Edge crossings in random linear arrangements

Lluís Alemany-Puig[1] **& 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.

Abstract. In spatial networks, vertices are arranged in some space and edges may cross. Here we consider the particular case of arranging vertices in a 1-dimensional lattice, where edges may cross when drawn above the vertex sequence as it happens in linguistic and biological networks. Here we investigate the distribution of edge crossings under the null hypothesis of a uniformly random arrangement of the vertices. We generalize the existing formula for the expectation of this number in trees to any graph. Furthermore, we present a characterization of the algebraic structure of the variance of crossings in an arbitrary space and apply it to derive a general expression for the variance of the number of crossings in random linear arrangements. We provide compact formulas of the expectation and variance of the number of crossings in complete graphs, cycle graphs, one-regular graphs and various kinds of trees (star trees, quasi-star trees and linear trees). In these graphs, the scaling of expectation and variance as a function of network size is asymptotically power-law-like. Our work paves the way for applications of the null hypothesis above in 1-dimension or investigating the distribution of the number of crossings in lattices of higher dimension.

Keywords: crossings in linear arrangements, variance of crossings.

PACS numbers: 89.75.Hc Networks and genealogical trees
89.75.Fb Structures and organization in complex systems
89.75.Da Systems obeying scaling laws

Contents

1 Introduction **3**

2 A mathematical theory of crossings **4**

2.1 The number of crossings of a complete graph 5
2.2 The potential number of crossings of a graph 6
2.3 When is $|Q|$ maximum? 7
2.4 When is $|Q|$ minimum? 8
2.5 Theoretical graphs 10

3 The expected number of crossings of a uniformly random linear arrangement **13**

3.1 The scaling of $\mathbb{E}[C]$ as function of n. 14

4 The variance of the number of crossings of a uniformly random linear arrangement **14**

4.1 The types of products 16
4.2 The variance of C as the function of the number of products 19
4.3 Preliminaries 22
4.4 General results 23
4.5 Theoretical formulae 28
4.5.1 $\tau = 2$, $\phi = 4$ 30
4.5.2 $\tau = 0$, $\phi = 0$ 30
4.5.3 $\tau = 1$, $\phi = 3$ 30
4.5.4 $\tau = 1$, $\phi = 2$ 32
4.5.5 $\tau = 0$, $\phi = 4$ 33
4.5.6 $\tau = 0$, $\phi = 3$ 33
4.5.7 $\tau = 0$, $\phi = 2$, Subtype 1 36
4.5.8 $\tau = 0$, $\phi = 2$, Subtype 2 40
4.5.9 $\tau = 0$, $\phi = 1$ 42
4.6 Variance of the number of crossings 43

5 Theoretical examples **44**

5.1 1-regular graphs 44
5.2 Quasi-star trees 45
5.3 Complete graphs 45
5.3.1 Variance 47
5.4 Complete bipartite graphs 47
5.4.1 Variance 49
5.5 Cycle graphs 50
5.5.1 Variance 54
5.5.2 Notes on the occurrences of edges 54

5.6 Linear trees . . . 55
5.6.1 Variance . . . 56
5.7 The scaling of $\mathbb{V}[C]$ as function of n. . . . 57
5.8 Summary . . . 58

6 Discussion 60

Appendix A Potential number of crossings of an edge 61

Appendix B Toy examples 61

Appendix C Protocol for testing 65

Appendix D Alternative proof for $f_{021} = 2n_G(\mathcal{L}_4 \oplus \mathcal{L}_2)$. 67

1. Introduction

The organization of many complex systems can be described with the help of a spatial network, where nodes and are embedded in some space [1]. When vertices are arranged in some space, edges may cross. A spatial graph without edge crossings is planar. In street networks, crossings are practically impossible by the construction of the network [1]. In road networks, crossings typically involve bridges and tunnels [2]. Thus, crossings in road networks are costly and consequently scarce. Crossings are also scarce in syntactic dependency networks, networks linking the words of a sentence via syntactic dependencies [3]. However, whether syntactic dependency crossings are inherently costly is a matter of debate [4].

Here we focus on in the particular case of embedding vertices in a one-dimensional lattice, where edges may cross when drawn above the lattice, as it happens in sentences [3] and RNA structures [5]. In the case of RNA structures, vertices are nucleotides A, G, U, and C, while edges are Watson-Crick (A-U, G-C) and (U-G) base pairs [5]. Here we aim at investigating the distribution of edge crossings under the null hypothesis of a uniformly random arrangement of the vertices. In trees, $\mathbb{E}[C]$, the expected number of crossings in such a random arrangement has been shown to be $|Q|/3$, where $|Q|$ is the size of the set Q, the pairs of edges of a network that may potentially cross [4]. A pair of edges belongs to Q if there is at least one linear arrangement where they cross. A pair of edges belongs to Q if the edges do not share vertices. In trees [6, 7],

$$|Q| = \frac{n}{2}(n - 1 - \langle k^2 \rangle), \tag{1}$$

where $\langle k^2 \rangle$ is the second moment of degree about zero. In many cases, the syntactic dependency networks of sentences are not trees [4] and RNA secondary structures are graphs where degrees do not exceed one and are then usually disconnected [5].

Consider a graph of n vertices and m edges whose vertices are arranged linearly with a function π that given a vertex returns its position in sequence of length n. The

1st and the last vertex in that ordering are located in position 1 and n, respectively. Throughout this article we use letters $s, t, ..., z$ to indicate vertices. Suppose that the vertices s and t and the vertices u and v are linked. Without any loss of generality suppose that s precedes t, i.e. $\pi(s) < \pi(t)$, and u precedes v, i.e. $\pi(u) < \pi(v)$. Then their edges cross if and only if one of the two following conditions is met

$$\pi(s) < \pi(u) < \pi(t) < \pi(v), \text{ or} \tag{2}$$

$$\pi(u) < \pi(s) < \pi(v) < \pi(t). \tag{3}$$

Let C the number of edge crossings produced by one of the $n!$ linear arrangement of the vertices of the graph. Figure 3 shows a couple of linear arrangements with $C = 1$ on top and linear arrangements with $C = 0$ below.

In this article we derive $\mathbb{E}[C]$ for general graphs and investigate the distribution of crossings under the null hypothesis by means of $\mathbb{V}[C]$. Such knowledge of the distribution of crossings under the null hypothesis has potentially many applications in biology and linguistics, e.g., it would facilitate the development of efficient tests of the significance of C.

The current is a piece of a broader research program on the statistical properties of measures on linear arrangements. Recently, the distribution of D, the sum of edge lengths in random linear arrangements, has been investigated [8]. The present article can be seen as a continuation where D is replaced by C, bearing in mind that the the analysis of C is more complex as we will see. Research on such a program can be classified according to the target: D [9, 10, 8], C [3] or the interplay between D and C [11, 12, 13, 14]. In some cases, the aspects are considered simultaneously [6, 15].

The remainder of the article is organized as follows. Section 2 presents a derivation of the number of crossings in complete graphs whose vertices are arranged linearly, reviews the concept of Q, extending it to general graphs and investigating graphs that minimize or maximize $|Q|$. This presents the specific graphs for which compact formulae of $\mathbb{E}[C]$ and $\mathbb{V}[C]$ are derived, in later sections. Section 3 presents a general expression for $\mathbb{E}[C]$ in graphs as well compact formulae for specific graphs. Section 4 analyses the mathematical structure of $\mathbb{V}[C]$ providing a general expression for $\mathbb{V}[C]$. Section 5 provides compact formulae for specific graphs. Section 6 summarizes and discusses the findings of previous sections.

2. A mathematical theory of crossings

The number of different pairs of edges that can be formed is

$$P = \binom{m}{2}.$$

Obviously,

$$C \leq P.$$

In $\mathcal{K}_n$, i.e. a complete graph of n vertices,

$$m = \binom{n}{2}, \tag{4}$$

and then the number of pairs of edges that can be formed is

$$\begin{aligned} P(\mathcal{K}_n) &= \binom{\frac{n(n-1)}{2}}{2} \\ &= \frac{1}{8}(n+1)n(n-1)(n-2). \end{aligned} \tag{5}$$

As m is maximum in a complete graph, we have that

$$C \leq P(\mathcal{K}_n).$$

We will show that the actual number of crossings of a complete graph is actually 3 times smaller than $P(\mathcal{K}_n)$.

2.1. The number of crossings of a complete graph

In complete graphs, the number of crossings does not depend on the linear arrangement because all vertices have maximum degree. Therefore we can refer to the number of crossings of a complete graph without specifying the linear arrangement that produces it.

In linear arrangement, we define the shadow of an edge as the vertices that are placed in-between the first vertex of the edge and the last vertex of the edge. In a graph of $n \geq 1$ vertices, d, the length of an edge, satisfies $1 \leq d \leq n-1$.

In an arbitrary graph, $f(d)$, the number of edges of length d, satisfies [8]

$$f(d) \leq f_{max}(d) = n - d, \tag{6}$$

and $C(d)$, the number of edges that cross an edge of length d, satisfies [6]

$$C(d) \leq C_{max}(d) = (d-1)(n-d-1). \tag{7}$$

Notice that $d-1$ is the number of vertices of the shadow of an edge of length d and $n-d-1$ is the number of vertices excluding the vertices in the shadow and the vertices of the edge. Therefore, $C(d)$ cannot exceed $(d-1)(n-d-1)$. In addition, the number of crossings satisfies

$$C \leq \frac{1}{2}\sum_{d=1}^{n-1} f_{max}(d)C_{max}(d). \tag{8}$$

As

$$C(\mathcal{K}_n) = \frac{1}{2}\sum_{d=1}^{n-1} f_{max}(d)C_{max}(d), \tag{9}$$

C is maximized by a complete graph.

Applying equations 6 and 7 to equation 9, one obtains

$$C(\mathcal{K}_n) = \frac{1}{2}\sum_{d=1}^{n-1}(n-d)(d-1)(n-d-1)$$
$$= \frac{1}{24}n(n-1)(n-2)(n-3)$$

for $n \geq 4$. Noting that $C = 0$ for $n < 4$, we get, for $n \geq 0$,

$$C(\mathcal{K}_n) = \binom{n}{4}. \tag{10}$$

The same value of $C(\mathcal{K}_n)$ has been derived recently using a different approach [16]. Equations 10 and 5 allow one to calculate the ratio

$$\frac{P(\mathcal{K}_n)}{C(\mathcal{K}_n)} = \frac{3(n+1)}{n-3}.$$

For sufficiently large n, $\frac{P(\mathcal{K}_n)}{C(\mathcal{K}_n)} \approx 3$. In sum, taking into account the spatial constraints of linear arrangements, it turns out that the actual number of crossings in a complete graph is about 3 times smaller than its number of edge pairs.

2.2. The potential number of crossings of a graph

We define Q as the set of pairs of edges that may potentially cross in a linear arrangement of the vertices [4]. A pair of edges belongs to Q if and only if there is at least one linear arrangement where the two edges cross. Obviously,

$$C \leq |Q|. \tag{11}$$

where $|Q|$ is the cardinality of Q. $|Q|$ can be defined equivalently as the number of independent pairs of edges of a graph [17]. Two edges are independent or disjoint if and only if they do not have a common endpoint. Then $|Q|$ can be easily derived as the difference between the number of pairs of edges that can be formed, i.e. $\binom{m}{2}$ and the number of dependent pairs of different edges produced by every edge. A vertex of degree k produces $\binom{k}{2}$ dependent edges, i.e. [17]

$$|Q| = \binom{m}{2} - \sum_{s=1}^{n}\binom{k_s}{2}, \tag{12}$$

which is equivalent to [8]

$$|Q| = \frac{1}{2}\left[m(m+1) - n\left\langle k^2\right\rangle\right]. \tag{13}$$

Assuming that $m = n - 1$, e.g., a tree, equation 13 gives equation 1, which has already been derived for the particular case of trees [6, 7].

A k-regular graph, denoted as $\mathbf{k}_n$, is a graph where all nodes have degree k [18, p. 4] (and also [19]). In a k-regular graph, $\langle k\rangle = k$, $\langle k^2\rangle = k^2$ and $m = kn/2$. Thus equation 13 becomes

$$|Q(\mathbf{k}_n)| = \frac{1}{2}\left[\frac{kn}{2}\left(\frac{kn}{2}+1\right) - nk^2\right]$$
$$= \frac{1}{8}kn(k(n-4)+2). \tag{14}$$

In a complete graph, $k = n - 1$ and then

$$|Q(\mathcal{K}_n)| = \frac{1}{8}n(n-1)(n-2)(n-3)$$

for $n \geq 4$, in agreement with previous work [8]. Noting that $C = 0$ for $n < 4$ in general, we may finally write

$$|Q(\mathcal{K}_n)| = 3\binom{n}{4} \tag{15}$$

for $n \geq 0$. Recalling equation 10, it turns out that

$$C(\mathcal{K}_n) = \frac{|Q(\mathcal{K}_n)|}{3}. \tag{16}$$

Equation 15 is actually equivalent to one derived in previous work [20], i.e.

$$|Q(\mathcal{K}_n)| = \frac{1}{2}\binom{n}{2}\binom{n-2}{2}. \tag{17}$$

The last expression and equation 4 allow one to calculate the average number of crossings per edge of a complete graph easily, i.e.

$$\begin{aligned} C(\mathcal{K}_n)/m &= \frac{|Q(\mathcal{K}_n)|}{3m} \\ &= \frac{1}{6}\binom{n-2}{2}. \end{aligned}$$

Next subsections are concerned about the maxima and the minima of $|Q|$. These are not only relevant for crossing theory *per se* but also because $\mathbb{E}[C]$ is proportional to $|Q|$, as we will see in section 3.

2.3. When is $|Q|$ maximum?

C, the number of crossings of that graph can be defined as

$$C = \frac{1}{2}\sum_{(s,t)} C(s,t),$$

where $C(s,t)$ is the number of edge crossings involving edge $\{s,t\}$.

Let $A = \{a_{st}\}$ be the adjancency matrix of a graph, i.e. $a_{st} = 1$ if vertices u and v are connected and $a_{st} = 0$ otherwise. Then C can be defined equivalently as

$$C = \frac{1}{4}\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}C(s,t). \tag{18}$$

$C(s,t)$ cannot exceed $q(s,t)$, the potential number of crossings of the edge formed by s and t, namely the number of edges that do not share a vertex with the pair (s,t), or equivalently, the number of edges for which there exists at least one linear arrangement where they cross the edge formed by u and v. It is easy to see that

$$q(s,t) = m - k_s - k_t + 1, \tag{19}$$

where k_t is the degree of vertex t (see Appendix A for a detailed proof). Equation 19 is actually a generalization of a previous result, i.e.

$$q(s,t) = n - k_s - k_t$$

for trees, where $m = n - 1$ [7].

Applying equation 19 to equation 18, we get

$$|Q| = \frac{1}{4}\sum_{s=1}^{n}\sum_{t=1}^{n} a_{st}(m + 1 - k_s - k_t). \tag{20}$$

We derive now a tight upper bound of $|Q|$ from equation 20. As $a_{st} \in \{0, 1\}$ and $C(s,t) \in \{0, m - 1 - k_s - k_t\}$, we have that

$$a_{st}C(s,t) \leq m - 1 - k_s - k_t.$$

Then noting that $a_{tt} = 0$, equation 20 gives

$$\begin{aligned} |Q| &\leq \frac{1}{4}\sum_{s=1}^{n}\sum_{t=1, t\neq s}^{n} (m + 1 - k_s - k_t) \\ &= \frac{1}{4}\sum_{s=1}^{n}\sum_{t=1, t\neq s}^{n} (m + 1) - 2\sum_{s=1}^{n}\sum_{t=1, t\neq s}^{n} k_t \\ &= n(n-1)(m+1) - 4m(n-1) \\ &= \frac{n-1}{4}[(n-4)m + n]. \end{aligned}$$

Replacing m by its maximum value, namely that of a complete graph, we finally obtain

$$\begin{aligned} |Q| &\leq \frac{1}{8}n(n-1)(n-2)(n-3) \\ &= |Q(\mathcal{K}_n)|. \end{aligned}$$

Therefore, $|Q|$ is maximized by a complete graph.

