated are in agreement with Moon's results, i.e. $\mathbb{E}_{rsa}[\gamma_\omega]$ for $\omega \in \{00, 01, 04, 12, 24\}$, we have found that others are not, i.e.
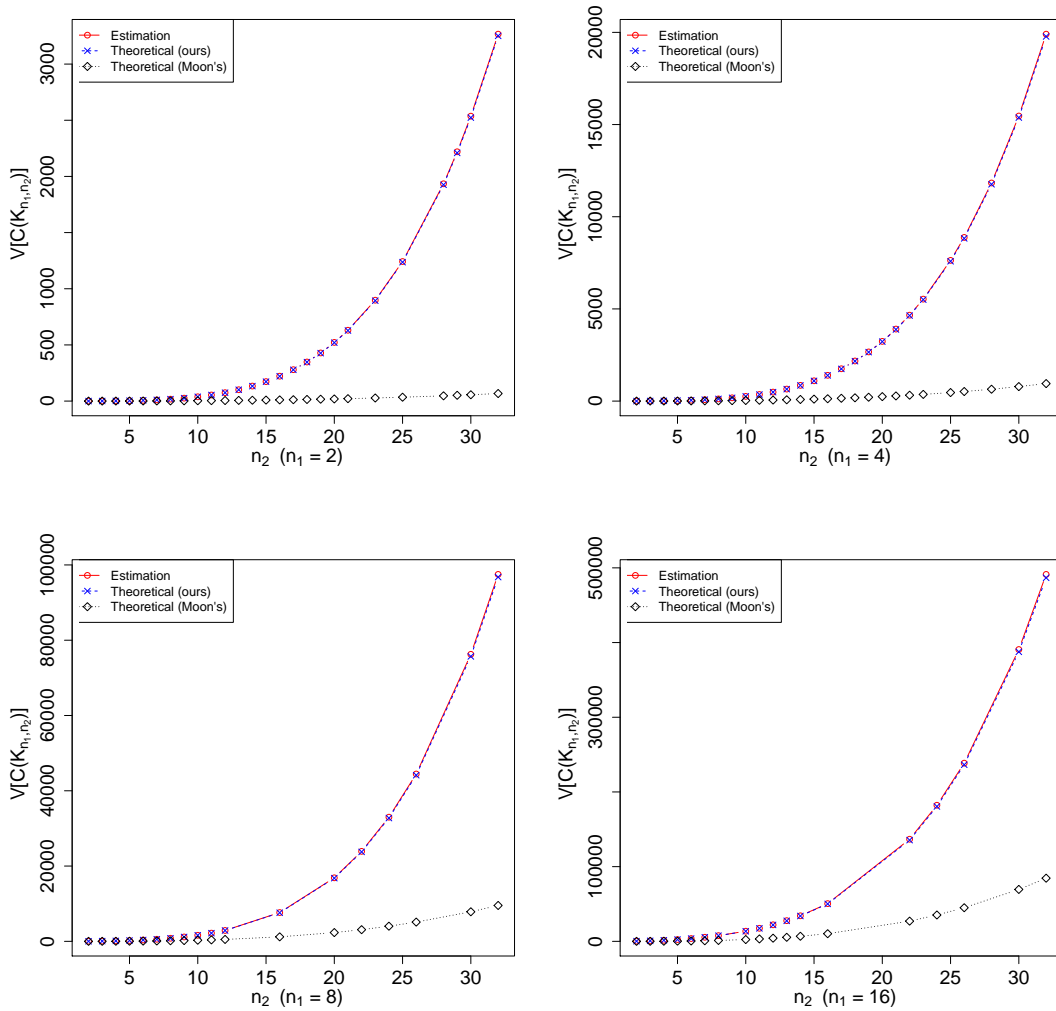
Figure 6: Comparison of $\mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$ (3), our proposal of $\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$ (29), and $\hat{\mathbb{V}}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$, a numerical estimate obtained via computer simulations using $N = 10^7$ random arrangements for each pair n_1, n_2. $\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$ denotes the variance of the number of edge crossings in random spherical arrangements of a complete bipartite graph as a function of n_1, with n_2 fixed. n_1 and n_2 are, respectively, the number of vertices of the first and second partitions.