2.4. When is $|Q|$ minimum?

In addition to $\mathcal{K}_n$ and $\mathbf{k}_n$, we will use specific notation to refer to other kinds of graphs: $\mathcal{S}_n$ for a star tree of n vertices, $\mathcal{C}_n$ for a cycle graph of n vertices (figure 1(a)). These graphs are related: e.g., $\mathbf{n\text{-}1}_n = \mathcal{K}_n$, $\mathcal{K}_3 = \mathcal{C}_3$ and $\mathcal{C}_n$ is a kind of $\mathbf{2}_n$. We define the operator $\oplus$ for the union of disjoint graphs, i.e. $G \oplus G'$ is the graph formed by two graphs, G and G', that do not share vertices [21]. For instance, $\mathcal{K}_2 \oplus \mathcal{K}_2$, a graph formed by two independent edges, is equivalent to $\mathbf{1}_4$. Isolated vertices, vertices of degree zero, are also referred as unlinked vertices.

Asking when $|Q|$ is minimum is equivalent to asking when $|Q| = 0$ in equation 13, which is in turn equivalent to

$$\langle k^2 \rangle = \frac{m(m+1)}{n}. \tag{21}$$

The minima of $|Q|$ cannot have crossings in a one-dimensional layout ($C \leq |Q| = 0$). Thus these minima are trivially a subset of outerplanar graphs (a graph is outerplanar if and only if its book thickness is one [22]). We show that the graphs satisfying equation 21 are actually sparser. Due to being outerplanar, the number of edges of the minima of $|Q|$ must satisfy [23]

$$m \leq 2n - 3.$$

Indeed, now we show that the graphs where $|Q| = 0$ are a subset of outerplanar graphs whose members satisfy

$$m \leq n$$

with equality if and only if $n = 3$.

We derive the kinds of graphs where $|Q| = 0$. The derivation is based on the following principle. Let G' be a subgraph of a graph G and let Q' be the set of pairs of edges that may potentially cross in G'. If $|Q'| > 0$ then $|Q| > 0$. Two kinds of graphs are vital for the derivation. First, cycle graphs, where all vertices have degree 2 and then $\langle k^2 \rangle = 4$ for $n \geq 3$ (for $n < 3$ a regular graph with vertices of degree 2 cannot be formed) and $m = n$. Applying these two properties to equation 13, one gets

$$|Q| = \frac{n}{2}(n - 3)$$

for $n \geq 3$. Namely, $|Q| = 0$ if $n = 3$ and $|Q| > 0$ for $n > 3$. The other kind of graph is a paw, namely a triangle (i.e. a cycle graph of tree vertices) with a leaf attached to it (figure 1(b)) [24]. This graph has $m = 4$. Its two vertices of degree 2, its single vertex of degree 1 and its single vertex of degree 3 give $\langle k^2 \rangle = 9/2$. Applying these properties to equation 13, one gets $|Q| = 1$ for that graph.

The derivation is as follows:

- A graph G where $|Q| = 0$ may have more than one connected component but must have all edges concentrated on one of the connected components. If a graph has two edges, e_1 and e_2, from different connected components, then $|Q| > 0$ because $\{e_1, e_2\} \in Q$. Notice that e_1 and e_2 cannot share vertices for belonging to different components.
- In a graph G where all edges belong to just one of the connected components, whether $|Q| = 0$ or not is determined by the connected component that contains the edges as all other components do not have edges. Let G' be the subgraph induced by the largest connected component of G.
  - Suppose that G' is a tree. Then $|Q'| = 0$ if and only if G' is a star tree [7].
  - If G' is not a tree then it must be a connected graph with cycles. There are only three possibilities:
    * G' has a cycle of 4 or more vertices and then $|Q'| > 0$ borrowing the results above on cycle graphs.
    * G' is just a cycle of 3 vertices then $|Q'| = 0$.
    * G' contains a cycle of 3 vertices and additional vertices. Then $|Q| > 0$ because it contains the paw graph (figure 1(b)).

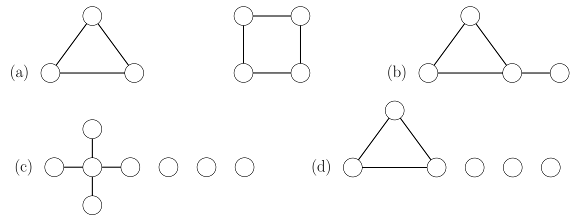


Figure 1: Drawings of simple graphs. (a) cycle graphs, (b) a paw (a triangle with a node attached), (c) a star tree with isolated vertices, and (d) a cycle with isolated vertices.

> Notice that if G' does not have any cycle then it has to be tree because it is connected. Therefore, the only cyclic G' that has $|Q| = 0$ is a triangle (a cycle graph of 3 vertices).

We conclude that $|Q| = 0$ is only possible in two kinds of graphs,

(i) $\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$, a forest consisting of a star tree of λ vertices and a series of $n - \lambda$ unlinked vertices (figure 1(c)).

(ii) $\mathcal{K}_3 \oplus \mathbf{0}_{n-3}$, a graph consisting of a complete graph of 3 vertices (namely a cycle graph of 3 vertices) and $n - 3$ isolated vertices (figure 1(d)).

We check the condition in equation 21 is satisfied by $\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$, that has a hub vertex of degree m, m vertices of degree 1 and $n - m - 1$ isolated vertices. Therefore,

$$\begin{aligned} \langle k^2 \rangle (\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}) &= \frac{1}{n} \sum_{s=1}^{n} k_s^2 \\ &= \frac{1}{n} m(m+1) \end{aligned} \tag{22}$$

as expected by equation 21. It is also easy to check that $\mathcal{K}_3 \oplus \mathbf{0}_{n-3}$ satisfies equation 21 because $m = 3$ and $\langle k^2 \rangle = 12/n$.

Now let us derive a tight upper bound of m for graphs where $|Q| = 0$. For $\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$ we have have $m \leq n - 1$ while for $\mathcal{K}_3 \oplus \mathbf{0}_{n-3}$ we have $m = 3$. Thus we have

$$m \leq \begin{cases} n - 1 & \text{if } n < 3 \\ \max(n-1, 3) & \text{if } n \geq 3 \end{cases}$$

and then

$$m \leq n$$

with equality if and only if $n = 3$.

2.5. Theoretical graphs

In this article we consider the following special graphs.

First, complete graphs. They are interesting for various reasons. They maximize m, $\langle k^2 \rangle$, C and $|Q|$; C is constant, $C < |Q|$ (for $n > 3$) and $\mathbb{V}[C] = 0$ because C is

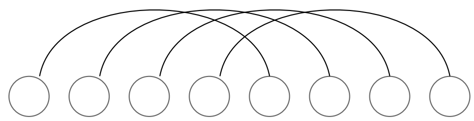


Figure 2: Linear arrangement of a 1-regular graph with $n = 8$ that maximizes the number of crossings ($C = |Q| = 6$).

constant. In addition, they are a chosen in many contexts for the ease with which they allow one to obtain theoretical results, e.g. random layouts on the surface of a sphere [20], neural networks [25] or social dynamics [26].

Second, complete bipartite graphs, $\mathcal{K}_{n_1,n_2}$, where n_1 and n_2 are the number of vertices of each partition. They are relevant for the close relationship between present article previous work on random layouts on the surface of a sphere [20]. It is easy to see that

$$\langle k^2 \rangle(\mathcal{K}_{n_1,n_2}) = m = n_1 n_2. \tag{23}$$

Applying this result to equation 13, one obtains

$$|Q(\mathcal{K}_{n_1,n_2})| = \frac{1}{2} n_1 n_2 (n_1 - 1)(n_2 - 1) = 2\binom{n_1}{2}\binom{n_2}{2} \tag{24}$$

after some routine work.

Third, $\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$. In these graphs $m = n - \lambda - 1$. They are interesting because $C = |Q| = 0$ (section 2.4).

Fourth, $\mathbf{1}_n$, 1-regular graphs. Trivially, $\langle k^2 \rangle = 1$. By definition, $m = n/2$ and n must be even. These graphs are relevant because all edges are independent and then

$$|Q| = \binom{m}{2},$$

which in turn implies that they maximize $|Q|$ given m. The fact that $m = n/2$ gives

$$|Q| = \binom{n/2}{2}.$$

We could have reached the same conclusion applying $k = 1$ to equation 14. Given an edge between vertices u and v, the initial position of the edge is $min(\pi(u), \pi(v))$ and its length is $|\pi(u) - \pi(v)|$, where $|\cdot|$ is the absolute value operator. 1-regular graphs achieve maximum C ($C = |Q|$) when the initial positions of each edge are consecutive and all edges have the same length, i.e. the length is $m = n/2$ (figure 2).

1-regular graphs are not the only graphs where all edges are independent. Indeed, the class of graphs where all edges are independent is formed by forests that result from the combination of a 1-regular graph (that could be empty) with an arbitrary number of unlinked vertices, i.e. $\mathbf{1}_{n_1} \oplus \mathbf{0}_{n_2}$ with $n = n_1 + n_2$. For simplicity, we restrict our analyses to pure 1-regular graphs.

Fifth, cycle graphs of n vertices, denoted as $\mathcal{C}_n$. They are interesting for being cyclic graphs with only one cycle, in opposition to complete graphs, where the number of cycles is maximized. As cycle graphs are two regular, equation 14 with $k = 2$ gives

$$|Q(\mathcal{C}_n)| = \frac{1}{3}n(n-3).$$

Notice that cycle graphs are interesting *a priori* for being like a linear tree but with "periodic boundary conditions", as in lattice field theory [27].

Sixth, trees because they are involved in spatial networks that have received a lot of attention [28, 29]. In this article we pay specific attention to kinds of trees that are relevant in crossing theory of trees [6, 15]:

- Star trees, $\mathcal{S}_n$. They are a special case of the forests above and then $|Q|$ is minimum, i.e. $|Q| = 0$, which in turn implies $C = 0$ [6]. In addition,

$$\langle k^2 \rangle(\mathcal{S}_n) = n - 1 \tag{25}$$

  is maximum among all trees with same n [6].
- Quasi-star trees, $\mathcal{Q}_n$, a graph in which $n-1$ of the vertices form a star tree and the n-th vertex forms an edge with one of the vertices but the central [15]. They interesting because they are a special case of the forests above and also because among all trees with same n,

$$|Q(\mathcal{Q}_n)| = n - 3$$

  is the smallest non-zero value of $|Q|$ while

$$\langle k^2 \rangle(\mathcal{Q}_n) = \frac{1}{n}(n^2 - 3n + 6)$$

  is the second largest possible value of $\langle k^2 \rangle$ [15].
- Linear trees, $\mathcal{L}_n$. They are interesting because, among all trees with same n,

$$|Q(\mathcal{L}_n)| = \binom{n-2}{2}$$

  is maximum, and

$$\langle k^2 \rangle(\mathcal{L}_n) = 4 - \frac{6}{n} \tag{26}$$

  is minimum [6].

A general interest of the special graphs above is that a compact or simpler formula for $\mathbb{V}[C]$ is easy to derive for the majority of them. Table 1 summarizes the properties of these graphs. Their values of $\langle k^2 \rangle$ and $|Q|$ have been presented above (either derived of borrowed from from previous work). Their values of $\mathbb{E}[C]$ and $\mathbb{V}[C]$ will be derived in the coming sections.

Table 1: A summary of the properties of the special graphs considered in this article as a function of n, the number of vertices. $\langle k^2 \rangle$ is the average of squared degrees, $|Q|$ is the number of pairs of independent edges; $\mathbb{E}[C]$ and $\mathbb{V}[C]$ are, respectively, the expectation and the variance of the number of crossings C in uniformly random linear arrangements.

Graph	$\langle k^2 \rangle$	$\lvert Q \rvert$	$\mathbb{E}[C]$	$\mathbb{V}[C]$
$\mathcal{K}_n$	$(n-1)^2$	$3\binom{n}{4}$	$\binom{n}{4}$	0
$\mathcal{K}_{n_1,n_2}$	$n_1 n_2$	$2\binom{n_1}{2}\binom{n_2}{2}$	$\frac{2}{3}\binom{n_1}{2}\binom{n_2}{2}$	$\frac{1}{90}\binom{n_1}{2}\binom{n_2}{2}((n_1+n_2)^2+n_1+n_2)$
1_m	1	$\binom{n/2}{2}$	$\frac{1}{3}\binom{n/2}{2}$	$\frac{1}{360}(n-2)n(n+6)$
$\mathcal{C}_n$	4	$\frac{n(n-3)}{2}$	$\frac{n(n-3)}{6}$	$\frac{1}{90}n(2n^2-n-30)$
$\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$	$\frac{1}{n}m(m+1)$	0	0	0
$\mathcal{S}_n$	$n-1$	0	0	0
$\mathcal{Q}_n$	$\frac{1}{n}(n^2-3n+6)$	$n-3$	$\frac{n}{3}-1$	$\frac{1}{18}n(n-3)$
$\mathcal{L}_n$	$4-6/n$	$\binom{n-2}{2}$	$\frac{1}{3}\binom{n-2}{2}$	$\frac{1}{90}n(2n^3-5n^2-22n+60)$

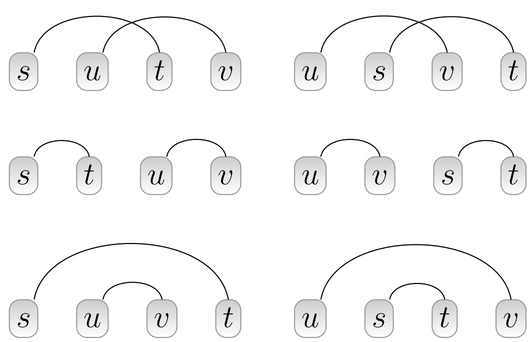


Figure 3: Linear arrangements of 4 different vertices forming two edges: $\{s,t\}$ and $\{u,v\}$.

3. The expected number of crossings of a uniformly random linear arrangement

The number of crossings of a graph for a concrete linear arrangement of its vertices can be defined as

$$C = \sum_{\{\{s,t\},\{u,v\}\} \in Q} C[\{s,t\},\{u,v\}]$$

where $C[\{s,t\},\{u,v\}]$ is an indicator variable such that $C[\{s,t\},\{u,v\}] = 1$ if the edges $\{s,t\}$ and $\{u,v\}$ cross and $C[\{s,t\},\{u,v\}] = 0$ otherwise. The expectation of C in a uniformly random linear arrangement of a graph is

$$\begin{aligned}\mathbb{E}\left[C\right] &= \mathbb{E}\left[\sum_{\{\{s,t\},\{u,v\}\}\in Q} C[\{s,t\},\{u,v\}]\right] \\ &= \sum_{\{\{s,t\},\{u,v\}\}\in Q} \mathbb{E}\left[C[\{s,t\},\{u,v\}] = 1 \mid \{\{s,t\},\{u,v\}\} \in Q\right]\end{aligned}$$

Assume without loss of generality that s precedes t and u precedes v in the linear arrangement. Then the edges formed by s and t and by u and v cross only in two relative orderings out of

$$\binom{4}{2}$$

as illustrated in figure 3. Therefore [7]

$$\mathbb{E}\left[C[\{s,t\},\{u,v\}] = 1 \mid \{\{s,t\},\{u,v\}\} \in Q\right] = \frac{2}{\binom{4}{2}} = \frac{1}{3}$$

and finally

$$\mathbb{E}\left[C\right] = \frac{|Q|}{3}. \tag{27}$$

In a complete graph, the number of crossings does not depend on the ordering of the vertices, therefore

$$\mathbb{E}\left[C\right] = C(\mathcal{K}_n).$$

Combining the previous equation with equation 16, we conclude that equation 27 also holds in complete graphs. Knowing $|Q|$ and applying equation 27, obtaining the value of $\mathbb{E}\left[C\right]$ for each of the special graphs considered in this article is straightforward given their already known values of $|Q|$ (table 1).