$\omega \in \{021, 022, 03, 13\}$. Our belief that the calculation of $\mathbb{E}_{rsa}\left[\gamma_{13}\right]$ by Moon is inaccurate is supported by the analyses in section 4 and the values that were obtained in section 3 (summarized in table 4). Moreover, it appears that Moon considered some of these type's contribution to the variance to be null, i.e., $\mathbb{E}_{rsa}\left[\gamma_\omega\right] = 0$ for $\omega \in \{021, 022, 03\}$, supported by the claim that [16] "*the two variables appearing are independent and hence the expectation of their product equals the product of their individual expectations, i.e., zero.*" We have found both via computer simulations and analytical calculations that this is not the case, hence proving that Moon's formula for $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$ is inaccurate, as
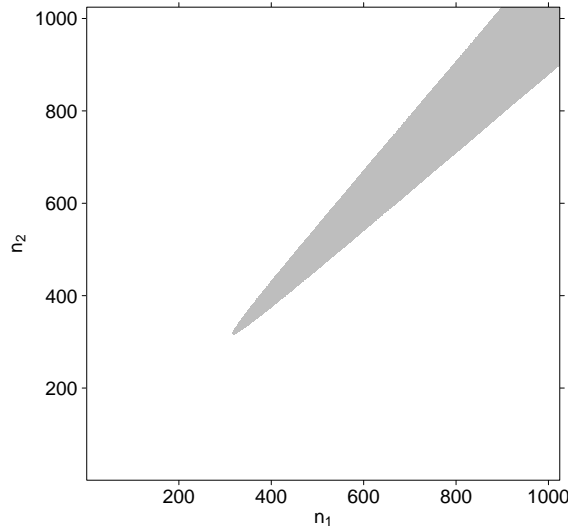
Figure 7: The sign of $\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right] - \mathbb{V}_{rsa}^{(M)}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right]$ as a function of n_1 and n_2 (white for positive and gray for strictly negative).

shown in sections 4.1 and 4.2.

We are aware of an erratum of Moon's article [17] where it is acknowledged that the initial derivation might be inaccurate and provides two arithmetic expressions for $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$, i.e.,

$$\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_n)\right] = \frac{\pi^2 - 8}{512\pi^2}n^6 + O\left(n^5\right) \tag{33}$$

$$\mathbb{V}_{rsa}\left[\mathbf{C}(\mathcal{K}_{n_1,n_2})\right] = \frac{\pi^2 - 8}{128\pi^2}n_1^2 n_2^2(n_1^2 + n_2^2) + O\left(n_1^2 n_2^2(n_1 + n_2)\right). \tag{34}$$

Although [17] does not provide any explanation concerning why the initial derivation was inaccurate, our asymptotic equations (31)-(32) are consistent with Moon's (33)-(34).

In this work, besides having contributed with an accurate expression for $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$, we have also shed light on the origins of the bias in Moon's formulae. Moreover, the values of $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$ coupled with the arithmetic expressions of the f_ω [2] pave the way towards the obtention of $\mathbb{V}_{rsa}\left[\mathbf{C}\right]$ in other types of graphs.

Our current revision of [16] offers various possibilities for future research. From the exact calculation of $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$ for $\omega \in \{021, 022, 03, 13\}$, to the main goal of [16], that was to show that the distribution of $\mathbf{C}$ is asymptotically normal.

It is worth noting that the computational resources needed to test the mathematical arguments and results of [16] were not available at that time. Crucially, they are needed to approximate numerically the values of $\mathbb{E}_{rsa}\left[\gamma_\omega\right]$ for $\omega \in \{021, 022, 03, 13\}$ (section 6). Perhaps not so surprisingly, computers are helping to validate and improve classic work from the 1960-70s (e.g., [6]). We hope that our research stimulates further research on crossings in spherical arrangements and other layouts.