Combining equation 1 and equation 27 one recovers the value of $\mathbb{E}\left[C\right]$ that has been obtained previously for trees [7], i.e.

$$\mathbb{E}\left[C\right] = \frac{n}{6}\left[n - 1 - \langle k^2\rangle\right]. \tag{28}$$

3.1. The scaling of $\mathbb{E}\left[C\right]$ as function of n.

Figure 4 shows $\mathbb{E}\left[C\right]$ as a function of n for the special graphs where $\mathbb{E}\left[C\right]$ depends only on the number of vertices of the graph. Notice that $\mathbb{E}\left[C\right]$ is constant in complete graphs and in $\mathcal{S}_\lambda \oplus \mathbf{0}_{n-\lambda}$. In bipartite graphs, $|Q|$ depends on both n_1 and n_2. According to Table 1, $\mathbb{E}\left[C\right]$ is expected to scale as $\sim n^\gamma$, with $\gamma = 1$ for quasi-star trees and $\gamma = 2$ for the remainder of graphs in figure 4. Figure 4 validates the theoretical $\mathbb{E}\left[C\right]$ by means of numerical estimates.

4. The variance of the number of crossings of a uniformly random linear arrangement

By definition, $\mathbb{V}\left[C\right]$, the variance of C in a uniformly random linear arrangement, is

$$\mathbb{V}\left[C\right] = \mathbb{E}\left[(C - \mathbb{E}\left[C\right])^2\right].$$

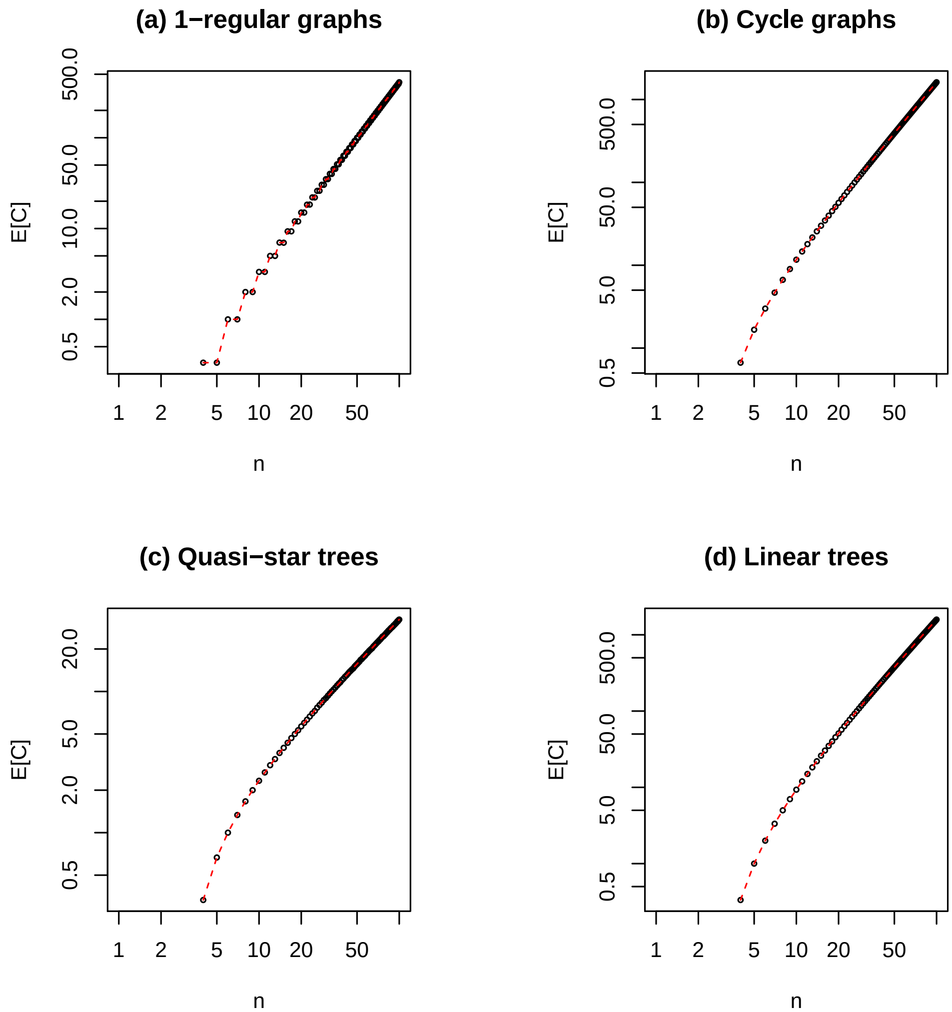


Figure 4: The mean of C, the number of crossings, as a function of n, the number of vertices of the graph. For every value n, the mean C is estimated over T different random linear arrangements ($T = n!$ for $n \leq 10$, then $T = 10^5$ for $n \geq 11$). When $T = n!$, the mean matches $\mathbb{E}\left[C\right]$. $\mathbb{E}\left[C\right]$, theoretical expectation according to Table 1 is also shown (red dashed line). As the scale is log-log, values with $n < 4$ are not shown because $E[C] = 0$.

We now present a derivation of C inspired by the derivation of the variance of C of the arrangement of the vertices of a complete graph on the surface of a sphere [20]. The number of crossings can be expressed in general as

$$C = \sum_{\{\{s,t\},\{u,v\}\} \in Q} \alpha(\{s,t\},\{u,v\}) \tag{29}$$

where $\alpha(\{s,t\},\{u,v\}) = 1$ if the edges $\{s,t\}$ and $\{u,v\}$ cross and $\alpha(\{s,t\},\{u,v\}) = 0$ otherwise. Expressing edges more compactly, we may equivalently write

$$C = \sum_{\{st,uv\}\in Q} \alpha(st,uv).$$

As $\mathbb{E}[C] = |Q|/3$, it is easy to see that

$$C - \mathbb{E}[C] = \sum_{\{st,uv\}\in Q} \beta(st,uv)$$

with

$$\beta(st,uv) = \alpha(st,uv) - 1/3,$$

and then

$$\mathbb{V}[C] = \mathbb{E}\left[\left(\sum_{\{st,uv\}\in Q} \beta(st,uv)\right)^2\right].$$

Expanding the previous expression, one finds that $\mathbb{V}[C]$ can be decomposed into a sum of $|Q \times Q| = |Q|^2$ summands of the form $\mathbb{E}[\beta(st,uv)\beta(wx,yz)]$, i.e.

$$\mathbb{V}[C] = \mathbb{E}\left[\sum_{\{st,uv\}\in Q}\sum_{\{wx,yz\}\in Q} \beta(st,uv)\beta(wx,yz)\right] \quad (30)$$

$$= \sum_{\{st,uv\}\in Q}\sum_{\{wx,yz\}\in Q} \mathbb{E}[\beta(st,uv)\beta(wx,yz)] \quad (31)$$

In general,

$$\mathbb{E}[\beta(st,uv)\beta(wx,yz)] = \mathbb{E}[\alpha(st,uv)\alpha(wx,yz)] - \frac{1}{3}\mathbb{E}[\alpha(st,uv)] - \frac{1}{3}\mathbb{E}[\alpha(wx,yz)] + \frac{1}{9}$$

As $\{st,uv\},\{wx,yz\} \in Q$,

$$\mathbb{E}[\alpha(st,uv)] = \mathbb{E}[\alpha(wx,yz)] = \frac{1}{3}.$$

Notice that $\mathbb{E}[\alpha(st,uv)] = 0$ if $\{st,uv\} \notin Q$. Then

$$\mathbb{E}[\beta(st,uv)\beta(wx,yz)] = \mathbb{E}[\alpha(st,uv)\alpha(wx,yz)] - \frac{1}{9} \quad (32)$$

As we will see, the analysis of the products $\beta(st,uv)\beta(wx,yz)$, or $\alpha(st,uv)\alpha(wx,yz)$, allows one to classify the combinations $(\{st,uv\},\{wx,yz\})$ in the double summation in equation 31 into 9 types.

4.1. The types of products

Suppose that $\eta = (\{e_1,e_2\},\{e_3,e_4\})$ is an element of $Q \times Q$. By definition, $\{e_1,e_2\},\{e_3,e_4\} \in Q$. The set of vertices of η is

$$\upsilon = \{e_1\} \cup \{e_2\} \cup \{e_3\} \cup \{e_4\}.$$

Table 2: The classification of the elements of $Q \times Q$ into types of products abstracting from the order of the elements of the pair, "code" is a meaningful but compact description of the type that result of from concatenating of τ and ϕ (except for types 6-7, where a third digit is required), $|v|$ is the number of different vertices of the type, τ is the size of the intersection between (e_1, e_2) and (e_3, e_4), ϕ is the number of edge intersections, p_i is the probability that $\alpha(e_1, e_2)\alpha(e_3, e_4) = 1$ for the i-th type in a uniformly random permutation of the vertices in V and $\mathbb{E}[\gamma_i] = p_i - 1/9$. We use the symbols s, t, u, v, w, x, y, z in column $(\{e_1, e_2\}, \{e_3, e_4\})$ to denote distinct vertices.

Type	Code	$\{\{e_1, e_2\}, \{e_3, e_4\}\}$	$\|v\|$	τ	ϕ	p_i	$\mathbb{E}[\gamma_i]$
0	00	$\{\{st, uv\}, \{wx, yz\}\}$	8	0	0	1/9	0
1	24	$\{\{st, uv\}, \{st, uv\}\}$	4	2	4	1/3	2/9
2	13	$\{\{st, uv\}, \{st, uw\}\}$	6	1	3	1/6	1/18
3	12	$\{\{st, uv\}, \{st, wx\}\}$	5	1	2	2/15	1/45
4	04	$\{\{st, uv\}, \{su, tv\}\}$	4	0	4	0	-1/9
5	03	$\{\{st, uv\}, \{su, vw\}\}$	5	0	3	1/12	-1/36
6	021	$\{\{st, uv\}, \{su, wx\}\}$	6	0	2	1/10	-1/90
7	022	$\{\{st, uv\}, \{sw, ux\}\}$	6	0	2	7/60	1/180
8	01	$\{\{st, uv\}, \{sw, xy\}\}$	7	0	1	1/9	0

One the one hand, the 4 edges of η contribute with at most 2 different vertices each. On the other hand, $\{e_1, e_2\}, \{e_3, e_4\} \in Q$ implies that $|e_1 \cap e_2| = |e_3 \cap e_4| = 0$. Then

$$4 \leq |v| \leq 8.$$

As a first approximation, we classify η based on two parameters. The first parameter is τ, the size of the intersection between the two sets of edges making an element of $Q \times Q$, i.e.

$$\tau = |\{e_1, e_2\} \cap \{e_3, e_4\}|.$$

The second parameter is ϕ, the number of edge intersections, i.e.

$$\phi = |e_1 \cap e_3| + |e_1 \cap e_4| + |e_2 \cap e_3| + |e_2 \cap e_4|. \tag{33}$$

Table 2 summarizes the 9 types of products, from 0 to 8. Type 0 and type 1 represent two extreme configurations: type 0 is the case where all the vertices are actually different while type 1 is the case where the pairs of edges are the same. Types 2-8 represent intermediate possibilities. Types 0 to 4 are found in the pioneering analysis by Moon [20] on complete graphs (type 0 is implicit in p. 506 and types 1-4 are enumerated as the types that have non-zero contribution in p. 506). Types 5-8 are novel.

Every combination of τ and ϕ yields a distinct type of product except $\tau = 0$ and $\phi = 2$, that yields two types (types 6 and 7). The latter follows from a further graph analysis that shows, in addition, the classification into 9 types is actually exhaustive. The analysis is based on an equivalence between $\eta = (\{e_1, e_2\}, \{e_3, e_4\})$ and a labeled weighted graph where

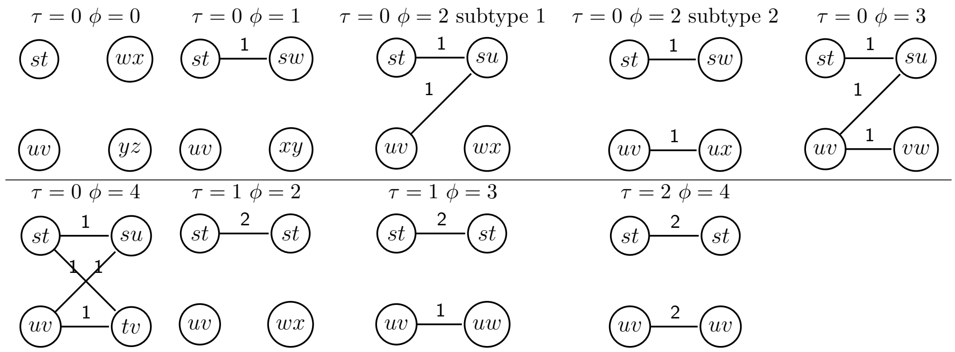


Figure 5: A summary of all the unlabeled weighted bipartite graphs produced by elements of $Q \times Q$, classified by τ and ϕ. In these bipartite graphs, vertices are edges in the original graph and edges indicate if two edges of the original graph are not independent. Edge weights indicate the number of vertices of the original graph shared by vertices of the bipartite graph.

(i) The set of vertices is e_1, e_2, e_3, e_4.

(ii) Two vertices e_i and e_j are linked if and only if $|e_i \cap e_j| > 0$.

(iii) The weight of an edge is $w(e_i, e_j) = |e_i \cap e_j|$. Therefore $w(e_i, e_j) \in \{1, 2\}$.

The graph is bipartite: one partition is $\{e_1, e_2\}$ and the other is $\{e_3, e_4\}$ (within each partition, edges cannot be formed because because $\{e_1, e_2\}, \{e_3, e_4\} \in Q$). We later formalize the type of η as a one to one mapping between the numbers 0 to 8 and the set of unlabeled bipartite graphs derived from all the labeled bipartite graphs produced by the elements of $Q \times Q$. All the unlabeled weighted bipartite graphs that can be produced by all the possible values of τ and ϕ are summarized in figure 5. In that figure, vertex labels are merely used to show that all these graphs are actually possible. The ensemble of labeled bipartite graphs that can be produced is, of course, larger. A detailed analysis follows.

By definition $0 \leq \tau \leq 2$. First, suppose that $\tau > 0$. Then, suppose without any loss of generality that $e_1 = e_3$. Then e_1 and e_3 are linked with a weight of 2. In turn, this implies that $|e_1 \cap e_4| = |e_2 \cap e_3| = 0$ because $\{e_1, e_2\}, \{e_3, e_4\} \in Q$. Now consider two cases:

- $\tau = 2$. Then $e_2 = e_4$ because $\tau = 2$ and e_2 and e_4 are linked with a weight of 2. Therefore, there is only one possible unlabeled bipartite graph and $\phi = 4$ as a result of plugging all the results above into the definition of ϕ (equation 33).
- $\tau = 1$. Then $2 \leq \phi$ and $|e_2 \cap e_4| < 2$, implying that there are only two possibilities: (1) e_2 and e_4 are unlinked with $\phi = 2$ or (2) linked with a weight of 1 and $\phi = 3$. Finally, we conclude that there is only one unlabeled bipartite graph for $\phi = 2$ and another one for $\phi = 3$.