6. Methods

Here we explain a way to generate points uniformly at random (u.a.r.) on the surface of the sphere (section 6.1), to determine if two arcs intersect (section 6.2), to speed up the intersection test (section 6.3), to estimate the values of $p_{rsa,\omega}$ and $\mathbb{E}_{rsa}[\gamma_\omega]$ (section 6.4), and, finally, a way to estimate $\mathbb{V}_{rsa}[\mathbf{C}]$ (section 6.5).

6.1. Generating random points

Generating points uniformly at random on the surface of a unit sphere was done by generating random polar coordinates following [20]. First, we generated random values for variables $u, v \in [0, 1]$. Then we generated the polar coordinates with

$$\theta = 2\pi u, \quad \phi = \cos^{-1}(2v - 1),$$

and, finally, the coordinates (x, y, z) of a point were calculated using

$$x = \sin(\theta)\cos(\phi), \quad y = \sin(\theta)\sin(\phi), \quad z = \cos(\theta).$$

The random values for generating points uniformly at random on the surface of a sphere were generated using the C++'s header *random*: we used the *default_random_engine* initialized with the C++'s *random_device*, a non-deterministic random number generator that uses hardware entropy source. This is used to seed the random engine once for each set of tests, that is, for a whole set of replicas. The *default_random_engine* is then used in the *uniform_real_distribution* object to generate floating-point pseudo-random numbers uniformly at random.

6.2. Arc-arc intersection test

In [16], Moon mentioned in passing possible ambiguities in the definition of edge crossing "*arising when different vertices coincide, when two vertices are diametrically opposed, when more than two vertices lie on the same great circle, or when more than two arcs intersect at the same point may be disregarded as they occur only with probability zero.*" The situation that "*more than two arcs intersect at the same point*", as explained in the introduction when introducing the generic arrangement $*$, contributes to $\mathbf{C}$ with $\binom{e}{2}$ crossings, where e is the number of arcs. The other two cases, still have probability zero but can appear when calculating arc-arc intersections computationally due to lack of numerical precision. Below we cover them in a formalization of the notion of arc-arc intersection.

Let $a_1 = (S, T)$ and $a_2 = (U, V)$ be the two arcs with points S, T, U, V pairwise distinct. Figure 8(a) illustrates an example of two arcs that cross and figure 8(b) two that do not cross. We adopt the following operational definition of arc intersection: arcs a_1 and a_2 intersect if, and only if, the triangles $t_1 = (O, S, T)$, $t_2 = (O, U, V)$ intersect at some point other than O.
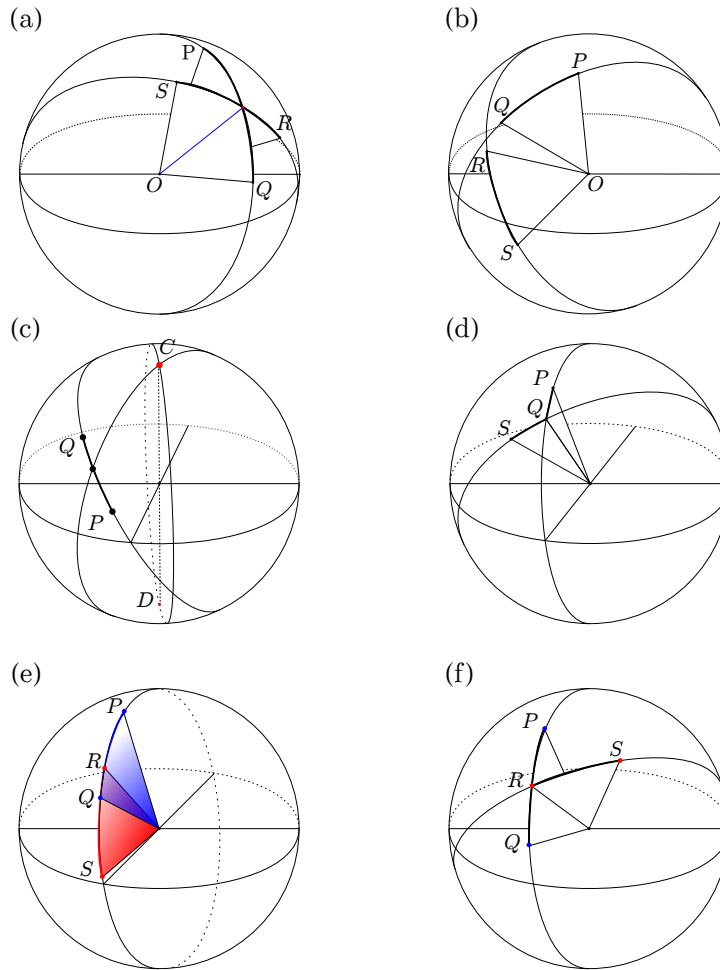
Figure 8: The different scenarios in which two great arcs might arcs cross. (a) Two arcs that cross. Their triangles intersect at a segment. (b) Two arcs that cross. Their triangles intersect only at point O. (c) Diametrically opposed points (**C** and D, in red), two possible great circles through **C** and D. One of the great circles crosses great arc PQ. (d) Two arcs sharing an endpoint. (e) 4 points lying on the same great circle (overlapping): intersection between the triangle of the arc PQ (blue) and the arc RS (red) at an infinite number of points. (f) 3 points lying on the same great circle. Intersection at a single point.

In 8(a) the triangles intersect at a segment while in figure 8(b), the triangles intersect in only at point O. If the intersection between the triangles is a single point then it must be O. There is an extreme case of intersection only at the origin that is worth mentioning: when an arc has its two endpoints diametrically opposed to each other (figure 8(c)) as pointed out by Moon. That is, the shortest Euclidean distance between them is exactly the diameter of the sphere. Such arc defines a degenerate triangle (actually a segment) that intersects the triangle of the other arc at O. *A priori*, any second arc, with its endpoints diametrically opposed or not, always intersects the first. However, we do not consider this as a valid intersection due to our operational definition.

In our implementation of the test, this situation is detected by testing whether any of the triangles t_1 or t_2 is degenerated, i.e., by testing whether t_1 or t_2 is actually a line segment.

Now we review two other situations considered by Moon that have probability zero in relation to our framework:

- The situation that *"more than two vertices lie on the same great circle"* can happen in two circumstances: when four points lie on the same great circle (the intersection between the two triangles is another triangle) then the arcs may overlap (figure 8(e)) and also when three points lie on the same great circle one of the endpoints of one arc may be between the endpoints of the other arc (figure 8(f)).

- When the intersection between the triangles is a segment then the arcs intersect at exactly one point, which might be one of the endpoints. Figure 8(a) and figure 8(f) shows two arcs crossing, not at an endpoint. Figure 8(d) illustrates *"when different vertices coincide"*, namely an arc-arc intersection at an endpoint.

We implemented the arc-arc intersection test using the CGAL library [1] (version 4.11.1). Computations were done using CGAL's kernel 'Exact_predicates_exact_constructions_kernel'.

6.3. Arc-arc intersection test speed-up

The kernel above provides a highly precise but time-consuming arc-arc intersection test. Because of the latter, we run a fast test to filter out those pairs of arcs that do not cross. This test is based on three sufficient conditions for non-intersection. If these conditions fail, then we run the time-consuming intersection test provided by CGAL.

The sufficient conditions are based on a division of the sphere's surface into 8 octants. Each octant is defined by the sign of the coordinates of the points in them. The octant of a point $A = (x, y, z)$ is $\mathrm{oct}(A) = (\mathrm{sgn}(x), \mathrm{sgn}(y), \mathrm{sgn}(z))$, where sgn is the sign function, i.e.

$$\mathrm{sgn}(x) = \begin{cases} \frac{x}{|x|} \text{ if } x \neq 0 \\ 1 \text{ if } x = 0. \end{cases}$$

We now define the three sufficient conditions for non-intersection. Any pair of disjoint arcs ST and UV does not cross

- If the arcs fall into different octants, namely $\mathrm{oct}(S) = \mathrm{oct}(T) \neq \mathrm{oct}(U) = \mathrm{oct}(V)$, then the arcs cannot intersect.