Second, we consider the case $\tau = 0$. Then

$$0 \leq |e_1 \cap e_3|, |e_1 \cap e_4|, |e_2 \cap e_3|, |e_2 \cap e_4| \leq 1 \tag{34}$$

which in turn implies that

$$0 \leq \phi \leq 4. \quad (35)$$

To obtain $\phi = 0$, there is only one possibility, i.e.

$$|e_1 \cap e_3| = |e_1 \cap e_4| = |e_2 \cap e_3| = |e_2 \cap e_4| = 0. \quad (36)$$

Two obtain $\phi = 4$ (the complementary of the case $\phi = 0$), there is also only one possibility, i.e.

$$|e_1 \cap e_3| = |e_1 \cap e_4| = |e_2 \cap e_3| = |e_2 \cap e_4| = 1. \quad (37)$$

Therefore $\phi = 0$ and $\phi = 4$ produce only one unlabeled bipartite graph each. To obtain $\phi = 1$, we should have $|e_i \cap e_j| = 1$ only in one pair and $|e_i \cap e_j| = 0$ in the remainder. To obtain $\phi = 3$ (the complementary of the case $\phi = 1$), we should have $|e_i \cap e_j| = 0$ only in one pair and $|e_i \cap e_j| = 1$ in the remainder.
The interesting case is $\phi = 2$. Suppose without any loss of generality that $|e_1 \cap e_3| = 1$, namely the bipartite graph has an edge between e_1 and e_3 with a link of 1. We have to link an additional pair of edges to achieve $\phi = 2$. There are only three possibilities,

(i) $|e_1 \cap e_4| = 1$ (e.g. $(\{st, uv\}, \{sw, tx\})$).
(ii) $|e_2 \cap e_3| = 1$ (e.g. $(\{st, uv\}, \{su, wx\})$).
(iii) $|e_2 \cap e_4| = 1$ (e.g. $(\{st, uv\}, \{sw, ux\})$).

where s, t, u, v, w, x are all distinct. The 1st and the 2nd configurations are symmetric (one gives the other swapping the contents of the two partitions), namely they represent the same unlabeled bipartite graph. The third yields a different unlabeled bipartite graph (notice that the degree sequence of the 1st and 2nd configurations differ from that of the 3rd). See figure 5 for examples of the only two different unlabeled bipartite graphs.

As a result of the arguments above, every type can be meaningfully described by a code of two digits that results from concatenating τ and ϕ as shown in Table 2. For instance, type 2 has the code 13. The only exception is 02 that requires an additional digit to distinguish the two unlabeled bipartite graphs it can produce. A by-product of the analysis above is that it shows that $0 \leq \phi \leq 4$ for any possible value of τ.

4.2. The variance of C as the function of the number of products

Let $\mathcal{T}(\{e_1, e_2\}, \{e_2, e_3\})$ be the type of product of $\beta(e_1, e_2)\beta(e_3, e_4)$. If $\mathcal{T}(\{e_1, e_2\}, \{e_2, e_3\}) = i$, then $\gamma_i = \beta(e_1, e_2)\beta(e_3, e_4)$, allowing one to express $\mathbb{V}[C]$ (equation 31) equivalently as

$$\mathbb{V}[C] = \sum_{i=0}^{8} f_i \mathbb{E}[\gamma_i], \quad (38)$$

where f_i is the number of products of type i, defined as

$$f_i = \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q \\ \mathcal{T}(q_1, q_2) = i}} 1. \quad (39)$$

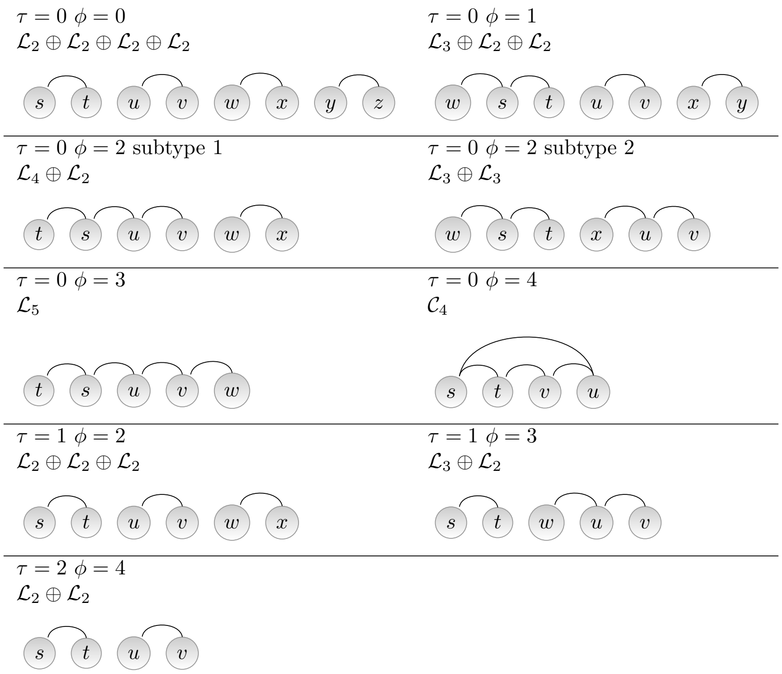


Figure 6: A summary of all the unipartite graphs generated by the bipartite graphs of figure 5, classified by τ and ϕ.

We can draw some initial conclusions from the analysis of the value of $|\upsilon_i|$ (the value of υ of type i) in table 2. For each type i, this value implies that

- $f_{00} = 0$ if $n < |\upsilon_{00}| = 8$.
- $f_{24} = f_{04} = 0$ if $n < |\upsilon_{24}| = |\upsilon_{04}| = 4$.
- $f_{13} = f_{021} = f_{022} = 0$ if $n < |\upsilon_{13}| = |\upsilon_{021}| = |\upsilon_{022}| = 6$.
- $f_{12} = f_{03}$ if $n < |\upsilon_{12}| = |\upsilon_{03}| = 5$.
- $f_{01} = 0$ if $n < |\upsilon_{01}| = 7$.

Furthermore, given the definition of the f_i's in equation 39 it is easy to see that any pair $\{q_1, q_2\}$, such that $q_1, q_2 \in Q$ with $q_1 \neq q_2$ is counted twice. In contrast, pairs such that $q_1 = q_2$ (i.e. type 24) are counted only once. Therefore, f_i is even for $i \neq 24$. It is easy to see that $f_{24} = |Q|$ in general, and that $f_{04} = 0$ in trees because the edges st, uv, su, tv define a cycle graph of 4 vertices.

Also, we have, by definition,

$$\sum_{i=0}^{8} f_i = |Q|^2. \qquad (40)$$

Moreover, the unipartite graphs generated by the bipartite graphs in figure 5 allow one to identify necessary conditions for finding a given type in a graph. These uniparte graph allow one to identify necessary conditions for finding a given type in a graph. For instance, as the uniparte graph of type 04 is a cycle graph of four vertices, a cycle is needed for $f_{04} > 0$. Then $f_{04} = 0$ in trees. Types 03 and 04 need paths of 4 vertices so that $f_{03}, f_{04} > 0$. Type 021 needs paths of 3 vertices so that $f_{021} > 0$. Types 022, 01 and 13 need paths of 2 vertices so that $f_{022}, f_{01}, f_{13} > 0$.

If the product $\alpha(e_1, e_2)\alpha(e_3, e_4)$ corresponds to a type i, we can replace $\mathbb{E}[\alpha(e_1, e_2)\alpha(e_3, e_4)]$ by p_i. p_i can be calculated exactly and easily be means of a computational procedure: p_i is the proportion of permutations of the vertices of v where $\alpha(..., ...)\alpha(..., ...) = 1$. The different values of p_i are summarized in Table 2. Some values are easy to predict. $p_0 = p_{00} = 1/9$ because

$$\mathbb{E}[\alpha(st, uv)\alpha(wx, yz)] = \mathbb{E}[\alpha(st, uv)]\,\mathbb{E}[\alpha(st, uv)] \tag{41}$$

and $\mathbb{E}[\alpha(st, uv)] = 1/3$. Notice also that

$$p_4 = p_{04} = 0$$

because a $C_{max} \leq 1$ for any graph of 4 vertices whose vertices are projected into a sequence of length 4 (equation 10 gives $C = 1$ for a complete graph). Projecting the vertices on a longer sequence does not change the conclusion.

In general, equation 32 gives

$$\begin{aligned}\mathbb{E}[\gamma_i] &= \mathbb{E}[\alpha(e_1, e_2)\alpha(e_3, e_4)] - \frac{1}{9} \\ &= p_i - \frac{1}{9}.\end{aligned}$$

The values of $\mathbb{E}[\gamma_i]$ are summarized in Table 2.

Applying the values of $\mathbb{E}[\gamma_i]$ in Table 2 to equation 38, allows one to express $\mathbb{V}[C]$ as a function of the amount of times every product appears, namely

$$\begin{aligned}\mathbb{V}[C] &= \sum_{i=1}^{7} f_i \mathbb{E}[\gamma_i] \\ &= \frac{1}{9}\left[2|Q| + \frac{1}{20}f_{022} + \frac{1}{5}f_{12} + \frac{1}{2}f_{13} - \left(\frac{1}{10}f_{021} + f_{04} + \frac{1}{4}f_{03}\right)\right].\end{aligned} \tag{42}$$

In a tree, $f_{04} = 0$ and then

$$\mathbb{V}[C] = \frac{1}{9}\left[2|Q| + \frac{1}{20}f_{022} + \frac{1}{5}f_{12} + \frac{1}{2}f_{13} - \left(\frac{1}{10}f_{021} + \frac{1}{4}f_{03}\right)\right] \tag{43}$$

with

$$f_{00} + f_{01} + f_{021} + f_{022} + f_{03} + f_{12} + f_{13} = |Q|(|Q| - 1). \tag{44}$$

In the coming sections, we derive expressions for the $f_i's$ in a simple graph and also present compact formulae for particular kinds of graphs that allow one to obtain compact formulae for $\mathbb{V}[C]$ via equation 42. Before we proceed, we give a chance to the reader to check detailed counts of the $f_i's$ and the calculation of $\mathbb{V}[C]$ for small graphs (Appendix B). They can help gain intuitions for the mathematical calculations to follow. In addition, these examples are component of the protocol to validate the formulae that will be derived (Appendix C).

4.3. Preliminaries

Before moving on to formalizing each type f_i and deriving general expressions for them, we first define the notation used.

We define A^x as the x-th power of the adjacency matrix A. $a_{st}^{(x)}$ represents the element (s,t) of the x-th power of the adjacency matrix and m_x the sum of one of half of the values of A^x excluding the diagonal, i.e.

$$m_x = \sum_{s<t} a_{st}^{(x)}.$$

where $m_1 = m$.

For any undirected simple graph $G = (V, E)$, let G_{-s} be the induced graph resulting from removing vertex s from G. More generally, we define $G_{-\mathcal{L}}$ as the induced graph resulting from the removal of the vertices in $\mathcal{L} \subseteq V$. Unless stated otherwise, we use $Q = Q(G)$, to refer to the set of pairs of independent edges of G.

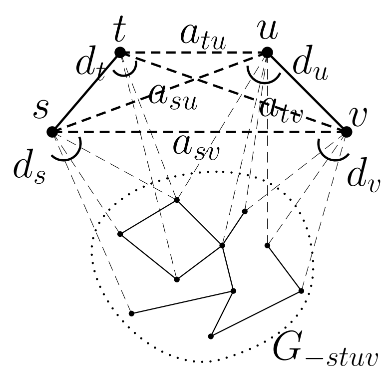


Figure 7: The effect of removing vertices s, t, u, v such that $\{st, uv\} \in Q$ from a graph G to produce the graph G_{-stuv}. Here d_s, d_t, d_u, d_v denote the degree of s, t, u, v in G without taking into account their neighbors equal to any of s, t, u, v.

Since it is used extensively, we denote $G_{-\{s,t,u,v\}}$ as simply G_{-stuv}. The number of edges in a graph G_{-stuv} is easy to calculate as a function of the number of edges in G. If s, t, u, v are four distinct vertices, then

$$|E(G_{-stuv})| = m - (k_s + k_t + k_u + k_v) + a_{st} + a_{su} + a_{sv} + a_{tu} + a_{tv} + a_{uv}. \quad (45)$$

By default, network features refer to G. Therefore, m, k_u, a_{uv}, ... in equation 45 refer to G. We use $\Gamma(s, -L) = \Gamma(s) \setminus L$ to denote the set of neighbors of $s \in L \subseteq V$ in $V(G_{-L})$. Notice that its size is:

$$|\Gamma(s, -L)| = k_l - |\{w \in L \mid \{w, l\} \in E\}| = k_l - \sum_{w \in L} a_{lw}. \quad (46)$$

Since $\Gamma(k, -\{s,t,u,v\})$, with $k \in \{s,t,u,v\}$, is used extensively in this work, we use $\Gamma(k, -stuv)$ instead. We use the shorthand "n-path" to refer to paths of n vertices. Finally, we use $n_G(F)$ to denote the number of subgraphs isomorphic to F in G.

The calculations of $f_{00}, f_{01}, ..., f_{24}$ to come, require a clear notation that states the vertices shared between each pair of elements of Q for an arbitrary graph G. Throughout this article, we need to use summations of the form

$$\sum_{\substack{s,t,u,v \in V:\\ \{st,uv\} \in Q(G)}} \sum_{\substack{w,x,y,z \in V:\\ \{wx,yz\} \in Q(G_{-\{s,t,u,v\}})}} \Box,$$

where below each summation operand there is a scope on top of a condition. The "$\Box$" represents any term. For the sake of brevity, we contract them as

$$\sum_{\{st,uv\} \in Q} \sum_{\{wx,yz\} \in Q(G_{-stuv})} \Box.$$

Notice that the scope is omitted in the new notation. This detail is crucial for the countings performed with the help of these compact summations. Likewise, if we want to denote when two elements of Q from each of the summations share one or more vertices, we use:

$$\sum_{\{st,uv\} \in Q} \sum_{\{sx,tz\} \in Q(G_{-uv})} \Box = \sum_{\substack{s,t,u,v \in V:\\ \{st,uv\} \in Q(G)}} \sum_{\substack{x,z \in V:\\ \{sx,tz\} \in Q(G_{-\{u,v\}})}} \Box.$$

This expression denotes the summation over the pairs of elements of Q in which the second one shares two vertices with the first one. Again, the expression to the left is a shorthand for the one to the right. For the sake of comprehensiveness, we also present two more compact definitions:

$$\sum_{\{st,uv\} \in Q} \sum_{\{st,yz\} \in Q(G_{-uv})} \Box = \sum_{\substack{s,t,u,v \in V:\\ \{st,uv\} \in Q(G)}} \sum_{\substack{y,z \in V:\\ \{st,yz\} \in Q(G_{-\{u,v\}})}} \Box,$$

$$\sum_{\{st,uv\} \in Q} \sum_{\{sv,yz\} \in Q(G_{-tu})} \Box = \sum_{\substack{s,t,u,v \in V:\\ \{st,uv\} \in Q(G)}} \sum_{\substack{y,z \in V:\\ \{sv,yz\} \in Q(G_{-tu})}} \Box.$$

4.4. General results

Before deriving general expressions for the amount of elements of $Q \times Q$ that exist of each type in section 4.5, we need to introduce some useful results, which prove extremely useful when simplifying the expressions in that section.

The first relates the amount of $\mathcal{L}_4$ and the elements of Q.