- If ST and UV are separated by one of the planes $x = 0$, $y = 0$, or $z = 0$. Let $\mathrm{oct}_c(P) = \mathrm{sgn}(c)$ for $c \in \{x, y, z\}$ of a point $P = (x, y, z)$. Formally, two arcs ST and UV are separated by one of the planes $x = 0$, $y = 0$, $z = 0$ when there exists a coordinate c such that $\mathrm{oct}_c(S) = \mathrm{oct}_c(T) \neq \mathrm{oct}_c(U) = \mathrm{oct}_c(V)$.

- If points U, V are located in the same half space defined by the plane through points S, T, O then the arcs do not intersect.

In order to achieve a higher reduction of the amount of arc-arc intersection tests, we apply the first two sufficient conditions defined above to $3l$, $l \geq 1$, transformations of the original set of points $\mathcal{S}$. The ith group of three transformations consists of rotations of all points around each axis separately by angle θ_i. More precisely, the set of $3l$ transformations of $\mathcal{S}$ is

$$\{R_{\theta_i}^{OX}(\mathcal{S}), R_{\theta_i}^{OY}(\mathcal{S}), R_{\theta_i}^{OZ}(\mathcal{S})\}_{i=1}^{l}$$

where $R_{\theta_i}^{e}(\mathcal{S}) = \{R_{\theta_i}^{e}(P) \mid P \in \mathcal{S}\}$ is the rotation of all points in S around axis e by an angle θ_i. The angles θ_i are defined as $\theta_i = i\pi/2(l+1)$ for all $i \in \{1, ..., l\}$. If one of the sufficient conditions defined above is true for a given pair of independent arcs in some of the transformations then the arcs do not intersect.

6.4. Estimating the expectation of types

The integrals used to calculate the values of the $p_{rsa, \omega}$ (section 3) have been approximated numerically using the computer algebra system Maple$^{\text{TM}}$‡ [15]. We used the function *integrate* with the parameter *method* set to *CubaDivonne* and *CubaCuhre* depending on the case.

Each $p_{rsa, \omega}$ was estimated numerically by generating $T = 10^7$ random spherical arrangements (section 6.1) of a $\mathcal{K}_{10}$. This implies that each $p_{rsa, \omega}$ was estimated over $T f_\omega$ replicas, where f_ω is given in table 3.

For each random layout, we calculated what pairs of arcs intersected (sections 6.2 and 6.3). Then we classified each pair of independent arcs, i.e., each of the elements in $Q \times Q$, and used this information to compute the $p_{rsa, \omega}$ and then $\mathbb{E}_{rsa}[\gamma_\omega]$.

Our simulations confirm the correctness of the results obtained in section 3. Our T simulations yielded the results presented in the middle columns of table 6. Since the values were obtained by classifying the different elements of $Q \times Q$ in a complete graph, each type was sampled at different amounts. This means that the precision at which they were obtained is different for every type, which we convey in the rightmost column of table 6. That column indicates the amount of iterations for which the last decimal of the estimate of $p_{rsa, \omega}$ did not change before reaching the end of the simulation.

6.5. Estimating $\mathbb{V}_{rsa}[\boldsymbol{C}]$ in a graph

Estimating $\mathbb{V}_{rsa}[\mathbf{C}]$ on a graph consists of

(i) Generating N random spherical layouts (section 6.1),

(ii) Calculating the value of $\mathbf{C}$ for each layout (using the arc-arc intersection test described in section 6.2 , optionally with the improvements explained in section 6.3),

(iii) Applying an unbiased estimator of variance to the N values of $\mathbf{C}$.

‡ Maple is a trademark of Waterloo Maple Inc.

Table 6: Numerical estimates of $p_{rsa,\,\omega}$ and $\mathbb{E}_{rsa}\,[\gamma_\omega]$ obtained via computer simulations, for all types $\omega \in \Omega$ (10). Values $\hat{p}_{rsa,\,\omega}$ have been truncated to the last decimal that did not change in the last amount iterations indicated in the column "Iterations" (a "-" indicates that the value remained the same throughout the whole simulation). f_ω is the amount of products of type ω that can be found in $\mathcal{K}_{10}$ (they are obtained from table 3 with $n = 10$). The number of samples used to calculate the numerical estimates is Tf_ω.

ω	$\hat{p}_{rsa,\,\omega}$	$\hat{\mathbb{E}}_{rsa}\,[\gamma_\omega]$	f_ω	Iterations
00	0.0156253	0.0000003	28350	4750
01	0.0156258	0.0000008	151200	5750
021	0.0126703	-0.0029546	75600	5600
022	0.0185812	0.0029562	75600	5550
03	0.010417	-0.005207	30240	49900
04	0.0000000	-0.0156250	1260	-
12	0.01858	0.00295	18900	>40000
13	0.0312507	0.0156257	15120	3650
24	0.125001	0.109376	630	39250

However, for large values of n (in complete graphs), or large values of n_1 and n_2 (in complete bipartite graphs), estimating the variance with $N = 10^7$ replicas turned out to be a rather time-consuming task. Therefore, for large n (or n_1 and n_2), the first two steps were organized into partitions and then parallelized (each partition was in charge of the processing a certain number of random layouts).

Acknowledgments

Appendix A. Relative frequencies of types

We recall the definition of the falling factorial, i.e. [4]

$$(n)_x = n(n-1)(n-2)...(n-x+1) = x!\binom{n}{x}.$$

We derive f_ω/q for all types of products with the help of the values of f_ω for complete graphs and complete bipartite graphs (table 3) and a property of the quotient of binomial

coefficients, namely

$$\frac{\binom{n}{x}}{\binom{n}{y}} = (n-y)_{x-y}\frac{y!}{x!}.$$

when $x \geq y$. The results are summarized in table 5.

Appendix A.1. Complete graphs

First,

$$\frac{f_{00}}{q} = \frac{630\binom{n}{8}}{3\binom{n}{4}} = \frac{1}{8}(n-4)_4 = \frac{4!}{8}\binom{n-4}{4} = 3\binom{n-4}{4}.$$

Second,

$$\frac{f_{01}}{q} = \frac{1260\binom{n}{7}}{3\binom{n}{4}} = 2(n-4)_3 = 2\cdot 3!\binom{n-4}{3} = 12\binom{n-4}{3}$$

$$\frac{f_{021}}{q} = \frac{f_{022}}{q} = \frac{360\binom{n}{6}}{3\binom{n}{4}} = 4(n-4)(n-5) = 8\binom{n-4}{2}.$$

As $f_{12} = f_{021}/4$, we also get

$$\frac{f_{12}}{q} = (n-4)(n-5) = 2\binom{n-4}{2}.$$

Fourth,

$$\frac{f_{03}}{q} = \frac{120\binom{n}{5}}{3\binom{n}{4}} = 8(n-4).$$

As $f_{13} = f_{03}/2$, we also get $f_{13}/q = 4(n-4)$. Fifth, $f_{04}/q = 2$ and $f_{24}/q = 1$ trivially.

Appendix A.2. Complete bipartite graphs

First,

$$\frac{f_{00}}{q} = \frac{1}{2}(n-2)_2(m-2)_2 = 2\binom{n-2}{2}\binom{m-2}{2}.$$

Second, the fact that

$$\frac{\binom{n}{4}\binom{m}{3}}{q} = \frac{1}{36}(n-2)_2(m-2)$$

gives

$$\frac{f_{01}}{q} = 144\left[\frac{\binom{n}{4}\binom{m}{3}}{q} + \frac{\binom{n}{3}\binom{m}{4}}{q}\right] = 4(n-2)(m-2)(n+m-6).$$

Third, the fact that

$$\frac{\binom{n}{3}\binom{m}{3}}{q} = \frac{1}{18}(n-2)(m-2)$$

gives

$$\frac{f_{021}}{q} = 72\frac{\binom{n}{3}\binom{m}{3}}{q} = 4(n-2)(m-2)$$

and

$$\frac{f_{12}}{q} = 36\frac{\binom{n}{3}\binom{m}{3}}{q} = 2(n-2)(m-2).$$

Fourth, the fact that

$$\frac{\binom{n}{2}\binom{m}{4}}{q} = \frac{1}{24}(m-2)_2$$

and that f_{12} is included in the definition of f_{022} as third summand, gives

$$\frac{f_{022}}{q} = 24\left[\frac{\binom{n}{2}\binom{m}{4}}{q} + \frac{\binom{n}{4}\binom{m}{2}}{q}\right] + f_{022}$$

$$= (m-2)_2 + (n-2)_2 + 2(n-2)(m-2)$$

$$= \begin{cases} 2\left[\binom{m-2}{2} + \binom{n-2}{2} + (n-2)(m-2)\right] \\ (m+n-5)(m+n-4). \end{cases}$$