Proposition 4.1. *The number of subgraphs isomorphic to $\mathcal{L}_4$, namely half the amount of paths of 4 vertices, in a graph G is*

$$n_G(\mathcal{L}_4) = \sum_{\{st,uv\} \in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv}) \quad (47)$$

$$= \frac{1}{2} \sum_{s=1}^{n} \sum_{\substack{t=1\\ t \neq s}}^{n} (a_{st}^{(3)} - a_{st}(2k_t - 1)) \quad (48)$$

$$= m_3 + m_1 - n\langle k^2 \rangle. \quad (49)$$

Proof. Let $\{st, uv\} \in Q$ be any element of Q. The expression $a_{su} + a_{sv} + a_{tu} + a_{tv}$ counts how many $\mathcal{L}_4$ we can make with these two edges. This is trivially true: since $\{st, uv\}$ is an element of Q then we know that $a_{st} = a_{uv} = 1$. To these two edges we only have to add one of the four edges in the summation (i.e., any of the edges that connect a vertex of st with another vertex of uv) to make a $\mathcal{L}_4$. By adding one at a time we always obtain a different $\mathcal{L}_4$. Therefore, if one edge exists then we can make a $\mathcal{L}_4$ which will be different from the other three potential $\mathcal{L}_4$ that we can make (figure 8). This means that, when counting how many of these edges exist, we are actually counting how many $\mathcal{L}_4$ exist.

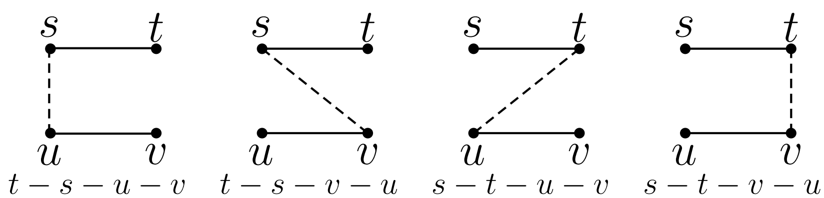


Figure 8: The four different $\mathcal{L}_4$ we can make with $\{st, uv\} \in Q$.

Let $\mathcal{L}_4(q_1)$ be the set of $\mathcal{L}_4$ into which $q_1 \in Q$ is mapped. The claim that the whole summation in equation 47 counts all $\mathcal{L}_4$ could be false for two reasons:

(i) Some path is not counted. Suppose that there exists a $\mathcal{L}_4$ with vertices $s-t-u-v$ in G that is not counted in the summation. If this was true then the pair of independent edges we can make using its vertices ($\{st, uv\}$) would not be in Q. But this cannot happen by definition of Q.

(ii) Some $\mathcal{L}_4$ is counted more than once. It can only happen when another element $q_2 \in Q$, $q_1 \neq q_2$, is mapped to a set of $\mathcal{L}_4$ so that at least one of them is also in $\mathcal{L}_4(q_1)$. Formally: it happens when there exists a $q_2 \in Q$ such that $\mathcal{L}_4(q_1) \cap \mathcal{L}_4(q_2) \neq \emptyset$. These potential q_2 need to have the same vertices as q_1. If $q_1 = \{st, uv\}$ then this implies that $q_2 = \{su, tv\}$ or $q_2 = \{sv, tu\}$. However, in $\mathcal{L}_4(q_2)$ there are no $\mathcal{L}_4$ that are also in $\mathcal{L}_4(q_1)$ (figure 8) because the $\mathcal{L}_4$ we can make with either possibility for q_2 are all different from those in $\mathcal{L}_4(q_1)$ (figure 9).

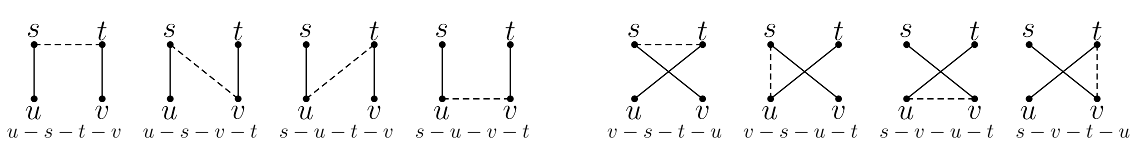


Figure 9: The two sets of $\mathcal{L}_4$ we can make with $\{su, tv\} \in Q$ (left) and $\{sv, tu\} \in Q$ (right).

Equation 48 follows from [30] and can be expressed as

$$\frac{1}{2} \sum_{s \neq t} (a_{st}^{(3)} - (2k_t - 1)a_{st}) = m_3 + m_1 - \sum_{s \neq t} a_{st} k_t = m_3 + m_1 - n\langle k^2 \rangle,$$

namely equation 49. Recall m_1 is the number of edges of the graph. □

Furthermore, the second result relates the amount of $\mathcal{L}_5$ and the elements in Q.

Proposition 4.2. *The number of subgraphs isomorphic to $\mathcal{L}_5$, namely half the amount of paths of 5 vertices, in a graph G is*

$$n_G(\mathcal{L}_5) = \sum_{\{st,uv\}\in Q} \left(\sum_{w_s\in\Gamma(s,-stuv)} (a_{uw_s} + a_{vw_s}) + \sum_{w_t\in\Gamma(t,-stuv)} (a_{uw_t} + a_{vw_t}) \right). \quad (50)$$

Proof. The proof is similar to the proof of proposition 4.1. For a fixed element $\{st, uv\} \in Q$ the inner summations of equation 50, i.e.

$$\sum_{w_s\in\Gamma(s,-stuv)} (a_{uw_s} + a_{vw_s}) + \sum_{w_t\in\Gamma(t,-stuv)} (a_{uw_t} + a_{vw_t}),$$

count the number of $\mathcal{L}_5$ we can make with edges st and uv such that $q = \{st, uv\} \in Q$. This set of $\mathcal{L}_5$ is denoted as $\mathcal{L}_5(q)$ and contains the $\mathcal{L}_5$ that follow a concrete pattern: they have at one end the vertices of one of the edges of q, a vertex w as its centroid, and have the vertices of the other edge of q at the other end. Vertex w is such that $w \neq s, t, u, v$ and is adjacent to its "previous" and "next" vertices in the $\mathcal{L}_5$. For example, $\mathcal{L}_5(q)$ may contain $t - s - w - u - v$ if $a_{sw} = a_{wu} = 1$. Therefore, the graphs $\mathcal{L}_5(q)$, for any $q \in Q$, have 4 different forms, with the vertices of q "at the beginning" and "at the end", and a vertex different from its vertices in the middle (figure 10).

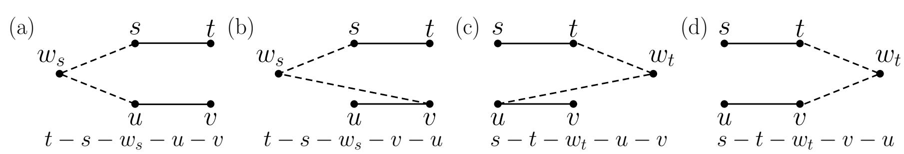


Figure 10: The four different families of $\mathcal{L}_5$ we can make with two edges st and uv such that $q = \{st, uv\} \in Q$, namely $\mathcal{L}_5(q)$. (a) w_s is a neighbor of s and u, (b) w_s is a neighbor of s and v, (c) w_t is a neighbor of t and u, (d) w_t is a neighbor of t and v.

Similarly, equation 50 could be wrong for two reasons: some path may be counted more than once or not counted at all.

All $\mathcal{L}_5$ are counted: by contradiction, given a fixed element $\{st, uv\} \in Q$, suppose that there is a $\mathcal{L}_5$, $s - t - w - u - v$, not counted in the inner summation. By definition of $\mathcal{L}_5$, $a_{st} = a_{tw} = a_{wu} = a_{uv} = 1$, the vertices are distinct and then the pairs of edges $\{st, wu\}$, $\{tw, uv\}$, $\{st, uv\} \in Q$. Therefore, if such $\mathcal{L}_5$ is not counted then none of the elements $\{st, wu\}$, $\{tw, uv\}$, $\{st, uv\}$ would be in Q.

Some $\mathcal{L}_5$ may be counted more than once: if this was true then for some $q_1 = \{st, uv\} \in Q$ there would exist a $q_2 \in Q$, $q_1 \neq q_2$, such that $\mathcal{L}_5(q_1) \cap \mathcal{L}_5(q_2) \neq \emptyset$. For this to happen, q_2 must be built upon the vertices s, t, u, v, w. These potential q_2 are constructed in two steps. Firstly, choose one of the subsets of 4 vertices out of the $\binom{5}{4} = 5$ possibilities for the set of vertices of q_2. Secondly, form one of the $\binom{4}{2}/2 = 3$ unordered pair of edges using the four vertices selected in the previous step. This gives a total of 15 possibilities. Since s, t, u, v, w are all distinct then any subset of size 4 also contains distinct vertices so any two disjoint pairs out of these 4 vertices make an element of Q. Table 3 contains all such combinations, and each of them is one possible value for q_2.

Table 3: All 15 elements of Q that can be made out of 5 distinct vertices s, t, u, v, w.

s,t,u,v	s,t,u,w	s,t,v,w	s,u,v,w	t,u,v,w
st, uv	st, uw	st, vw	su, vw	tu, vw
su, tv	su, tw	sv, tw	sv, uw	tv, uw
sv, tu	sw, tu	sw, tv	sw, uv	tw, uv

None of these combinations for q_2 are such that $\mathcal{L}_5(q_1) \cap \mathcal{L}_5(q_2) \neq \emptyset$, except the only combination that is equal to q_1 (which contradicts the assumption that $q_1 \neq q_2$). That is true because the combinations with st, or ts at one end of the $\mathcal{L}_5$, do not have uv, or vu at the other end, and the same applies to the reverse (these add up a total of 4 combinations). The other combinations always have one vertex of one edge attached to a vertex of the other edge, i.e., su, sv, tu, or tv (these add up to a total of 10 combinations), and these pairs are not at neither end of the $\mathcal{L}_5$'s in $\mathcal{L}_5(q_1)$. Therefore none of the $\mathcal{L}_5$ in $\mathcal{L}_5(q_2)$ are in $\mathcal{L}_5(q_1)$.

As a conclusion, all paths are counted, and no path is counted more than once. Therefore, equation 50 evaluates to exactly all $\mathcal{L}_5$ in G. □

The next proposition relates the number of cycles of four vertices in a graph with its set of independent edges Q:

Proposition 4.3. *The number of cycles of* 4 *vertices, namely* $\mathcal{C}_4$*, in a graph* G *is*

$$n_G(\mathcal{C}_4) = \frac{1}{2} \cdot \sum_{\{st,uv\} \in Q} (a_{sv}a_{ut} + a_{su}a_{tv}) \tag{51}$$

$$= \frac{1}{8}\left[tr(A^4) - 4n_G(\mathcal{L}_3) - 2n_G(\mathcal{L}_2)\right] \tag{52}$$

$$= \frac{1}{8}\left[tr(A^4) + 4|Q| - 2m^2\right]. \tag{53}$$

Proof. For each element $\{st, uv\} \in Q$ we can potentially make up to two different cycles:

- Cycle (s, t, v, u, s) if $a_{su}a_{tv} = 1$‡.
- Cycle (s, t, u, v, s) if $a_{tu}a_{sv} = 1$§.

‡ Notice that the cycles (t, v, u, s, t), (v, u, s, t, v), (u, s, t, v, u) are the same as cycle (s, t, v, u, s).
§ Notice that the cycles (t, u, v, s, t), (u, v, s, t, u), (v, s, t, u, v) are the same as cycle (s, t, u, v, s).

Notice that for every element $\{st, uv\} \in Q$ that forms a cycle, say (s, t, v, u, s), we have that $a_{su}a_{tv} = 1$ and thus we have another element $\{su, tv\} \in Q$.

For this other element $\{su, tv\} \in Q$ we can make the cycle (s, u, v, t, s) because $st \in E$ and $uv \in E$ (because we assumed that $\{st, uv\} \in Q$), which is isomorphic to (s, t, v, u, s). The same reasoning can be applied to the second potential cycle we can make. Therefore, each cycle of length 4 can be made from two different elements in Q. Finally, all cycles are counted in the summation: if there was some uncounted cycle then none of the two pairs of independent edges we can make with said cycle would be in Q and this cannot happen. To conclude, all cycles of 4 vertices are counted, but each of them is counted twice in the summation, hence the 1/2.

Equations 52 and 53 follow from previous results in [31, 32]. Suppose that $n_G(\mathcal{C}_4)$ is the number of different cycles of length 4 that are contained in G [31]. More technically, $n_G(\mathcal{C}_4)$ is the number of subgraphs of G that are isomorphic to $\mathcal{C}_4$.

We have that [31]

$$n_G(\mathcal{C}_4) = \frac{1}{8}\left[tr(A^4) - 4n_G(H_2) - 2n_G(H_1)\right] \tag{54}$$

with

$$n_G(H_1) = m, \qquad n_G(H_2) = \sum_{s=1}^{n} \binom{k_s}{2}.$$

A similar formula for $n_G(\mathcal{C}_4)$ was derived in the pioneering research by Harary and Manvel [32]. The fact that $n_G(H_1) = n_G(\mathcal{L}_2)$ and $n_G(H_2) = n_G(\mathcal{L}_3)$ [31], transforms equation 54 into equation 52. Recalling the definition of $|Q|$ in [17] (equation 12), we may write $n_G(\mathcal{C}_4)$ equivalently as

$$n_G(\mathcal{C}_4) = \frac{1}{8}\left[tr(A^4) + 4|Q| - 2m^2\right].$$

□

There is another useful result regarding the sum of the degrees of all vertices involved in the elements in Q:

Proposition 4.4. *The sum of the degrees of the vertices of every pair of independent edges in a graph G is*

$$K_G = \sum_{\{st,uv\}\in Q} (k_s + k_t + k_u + k_v) = n[(m+1)\langle k^2\rangle - \langle k^3\rangle] - 2\sum_{st\in E} k_s k_t$$

Proof. The proof is straightforward:

$$\begin{aligned} K_G &= \sum_{\{st,uv\}\in Q} (k_s + k_t + k_u + k_v) \\ &= \sum_{\{st,uv\}\in Q} (k_s + k_t) + \sum_{\{st,uv\}\in Q} (k_u + k_v) \\ &= \frac{1}{2}\left[\sum_{st\in E}\sum_{uv\in E(G_{-st})} (k_s + k_t) + \sum_{uv\in E}\sum_{st\in E(G_{-uv})} (k_u + k_v)\right] \\ &= \sum_{st\in E}(k_s + k_t) \sum_{uv\in E(G_{-st})} 1. \end{aligned}$$

Recalling the definition of $q(s,t)$, the number of edges that the may potentially cross the edge formed by s and t (equation 19), one obtains

$$\begin{aligned} K_G &= \sum_{st\in E}(k_s + k_t)q(s,t) \\ &= \sum_{st\in E}(k_s + k_t)(m + 1 - k_s - k_t) \\ &= n[(m+1)\langle k^2\rangle - \langle k^3\rangle] - 2\sum_{st\in E} k_s k_t. \end{aligned}$$

□

Notice $q(s,t) = |E(G_{-st})|$. Notice also that the term $\sum_{st\in E} k_s k_t$ is a sum of the product of the degrees at both ends of an edge, needed for the calculation of degree correlations [33].

4.5. Theoretical formulae

In the following subsections we formalize the number of products of each type for an arbitrary graph providing general expressions that are to be considered a first approach. The general formulas for the f_i's presented in this section are designed based on three non-exclusive principles: compactness, linking with standard graph theory and linking with the recently emerging subfield of crossing theory (section 2 and also [14]). The link with graph theory will be established showing that

$$f_i = a_i n_G(...), \tag{55}$$

where a_i is an even natural number (except $a_{24} = 1$) and "..." is some subgraph that can be some some elementary graph (a linear tree or a cycle graph) or a combinations of them with the operator $\oplus$, denoting independent union of graphs, that depend on i. The subgraph for each type of product is found in figure 6. The link with crossing theory will be established deriving alternative expressions for the f_i's that are a function of $|Q|$ and other network properties. An overview of the two types of expression that will be derived for the f_i's, is shown in table 4.

Table 4: Expressions for the f_i's linking them to graph theory (middle column), and some of the expressions as a function of $|Q|$ and other network features (some of the expressions are not included due to lack of space).

type	$f_i = a_i n_G(...)$	f_i	equation
00	$6n_G(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2)$	$\sum_{\{st,uv\}\in Q} \lvert Q(G_{-stuv})\rvert$	59
24	$n_G(\mathcal{L}_2 \oplus \mathcal{L}_2)$	$\lvert Q\rvert$	56
13	$2n_G(\mathcal{L}_3 \oplus \mathcal{L}_2)$	$K_G - 4\lvert Q\rvert - 2n_G(\mathcal{L}_4)$	63
12	$6n_G(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2)$	$2[(m+2)\lvert Q\rvert + n_G(\mathcal{L}_4) - K_G]$	67
04	$2n_G(\mathcal{C}_4)$	$\frac{1}{4}tr(A^4) - n_G(\mathcal{L}_3) - \frac{1}{2}n_G(\mathcal{L}_2)$	
		$\frac{1}{4}tr(A^4) + \lvert Q\rvert - \frac{1}{2}m^2$	70
03	$2n_G(\mathcal{L}_5)$		74, 76
021	$2n_G(\mathcal{L}_4 \oplus \mathcal{L}_2)$		84
022	$4n_G(\mathcal{L}_3 \oplus \mathcal{L}_3)$		89
01	$4n_G(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2)$	$\sum_{\{st,uv\}\in Q} \sum_{k\in\{s,t,u,v\}} \sum_{w\in\Gamma(k,-stuv)} \lvert E(G_{-stuvw})\rvert$	93

For simplicity, we focus our work on the f_i's that actually contribute to $\mathbb{V}[C]$, namely f_i with $i \notin \{00, 01\}$ because $\mathbb{E}[\gamma_{00}] = \mathbb{E}[\gamma_{01}] = 0$.

We will follow two approaches for the calculations of the $f_i's$ in the coming subsections. The first one consists of instantiating equation 39 with the help of table 2 and figure 5. The second approach consists of producing an equation of the form of equation 55 by means of figure 6 and further information inherited from the first approach. All the initial definitions of the $f_i's$ that will follow stem from equation 39 although it is only mentioned for the first types so as not to become too repetitive.

4.5.1. $\tau = 2$, $\phi = 4$ Thanks to equation 39,

$$f_{24} = \sum_{\{st,uv\} \in Q} \sum_{\substack{w,x,y,z \in V: \\ w=s,x=t,y=u,z=v \\ \{wx,yz\} \in Q}} 1 = \sum_{\{st,uv\} \in Q} 1 = |Q|. \tag{56}$$

Then f_{24} can be calculated easily thanks to the definition of $|Q|$ in equation 13. By definition of Q,

$$f_{24} = n_G(\mathcal{L}_2 \oplus \mathcal{L}_2). \tag{57}$$

4.5.2. $\tau = 0$, $\phi = 0$ Thanks to equation 39,

$$f_{00} = \sum_{\{st,uv\} \in Q} \sum_{\{wx,yz\} \in Q(G_{-stuv})} 1. \tag{58}$$

Noting that the inner summation defines the size of the set of pairs of independent edges of G_{-stuv}, the expression above can be simplified, and then we obtain

$$f_{00} = \sum_{\{st,uv\} \in Q} |Q(G_{-stuv})|. \tag{59}$$

The last result comes to say that for any element $\{st, uv\} \in Q$ we only need to calculate the size of $Q(G_{-stuv})$, the set of pairs of independent edges of G_{-stuv}.

The summation in equation 59 counts over combinations of a pair of edges $\{st, uv\}$ from Q with any two other independent edges $\{wx, yz\}$, defining a set $H = \{st, uv, wx, yz\}$ of independent edges. Each of these sets defines some subgraph $\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$. Every distinct set H is produced by

$$\binom{|H|}{2} = 6$$

elements of Q. Therefore, equation 59 gives

$$f_{00} = 6n_G(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2). \tag{60}$$

4.5.3. $\tau = 1$, $\phi = 3$ This type deals with the pairs of edges sharing exactly one edge ($\tau = 1$) and that have 3 vertices in common ($\phi = 3$). Via equation 39, a possible formalization of f_{13} is

$$f_{13} = \sum_{\{st,uv\} \in Q} \left(\sum_{\{st,uw\} \in Q(G_{-v})} 1 + \sum_{\{st,vw\} \in Q(G_{-u})} 1 + \sum_{\{uv,sw\} \in Q(G_{-t})} 1 + \sum_{\{uv,tw\} \in Q(G_{-s})} 1 \right). \quad (61)$$

The first inner summation in the previous equation denotes the amount of vertices neighbors of u in G that are not s, t, v, the second the amount of vertices neighbors of v in G that are not s, t, u, and so on (figure 11 for an illustration). In that figure, and similar figures to follow, solid thick lines indicate existing edges, solid thin lines indicate real edges of a hypothetical graph and dashed lines indicate potential edges, namely edges that may not exist.

A formal definition for the first inner summation is, given a fixed $\{st, uv\} \in Q$,

$$\sum_{\{st,uw\} \in Q(G_{-v})} 1 = |\Gamma(u, -stuv)|.$$

Likewise for the other inner summations. Then, equation 62 becomes

$$f_{13} = \sum_{\{st,uv\} \in Q} |\Gamma(u, -stuv)| + \sum_{\{st,uv\} \in Q} |\Gamma(v, -stuv)| + \sum_{\{st,uv\} \in Q} |\Gamma(s, -stuv)| + \sum_{\{st,uv\} \in Q} |\Gamma(t, -stuv)|. \quad (62)$$

Applying

$$|\Gamma(u, -stuv)| = k_u - (a_{uv} + a_{us} + a_{ut})$$

equation 62, becomes (recall that $\{st, uv\} \in Q$ implies $a_{st} = a_{uv} = 1$):

$$f_{13} = \sum_{\{st,uv\} \in Q} (k_s + k_t + k_u + k_v) - 4|Q| - 2 \sum_{\{st,uv\} \in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv}).$$

Applying propositions 4.1 and 4.4 we can rewrite the previous result as

$$f_{13} = K_G - 4|Q| - 2n_G(\mathcal{L}_4). \quad (63)$$

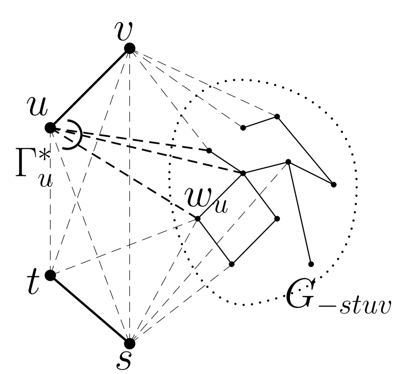


Figure 11: Illustration of the first inner summation in equation 62. In the figure, $\{st, uv\} \in Q$, and $w_u \in \Gamma_u^* = \Gamma(u, -stuv)$.

f_{13} counts over subgraphs $\mathcal{L}_3 \oplus \mathcal{L}_2$. This can be seen by analyzing the summations of equation 62: (1) the first summation counts the number of combinations of a 2-path (s,t) with a 3-path (w,u,v) , (2) the second summation counts the number of combinations of a 2-path (s,t) with a 3-path (u,v,w) , (3) the third summation counts the number of combinations of a 2-path (u,v) with a 3-path (w,s,t) , and (4) the fourth summation counts the number of combinations of a 2-path (u,v) with a 3-path (s,t,w). Figure 11 shows an example of summation (1): the edge st is independent of all $\mathcal{L}_3$ of the form (v,u,w_u), where $w_u \in V(G_{-stuv})$. This can also be seen in the corresponding unipartite graph in figure 6. Then, since summands (1) and (2) count the same subgraphs as summands (3) and (4) (the first and the second summand give the third and the fourth summand exchanging s and t by u and v) we have that

$$f_{13} = 2n_G(\mathcal{L}_3 \oplus \mathcal{L}_2). \tag{64}$$

We outline an alternative argument that leads to the same conclusion. Assume $\{(s,t),(u,v,w)\}$, a $\mathcal{L}_3 \oplus \mathcal{L}_2$, is a sugraph of G. Then, the only elements of $Q \times Q$ classified as type 13 with these vertices are $(\{st,uv\},\{st,vw\})$ and its reverse. It is easy to see that that there are other elements of $Q \times Q$ of type 13 with the same vertices but they do not correspond to $\{(s,t),(u,v,w)\}$. Therefore, equation 39 counts two elements of $Q \times Q$ for a single $\mathcal{L}_3 \oplus \mathcal{L}_2$. This argumentation is also used in some of the types to follow.

4.5.4. $\tau = 1$, $\phi = 2$ In this type, a type 13, one edge is shared, but this time only two vertices are equal. Therefore, we can formalize f_{12} as

$$f_{12} = \sum_{\{st,uv\}\in Q} \left(\sum_{\{st,wx\}\in Q(G_{-uv})} 1 + \sum_{\{uv,wx\}\in Q(G_{-st})} 1 \right). \tag{65}$$

There is a simpler expression for this type that holds for any graph. Each summation inside the expression above counts the amount of pairs of vertices different from s,t,u,v. Therefore, each summation counts the amount of edges of G_{-stuv}. Therefore, the expression in equation 65 can be simplified as

$$f_{12} = 2 \sum_{\{st,uv\}\in Q} |E(G_{-stuv})|. \tag{66}$$

Applying equation 45, we can rewrite the previous expression equivalently as

$$f_{12} = 2 \left(|Q|(m+2) + \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv}) - \sum_{\{st,uv\}\in Q} (k_s + k_t + k_u + k_v) \right)$$

which, thanks to propositions 4.1 and 4.4, leads to

$$f_{12} = 2[(m+2)|Q| + n_G(\mathcal{L}_4) - K_G], \tag{67}$$

The factor of 2 indicates explicitly that f_{12} is an even number as expected (section 4.2).

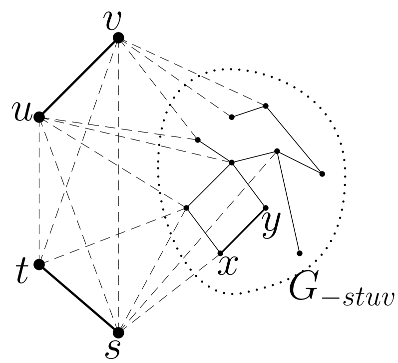


Figure 12: Illustration of the summations in equation 65, the first of which represents the elements in the Cartesian product of edge st and the edges in G_{-stuv}.

In equation 66, the summation is counting over configurations that are produced combining two edges from Q with a third independent edge, giving three independent edges, i.e. (s,t), (u,v), (w,x). The set containing these three edges, i.e. $H = \{(s,t),(u,v),(w,x)\}$ defines some subgraph $\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$. The summation in equation 66 visits the same set H

$$\binom{|H|}{2} = 3$$

times. Therefore, the summation in equation 66 matches $3n_G(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2)$ and then

$$f_{12} = 6n_G(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2). \tag{68}$$

4.5.5. $\tau = 0$, $\phi = 4$ All pairs of elements of Q classified in this type share no edges. However, they have 4 vertices in common. This allows a brief formalization for f_{04}. Before presenting it, take notice that, given $\{st, uv\} \in Q$, the existence of edges su, tv implies that $\{su, tv\} \in Q$. Similarly, the existence of sv, tu implies $\{sv, tu\} \in Q$. Then

$$f_{04} = \sum_{\{st,uv\}\in Q} \left(\sum_{\{su,tv\}\in Q} 1 + \sum_{\{sv,tu\}\in Q} 1 \right) = \sum_{\{st,uv\}\in Q} (a_{su}a_{tv} + a_{sv}a_{tu}). \tag{69}$$

Therefore, by proposition 4.3 we can rewrite equation 69 as

$$\begin{aligned} f_{04} &= 2n_G(\mathcal{C}_4) \\ &= \frac{1}{4}tr(A^4) - n_G(\mathcal{L}_3) - \frac{1}{2}n_G(\mathcal{L}_2) \\ &= \frac{1}{4}tr(A^4) + |Q| - \frac{1}{2}m^2. \end{aligned} \tag{70}$$

4.5.6. $\tau = 0$, $\phi = 3$ This type denotes those pairs in $Q \times Q$ that do not share an edge completely but that have 3 vertices in common. Given an element $\{st, uv\} \in Q$, the other possible elements of Q that make the pair follow this type's characterization are:

$$\begin{array}{lllll} \{su, tw\} & \{su, vw\} & \{sv, tw\} & \{sv, uw\} & \\ \{tu, sw\} & \{tu, vw\} & \{tv, sw\} & \{tv, uw\} & w \neq s,t,u,v \end{array}$$

Therefore, to calculate the value of f_{03}, we have to count how many elements of the previous list there are in the graph. Formally,

$$f_{03} = \sum_{\{st,uv\}\in Q} (\varphi_{sut} + \varphi_{svt} + \varphi_{tus} + \varphi_{tvs} + \varphi_{svu} + \varphi_{tvu} + \varphi_{tuv} + \varphi_{suv}), \tag{71}$$

where φ_{sut}, φ_{svt}, ... are functions with implicit parameter $\{st, uv\} \in Q$ and explicit parameters are three of the four vertices s,t,u or v. These $\varphi_{\dots}$ are defined as

$$\varphi_{xyz} = a_{xy}|\Gamma(z, -stuv)|. \tag{72}$$

where $x, y, z \in \{s, t, u, v\}$, with s, t, u, v the vertices of the implicit parameter. Then

$$\begin{aligned} f_{03} = & \sum_{\{st,uv\}\in Q} a_{su}|\Gamma(t, -stuv)| + \sum_{\{st,uv\}\in Q} a_{sv}|\Gamma(t, -stuv)| + \sum_{\{st,uv\}\in Q} a_{tu}|\Gamma(s, -stuv)| \\ & + \sum_{\{st,uv\}\in Q} a_{tv}|\Gamma(s, -stuv)| + \sum_{\{st,uv\}\in Q} a_{sv}|\Gamma(u, -stuv)| + \sum_{\{st,uv\}\in Q} a_{tv}|\Gamma(u, -stuv)| \\ & + \sum_{\{st,uv\}\in Q} a_{tu}|\Gamma(v, -stuv)| + \sum_{\{st,uv\}\in Q} a_{su}|\Gamma(v, -stuv)|. \end{aligned} \tag{73}$$

Looking at each $\varphi_{sut}, \varphi_{tus}, \cdots, \varphi_{suv}$ separately we see that, given $\{st, uv\} \in Q$, φ_{sut} counts the amount of neighbors of t in G_{-stuv} if $su \in E$, φ_{tus} counts the amount of neighbors of s in G_{-stuv} if $tu \in E$, and so on (figure 13).

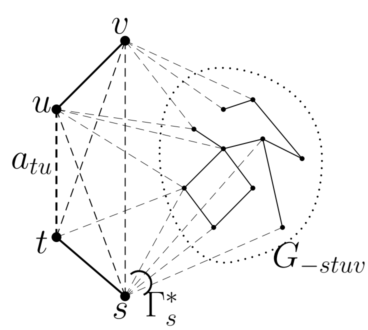


Figure 13: Illustration of φ_{tus}, that is exactly $|\Gamma^*_s| = |\Gamma(s, -stuv)|$ (the amount of neighbors of s in G_{-stuv}), provided that the edge tu exists (i.e. $a_{tu} = 1$).