Finally, one has

$$\frac{f_{04}}{q} = \frac{f_{24}}{q} = 1$$

and also

$$\frac{f_{03}}{q} = \frac{f_{13}}{q} = 2(n+m-4)$$

trivially.

References

[1] CGAL, Computational Geometry Algorithms Library. http://www.cgal.org.

[2] L. Alemany-Puig and R. Ferrer-i-Cancho. Edge crossings in random linear arrangements. *Journal of Statistical Mechanics*, page 023403, 2020.

[3] M. Barthélemy. *Morphogenesis of Spatial Networks*. Springer, Cham, Switzerland, 2018.

[4] B. Bollobás. *Modern graph theory*. Springer-Verlag, 1998.

[5] W. Y. C. Chen, H. S. W. Han, and C. M. Reidys. Random k-noncrossing RNA structures. *Proceedings of the National Academy of Sciences*, 106(52):22061–22066, 2009.

[6] J. L. Esteban and R. Ferrer-i-Cancho. A correction on Shiloach's algorithm for minimum linear arrangement of trees. *SIAM Journal of Computing*, 46(3):1146–1151, 2015.

[7] R. Ferrer-i-Cancho. Random crossings in dependency trees. *Glottometrics*, 37:1–12, 2017.

[8] R. Ferrer-i-Cancho, C. Gómez-Rodríguez, and J. L. Esteban. Are crossing dependencies really scarce? *Physica A: Statistical Mechanics and its Applications*, 493:311–329, 2018.

[9] William Anthony Granville, P. F. Smith, and J. S Mikesh. *Plane and spherical trigonometry*. Ginn and Company, Boston, 1908.

[10] R. Guimerá and L.A.N. Amaral. Modeling the world-wide airport network. *European Physical Journal B*, 38:381–385, 2004.

[11] R. Guimerà, S. Mossa, A. Turtschi, and L. A. N. Amaral. The worldwide air transportation network: Anomalous centrality, community structure, and cities' global roles. *Proceedings of the National Academy of Sciences (USA)*, 102(22):7794–7799, 2005.

[12] Sandra Kübler, Ryan McDonald, and Joakim Nivre. Dependency parsing. 1(1):1–127, 2009.

[13] W. Li and X. Cai. Statistical analysis of airport network of China. *Phys. Rev. E*, 69(4):046106, 2004.

[14] H. Liu, C. Xu, and J. Liang. Dependency distance: A new perspective on syntactic patterns in natural languages. *Physics of Life Reviews*, 21:171–193, 2017.

[15] Maplesoft, a division of Waterloo Maple Inc. Maple.

[16] J. W. Moon. On the distribution of crossings in random complete graphs. *Journal of the Society for Industrial and Applied Mathematics*, 13(2):506–510, 1965.

[17] J. W. Moon. Erratum: On the distribution of crossings in random complete graphs. *Journal of the Society for Industrial and Applied Mathematics*, 32(3):706, 1977.

[18] Barry L. Piazza, Richard D. Ringeisen, and Sam K. Stueckle. Properties of nonminimum crossings for some classes of graphs. In Yousef Alavi et al., editor, *Proc. 6th Quadrennial Int. 1988 Kalamazoo Conf. Graph Theory Combin. Appl.*, volume 2, pages 975–989, New York, 1991. Wiley.

[19] David Temperley and Daniel Gildea. Minimizing syntactic dependency lengths: Typological/Cognitive universal? *Annual Review of Linguistics*, 4(1):67–80, 2018.

[20] Eric W. Weisstein. chapter Sphere point picking. Chapman & Hall/CRC, Boca Raton, 2nd edition, 2003.