Our goal is to obtain a simpler expression for equation 71. We do this by simplifying first the sum of all the $\varphi_{\dots}$. For this, we apply equation 46 to a pairwise sum of these $\varphi_{\dots}$ so as to obtain a series of expressions that are easier to evaluate

$$\begin{aligned} \varphi_{tus} + \varphi_{tvs} &= (a_{tu} + a_{tv})(k_s - 1 - a_{su} - a_{sv}), \\ \varphi_{sut} + \varphi_{svt} &= (a_{su} + a_{sv})(k_t - 1 - a_{tv} - a_{tu}), \\ \varphi_{svu} + \varphi_{tvu} &= (a_{sv} + a_{tv})(k_u - 1 - a_{ut} - a_{us}), \\ \varphi_{suv} + \varphi_{tuv} &= (a_{su} + a_{tu})(k_v - 1 - a_{vt} - a_{vs}). \end{aligned}$$

When adding all of them together, we can simplify the expression a bit more

$$
\begin{aligned}
&(a_{tu}+a_{tv})(k_s-1)-(a_{tu}+a_{tv})(a_{su}+a_{sv})+(a_{su}+a_{sv})(k_t-1)-(a_{su}+a_{sv})(a_{tu}+a_{tv})+\\
&+(a_{sv}+a_{tv})(k_u-1)-(a_{vs}+a_{vt})(a_{ut}+a_{us})+(a_{su}+a_{tu})(k_v-1)-(a_{us}+a_{ut})(a_{vt}+a_{vs})=\\
&=(a_{tu}+a_{tv})(k_s-1)+(a_{su}+a_{sv})(k_t-1)+(a_{sv}+a_{tv})(k_u-1)+(a_{su}+a_{tu})(k_v-1)\\
&-2((a_{tu}+a_{tv})(a_{su}+a_{sv})+(a_{vs}+a_{vt})(a_{ut}+a_{us})).
\end{aligned}
$$

Upon expansion of the positive part of the expression, we obtain

$$
k_s(a_{tu}+a_{tv})+k_t(a_{su}+a_{sv})+k_u(a_{sv}+a_{tv})+k_v(a_{su}+a_{tu})-2(a_{tu}+a_{tv}+a_{sv}+a_{su}),
$$

and, upon expansion of the negative part,

$$
\begin{aligned}
&-2((a_{tu}+a_{tv})(a_{su}+a_{sv})+(a_{vs}+a_{vt})(a_{ut}+a_{us}))=\\
=&-4(a_{vs}a_{ut}+a_{vt}a_{us})-2(a_{tu}+a_{vs})(a_{su}+a_{tv}).
\end{aligned}
$$

Using the two previous expressions, the expression for f_{03} becomes

$$
\begin{aligned}
f_{03}=&\sum_{\{st,uv\}\in Q}(k_s(a_{tu}+a_{tv})+k_t(a_{su}+a_{sv})+k_u(a_{sv}+a_{tv})+k_v(a_{su}+a_{tu}))\\
&-2\sum_{\{st,uv\}\in Q}(a_{tu}+a_{tv}+a_{sv}+a_{su})-4\sum_{\{st,uv\}\in Q}(a_{vs}a_{ut}+a_{vt}a_{us})\\
&-2\sum_{\{st,uv\}\in Q}(a_{tu}+a_{vs})(a_{su}+a_{tv}),
\end{aligned}
$$

which, using the results in propositions 4.1 and 4.3, can be simplified further, yielding

$$
\begin{aligned}
f_{03}=&\sum_{\{st,uv\}\in Q}(k_s(a_{tu}+a_{tv})+k_t(a_{su}+a_{sv})+k_u(a_{sv}+a_{tv})+k_v(a_{su}+a_{tu}))\\
&-2n_G(\mathcal{L}_4)-8n_G(\mathcal{C}_4)-2\sum_{\{st,uv\}\in Q}(a_{tu}+a_{vs})(a_{su}+a_{tv}). \qquad (74)
\end{aligned}
$$

We can express this type as the amount of a certain type of subgraph with the help of figure 13 and the corresponding unipartite graph in figure 6. Figure 13 shows the interpretation of the value φ_{tus} (formalized in equation 72) which, given an element $\{st, uv\} \in Q$ and the existence of edge a_{tu}, counts the amount of neighbors of s in G_{-stuv}, namely $|\Gamma_s^*| = |\Gamma(s, -stuv)|$. Notice that if $\{st, uv\} \in Q$, and assuming that $a_{tu} = 1$, then we have a 4-path: (v, u, t, s), and that by appending any vertex $w \in \Gamma_s^*$ to it we can make a 5-path (v, u, t, s, w). The same reasoning applies to the other $\varphi_{...}$. The aforementioned unipartite graph can only make clearer the fact that f_{03} counts 5-paths.

Table 5: The pattern of the 5-paths counted by each summand in equation 73 for a given $\{st, uv\} \in Q$. * is used to indicate an arbitrary neighbor of the nearest vertex in the path different from s, t, u, v.

$(*, s, t, u, v)$	$(s, t, u, v, *)$
$(*, s, t, v, u)$	$(s, t, v, u, *)$
$(*, t, s, v, u)$	$(t, s, v, u, *)$
$(*, t, s, u, v)$	$(t, s, u, v, *)$

Within each summation in equation 73, each summand counts over different 5-paths and all the 5-paths counted within the summation are different. The patterns of the 5-paths counted in each summation are shown in table 5. Notice that the paths in the right column are obtained by shifting the vertices of the paths in the left column. Crucially, the edges of Q yielding the path are consecutive in the path (i.e. they correspond to four consecutive vertices in the path). Therefore any 5-path of the graph is counted exactly by two different summations in equation 73. Therefore,

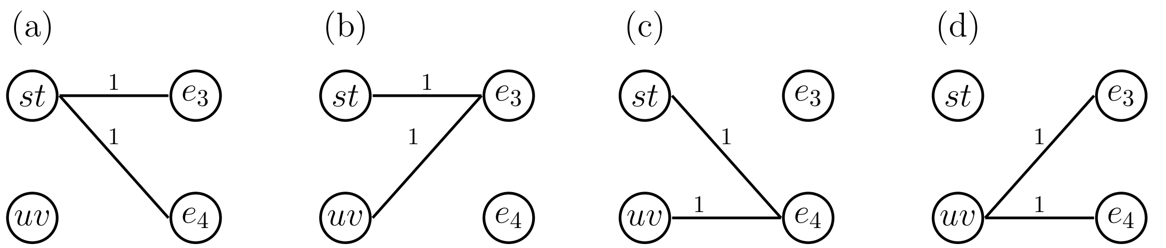
$$f_{03} = 2n_G(\mathcal{L}_5). \tag{75}$$

Then proposition 4.2 gives another formula for f_{03}, i.e.

$$f_{03} = 2 \sum_{\{st,uv\}\in Q} \left(\sum_{w_s\in\Gamma(s,-stuv)} (a_{uw_s} + a_{vw_s}) + \sum_{w_t\in\Gamma(t,-stuv)} (a_{uw_t} + a_{vw_t}) \right) \tag{76}$$

Since $n_G(\mathcal{L}_5) = f_{03}/2$, it is easy to see that equations 74 and 76 are an improvement over the formula derived for $n_G(\mathcal{L}_5)$ in previous work [34] because they provide a more compact expression, and equation 74 unveils the contribution of $n_G(\mathcal{L}_4)$ and $n_G(\mathcal{C}_4)$.

4.5.7. $\tau = 0$, $\phi = 2$, Subtype 1 Type 021 is defined by an unlabeled bipartite graph (figure 5). Figure 14 shows the 10 possible forms that elements of Q such that when paired with $\{st, uv\} \in Q$ yield a pair of $Q \times Q$ classified as 021 can take by labeling the left partition with st, uv and considering all the possible labelings of the right partition of that type. However, by symmetry between the forms in figure 14(b) and those of figure 14(c), there are only 6 unique forms.



(a) $\{sw, tx\} \in Q(G_{-uv})$

(b) $\{su, wx\} \in Q(G_{-tv})$, $\{sv, wx\} \in Q(G_{-tu})$, $\{tu, wx\} \in Q(G_{-sv})$, $\{tv, wx\} \in Q(G_{-su})$

(c) $\{su, yz\} \in Q(G_{-tv})$, $\{sv, yz\} \in Q(G_{-tu})$, $\{tu, yz\} \in Q(G_{-sv})$, $\{tv, yz\} \in Q(G_{-su})$

(d) $\{uw, vx\} \in Q(G_{-st})$

Figure 14: Elements of Q such that when paired with element $\{st, uv\} \in Q$, the pair is classified as type 021. Elements in (b) and (c) are symmetric.

Then we can formalize f_{021} as

$$\begin{aligned} f_{021} = & \sum_{\{st,uv\}\in Q} \varphi_{st} + \sum_{\{st,uv\}\in Q} \varphi_{uv} \\ & + \sum_{\{st,uv\}\in Q} \varepsilon_{su} + \sum_{\{st,uv\}\in Q} \varepsilon_{sv} + \sum_{\{st,uv\}\in Q} \varepsilon_{tu} + \sum_{\{st,uv\}\in Q} \varepsilon_{tv}, \end{aligned} \tag{77}$$

where φ_{xy} and ε_{xy} are auxiliary functions with implicit parameter $\{st, uv\} \in Q$ and explicit parameters xy, that are defined as

$$\begin{aligned}\varphi_{xy} &= \sum_{w\in\Gamma(x,-stuv)} \sum_{w'\in\Gamma(y,-stuvw)} 1 \\ &= \sum_{w\in\Gamma(x,-stuv)} |\Gamma(y,-stuvw)|, \\ \varepsilon_{xy} &= a_{xy}|E(G_{-stuv})|, \qquad x, y \in \{s, t, u, v\}.\end{aligned} \tag{78}$$

The functions φ_{st} and φ_{uv} count the elements of the form of those illustrated in figures 14(a) and 14(d). The first function counts, for each neighbor of s, $w_s \neq t, u, v$, the number of neighbors of t, $w_t \neq s, t, u, v, w_s$. Likewise for the second function. On the other hand, the values ε_{su}, ε_{sv}, ε_{tu}, ε_{tv} count the edges $xy \in E$, $x, y \neq s, t, u, v$ such that when paired with su, sv, tu, tv form an element of Q whose form is that of those elements illustrated in figures 14(b) and 14(c). These amounts are counted only if such edges exist in the graph, hence the a_{su} for ε_{su}, and likewise for the other $\varepsilon_{..}$.

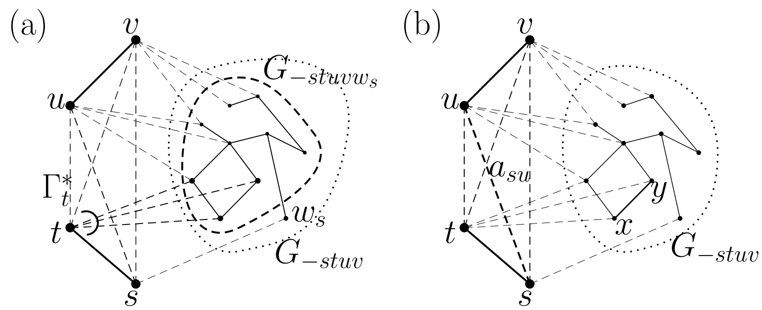


Figure 15: Illustration of (a) φ_{st}, and (b) ε_{su}. In (a), w_s represents the only neighbor of s different from t, u, v. Therefore, in this case, φ_{st} is exactly the amount of vertices in G_{-stuvw_s} neighbors of t, indicated with $\Gamma_t^* = \Gamma(t, -stuvw_s)$. ε_{su} requires the existence of an edge between s and u, indicated with a_{su}, and is equal to the amount of edges in G_{-stuv}.

We can take common factor in equation 78 and simplify it. We obtain

$$f_{021} = \sum_{\{st,uv\}\in Q} (\varphi_{st} + \varphi_{uv} + (a_{su} + a_{sv} + a_{tu} + a_{tv})|E(G_{-stuv})|), \tag{79}$$

and now we can use the results in equations 45 and 46 to calculate $|E(G_{-stuv})|$ and φ_{st} and φ_{uv}:

$$\begin{aligned}(a_{su} + a_{sv} + a_{tu} + a_{tv})|E(G_{-stuv})| &= (a_{su} + a_{sv} + a_{tu} + a_{tv})m \\ &- (a_{su} + a_{sv} + a_{tu} + a_{tv})(k_s + k_t + k_u + k_v) \\ &+ (a_{su} + a_{sv} + a_{tu} + a_{tv})^2 \\ &+ 2(a_{su} + a_{sv} + a_{tu} + a_{tv}).\end{aligned}$$

Recall that for $\{st, uv\} \in Q$, $a_{st} + a_{uv} = 2$. Adding this over all elements $\{st, uv\} \in Q$ and using the result in proposition 4.1 we obtain

$$\begin{aligned}
& m \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv}) \\
& - \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv})(k_s + k_t + k_u + k_v) \\
& + \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv})^2 + 2 \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv}) = \\
& = (m+2) n_G(\mathcal{L}_4) \\
& - \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv})(k_s + k_t + k_u + k_v) \\
& + \sum_{\{st,uv\}\in Q} (a_{su} + a_{sv} + a_{tu} + a_{tv})^2.
\end{aligned}$$

Now we derive simpler expressions for φ_{st} and φ_{uv}

$$\begin{aligned}
|\Gamma(t, -stuvw_s)| &= k_t - \sum_{\pi\in\{s,t,u,v,w_s\}} a_{t\pi} = k_t - (a_{ts} + a_{tu} + a_{tv} + a_{tw_s}), \\
\varphi_{st} &= \sum_{w_s\in\Gamma(s,-stuv)} (k_t - (a_{ts} + a_{tu} + a_{tv} + a_{tw_s})) = \\
&= \sum_{w_s\in\Gamma(s,-stuv)} (k_t - (a_{ts} + a_{tu} + a_{tv})) - \sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s} = \\
&= (k_s - a_{st} - a_{su} - a_{sv})(k_t - a_{ts} - a_{tu} - a_{tv}) - \sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s} = \\
&= (k_s - a_{su} - a_{sv} - 1)(k_t - a_{tu} - a_{tv} - 1) - \sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s}.
\end{aligned} \tag{80}$$

Likewise for φ_{uv}

$$\varphi_{uv} = (k_u - a_{us} - a_{ut} - 1)(k_v - a_{vs} - a_{vt} - 1) - \sum_{w_u\in\Gamma(u,-stuv)} a_{vw_u}. \tag{81}$$

The negative summations in φ_{st}, equation 80 (and φ_{uv}, equation 81) represent the amount of vertices from G_{-stuv} neighbors of s (of u) in G that are also neighbors of t (of v), in G. Therefore, the triangles formed by vertices s, t, w_s and u, v, w_u respectively.

We obtain another intermediate result

$$
\begin{aligned}
f_{021} = & (m+2)n_G(\mathcal{L}_4) \\
& + \sum_{\{st,uv\}\in Q} (a_{su}+a_{sv}+a_{tu}+a_{tv})^2 \\
& + \sum_{\{st,uv\}\in Q} (k_s - a_{su} - a_{sv} - 1)(k_t - a_{tu} - a_{tv} - 1) \\
& + \sum_{\{st,uv\}\in Q} (k_u - a_{us} - a_{ut} - 1)(k_v - a_{vs} - a_{vt} - 1) \\
& - \sum_{\{st,uv\}\in Q} (a_{su}+a_{sv}+a_{tu}+a_{tv})(k_s+k_t+k_u+k_v) \\
& - \sum_{\{st,uv\}\in Q} \left(\sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s} + \sum_{w_u\in\Gamma(u,-stuv)} a_{vw_u} \right).
\end{aligned}
\tag{82}
$$

We can benefit from expanding the subexpressions

$$
(a_{su}+a_{sv}+a_{tu}+a_{tv})^2 = 2(a_{su}+a_{tu})(a_{sv}+a_{tv})+2(a_{su}a_{tu}+a_{sv}a_{tv})+a_{su}+a_{sv}+a_{tu}+a_{tv}. \tag{83}
$$

Equation 83 is used later in this work. Furthermore,

$$
\begin{aligned}
& (k_s - a_{su} - a_{sv} - 1)(k_t - a_{tu} - a_{tv} - 1) + (k_u - a_{us} - a_{ut} - 1)(k_v - a_{vs} - a_{vt} - 1) = \\
& = (k_sk_t + k_uk_v) - (k_s+k_t+k_u+k_v) - \\
& \quad - [k_s(a_{tu}+a_{tv}) + k_t(a_{su}+a_{sv}) + k_u(a_{sv}+a_{tv}) + k_v(a_{su}+a_{tu})] \\
& \quad + (a_{su}+a_{sv})(a_{tu}+a_{tv}) + (a_{su}+a_{tu})(a_{sv}+a_{tv}) \\
& \quad + 2(a_{su}+a_{sv}+a_{tu}+a_{tv}) + 2.
\end{aligned}
$$

We can use these two expressions to expand equation 82 to obtain

$$
\begin{aligned}
f_{021} = & (m+5)n_G(\mathcal{L}_4) + 2n_G(\mathcal{C}_4) + 2|Q| - K_G \\
& + 3\sum_{\{st,uv\}\in Q} (a_{su}+a_{tu})(a_{sv}+a_{tv}) + 3\sum_{\{st,uv\}\in Q} (a_{su}a_{tu}+a_{sv}a_{tv}) \\
& + \sum_{\{st,uv\}\in Q} (k_sk_t + k_uk_v) \\
& - \sum_{\{st,uv\}\in Q} (k_s(a_{tu}+a_{tv}) + k_t(a_{su}+a_{sv}) + k_u(a_{sv}+a_{tv}) + k_v(a_{su}+a_{tu})) \\
& - \sum_{\{st,uv\}\in Q} (a_{su}+a_{sv}+a_{tu}+a_{tv})(k_s+k_t+k_u+k_v) \\
& - \sum_{\{st,uv\}\in Q} \left(\sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s} + \sum_{w_u\in\Gamma(u,-stuv)} a_{vw_u} \right).
\end{aligned}
\tag{84}
$$

We split the r.h.s. of equation 79 into two halves: the φ's and the ε's. On the one hand,

$$
\sum_{\{st,uv\}\in Q} \varphi_{st} + \sum_{\{st,uv\}\in Q} \varphi_{uv} = n_G(\mathcal{L}_4 \oplus \mathcal{L}_2) \tag{85}
$$

because the 1st summation is counting all the $\mathcal{L}_4 \oplus \mathcal{L}_2$ such that $\mathcal{L}_2$ is the edge uv and $\mathcal{L}_4$ is the path (w, s, t, w') and the 2nd summation is counting all the $\mathcal{L}_4 \oplus \mathcal{L}_2$ such that $\mathcal{L}_2$ is the edge st and $\mathcal{L}_4$ is the path (w, u, v, w'). On the other hand,

$$\sum_{\{st,uv\}\in Q} \varepsilon_{su} + \sum_{\{st,uv\}\in Q} \varepsilon_{sv} + \sum_{\{st,uv\}\in Q} \varepsilon_{tu} + \sum_{\{st,uv\}\in Q} \varepsilon_{tv} = n_G(\mathcal{L}_4 \oplus \mathcal{L}_2) \quad (86)$$

because each summation is counting all the $\mathcal{L}_4 \oplus \mathcal{L}_2$ such that the $\mathcal{L}_4$ is build on the vertices s, t, u, v and $\mathcal{L}_2$ is any edge that is not formed by these vertices. Every summation is in charge of one of the four different ways in which a distinct $\mathcal{L}_4$ can be produced linking one vertex of the edge st with a vertex of the edge uv. Therefore,

$$f_{021} = 2n_G(\mathcal{L}_4 \oplus \mathcal{L}_2). \quad (87)$$

A detailed proof of equation 87 with a technique that is applied to types 022 and 01 is can be found in Appendix D.

4.5.8. $\tau = 0$, $\phi = 2$, Subtype 2 We follow the same approach as for type 021 (this second subtype is simpler to formalize). Figure 16 shows all the elements of Q such that when paired with $\{st, uv\} \in Q$ yield a pair of $Q \times Q$ classified as type 022 for each of the two labeled bipartite graphs of that type. This gives 8 configurations that are constructed by making two new independent edges, one edge linking a new vertex, say w, to one of the vertices of st and another edge linking another vertex, say x, to edge uv, $(w, x \neq s, t, u, v)$, so that the pair of new edges belongs to Q. However, only 4 configurations are distinct by symmetry: x and w are interchangeable. Therefore, the elements of $Q \times Q$ defined in figure 16(a) are the same as those of 16 (b). As a result of this analysis, f_{022} can be defined as

$$f_{022} = \sum_{\{st,uv\}\in Q} \varphi_{su} + \sum_{\{st,uv\}\in Q} \varphi_{sv} + \sum_{\{st,uv\}\in Q} \varphi_{tu} + \sum_{\{st,uv\}\in Q} \varphi_{tv}, \quad (88)$$

where φ_{xy} is an auxiliary function defined as in equation 78. φ_{su} can be understood from the case of φ_{st} in figure 15(a).

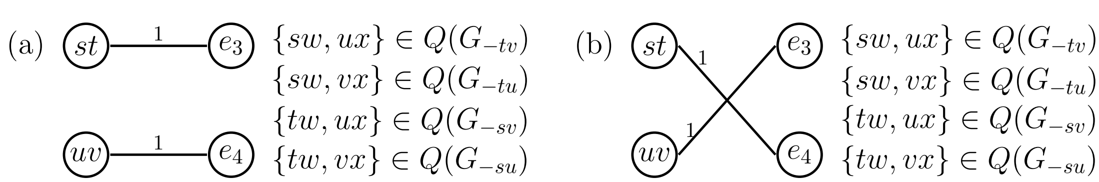


Figure 16: Elements of Q such that when paired with element $\{st, uv\} \in Q$, the pair is classified as type 022. (a) One of the bipartite graphs of type 022. (b) The other bipartite graph of type 022. The elements in (a) are symmetric to those of (b).

We aim at explaining the definition of f_{022} above. First, expand, like it is done in section 4.5.7, the expressions for φ_{su}, φ_{sv}, φ_{tu}, φ_{tv} as

$$\begin{aligned}
\varphi_{su} &= (k_s - a_{su} - a_{sv} - 1)(k_u - a_{us} - a_{ut} - 1) - \sum_{w_s \in \Gamma(s,-stuv)} a_{uw_s}, \\
\varphi_{sv} &= (k_s - a_{su} - a_{sv} - 1)(k_v - a_{vs} - a_{vt} - 1) - \sum_{w_s \in \Gamma(s,-stuv)} a_{vw_s}, \\
\varphi_{tu} &= (k_t - a_{tu} - a_{tv} - 1)(k_u - a_{us} - a_{ut} - 1) - \sum_{w_t \in \Gamma(t,-stuv)} a_{uw_t}, \\
\varphi_{tv} &= (k_t - a_{tu} - a_{tv} - 1)(k_v - a_{vs} - a_{vt} - 1) - \sum_{w_t \in \Gamma(t,-stuv)} a_{vw_t}.
\end{aligned}$$

Inserting these expressions back to equation 88 and taking common factors out, one obtains

$$\begin{aligned}
f_{022} = &\sum_{\{st,uv\} \in Q} (k_s + k_t - a_{su} - a_{sv} - a_{tu} - a_{tv} - 2)(k_u + k_v - a_{us} - a_{ut} - a_{vs} - a_{vt} - 2) \\
&- \sum_{\{st,uv\} \in Q} \left(\sum_{w_s \in \Gamma(s,-stuv)} (a_{uw_s} + a_{vw_s}) + \sum_{w_t \in \Gamma(t,-stuv)} (a_{uw_t} + a_{vw_t}) \right),
\end{aligned}$$

which can be further developed and then simplified, using proposition 4.2, as

$$\begin{aligned}
f_{022} = &-n_G(\mathcal{L}_5) + \sum_{\{st,uv\} \in Q} (k_s + k_t - 2)(k_u + k_v - 2) \\
&+ \sum_{\{st,uv\} \in Q} (a_{us} + a_{ut} + a_{vs} + a_{vt})^2 \\
&+ 4 \sum_{\{st,uv\} \in Q} (a_{us} + a_{ut} + a_{vs} + a_{vt}) \\
&- \sum_{\{st,uv\} \in Q} (k_s + k_t + k_u + k_v)(a_{us} + a_{ut} + a_{vs} + a_{vt}).
\end{aligned}$$

Using equation 83, the result in proposition 4.1, and by expanding $(k_s + k_t - 2)(k_u + k_v - 2)$, we obtain

$$\begin{aligned}
f_{022} = &\, 5n_G(\mathcal{L}_4) - n_G(\mathcal{L}_5) + 4|Q| - 2K_G + \sum_{\{st,uv\} \in Q} (k_s + k_t)(k_u + k_v) \\
&+ 2 \sum_{\{st,uv\} \in Q} (a_{su} + a_{tu})(a_{sv} + a_{tv}) + 2 \sum_{\{st,uv\} \in Q} (a_{su}a_{tu} + a_{sv}a_{tv}) \\
&- \sum_{\{st,uv\} \in Q} (a_{su} + a_{tu} + a_{sv} + a_{tv})(k_s + k_t + k_u + k_v).
\end{aligned} \tag{89}$$

Now, notice that this type counts pairs of $\mathcal{L}_3 \oplus \mathcal{L}_3$: given a fixed $\{st, uv\} \in Q$, the value φ_{su}, for example, counts the neighbors of s in G_{-stuv} and the neighbors of u in

G_{-stuvw}, where $w \in \Gamma(s, -stuv)$. Similarly for the other $\varphi_{..}$. Therefore, the form of the subgraphs counted by each summation of equation 88 for a fixed $\{st, uv\} \in Q$ are

$$\varphi_{su} : \{(z_s, s, t), (z_u, u, v)\}, \qquad \varphi_{sv} : \{(z_s, s, t), (u, v, z_v)\},$$
$$\varphi_{tu} : \{(s, t, z_t), (z_u, u, v)\}, \qquad \varphi_{tv} : \{(s, t, z_t), (u, v, z_v)\}.$$

It is easy to see that a fixed $\mathcal{L}_3 \oplus \mathcal{L}_3$ of the form $H = \{(s, t, u), (v, w, x)\}$ is counted four times in equation 88. First, notice that we can make exactly four elements of Q using the edges of H

$$q_1 = \{st, vw\},\ q_2 = \{st, wx\},\ q_3 = \{tu, vw\},\ q_4 = \{tu, wx\}.$$

Given one of them, say q_1, our graph H will only be counted by φ_{tw} because for q_1 the subgraphs counted in equation 88 are of the form

$$\varphi_{sv} : \{(z_s, s, t), (z_v, v, w)\}, \qquad \varphi_{sw} : \{(z_s, s, t), (v, w, z_w)\},$$
$$\varphi_{tv} : \{(s, t, z_t), (z_v, v, w)\}, \qquad \varphi_{tw} : \{(s, t, z_t), (v, w, z_w)\}.$$

Similarly, for each of the other q_i's, there is a unique $\varphi_{..}$ where H is counted. Then, when calculating f_{022} with equation 88, every distinct H will be counted 4 times, i.e.

$$f_{022} = 4n_G(\mathcal{L}_3 \oplus \mathcal{L}_3). \tag{90}$$

4.5.9. $\tau = 0$, $\phi = 1$ Finally, f_{01} can be formalized as

$$f_{01} = \sum_{\{st,uv\}\in Q} \varphi_s + \sum_{\{st,uv\}\in Q} \varphi_t + \sum_{\{st,uv\}\in Q} \varphi_u + \sum_{\{st,uv\}\in Q} \varphi_v, \tag{91}$$

where φ_k is a function with an implicit parameter $\{st, uv\} \in Q$ and an explicit parameter $k \in \{s, t, u, v\}$, that is defined as

$$\begin{aligned} \varphi_k &= \sum_{\{kw,xy\}\in Q(G_{-stuv\setminus k})} 1 \qquad (92) \\ &= \sum_{w\in\Gamma(k,-stuv)} \sum_{xy\in E(G_{-stuvw})} 1 \\ &= \sum_{w\in\Gamma(k,-stuv)} |E(G_{-stuvw})| \end{aligned}$$

and then

$$f_{01} = \sum_{\{st,uv\}\in Q} \sum_{k\in\{s,t,u,v\}} \sum_{w\in\Gamma(k,-stuv)} |E(G_{-stuvw})|. \tag{93}$$

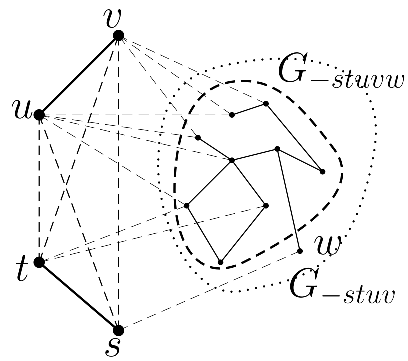


Figure 17: Illustration of φ_s. Here w represents the only neighbor of s different from s, t, u, v. Thus φ_s is exactly the amount of edges in G_{-stuvw}.

The fact that equation 91 counts over $\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$ is readily seen by noting that, for a fixed $\{st, uv\} \in Q$, the summations of the $\varphi_{\cdot}$ count subgraphs of the form

$$\varphi_s : \{(z_s, s, t), (u, v), (z_1, z_2)\}, \qquad \varphi_t : \{(s, t, z_t), (u, v), (z_1, z_2)\}$$
$$\varphi_u : \{(s, t), (z_u, u, v), (z_1, z_2)\}, \qquad \varphi_v : \{(s, t), (u, v, z_v), (z_1, z_2)\}.$$

Now, consider some $\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2$ of the form $H = \{(s, t, u), (v, w), (x, y)\}$. We can make exactly four elements of Q with its edges, i.e.

$$q_1 = \{st, vw\},\ q_2 = \{st, xy\},\ q_3 = \{tu, vw\},\ q_4 = \{tu, xy\}.$$

Given one of them, say q_1, our graph H will only be counted in only one of the $\varphi_{\cdot}$ (by φ_t in the example) because for q_1 the subgraphs counted in equation 91 are of the form

$$\varphi_s : \{(z_s, s, t), (v, w), (z_1, z_2)\}, \qquad \varphi_t : \{(s, t, z_t), (v, w), (z_1, z_2)\}$$
$$\varphi_v : \{(s, t), (z_v, v, w), (z_1, z_2)\}, \qquad \varphi_w : \{(s, t), (v, w, z_w), (z_1, z_2)\}.$$

Similarly, for each of the other q_i's, there is a unique $\varphi_{\cdot}$ where H is counted. Then, when calculating f_{01} with equation 91, every distinct H will be counted 4 times, i.e.

$$f_{01} = 4 n_G(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2). \tag{94}$$

4.6. Variance of the number of crossings

The aim of this paper is to provide a general formula for the variance of C in any given graph when its vertices are placed in a linear arrangement. Thanks to equation 38 and the results in the previous sections, now we are able to give such expression:

$$\begin{aligned}\mathbb{V}[C] = &\frac{2}{45}(m+2)|Q| - \frac{1}{180}n_G(\mathcal{L}_5) - \frac{2m+7}{180}n_G(\mathcal{L}_4) - \frac{3}{45}n_G(\mathcal{C}_4) + \frac{1}{90}K_G \\ &- \frac{1}{60}\sum_{\{st,uv\}\in Q}(k_s(a_{tu}+a_{tv}) + k_t(a_{su}+a_{sv}) + k_u(a_{sv}+a_{vt}) + k_v(a_{su}+a_{tu})) \\ &+ \frac{1}{180}\sum_{\{st,uv\}\in Q}(a_{su}+a_{sv}+a_{tu}+a_{tv})(k_s+k_t+k_u+k_v) \\ &+ \frac{1}{180}\sum_{\{st,uv\}\in Q}(k_s+k_t)(k_u+k_v) - \frac{1}{90}\sum_{\{st,uv\}\in Q}(k_sk_t+k_uk_v) \\ &+ \frac{1}{30}\sum_{\{st,uv\}\in Q}(a_{tu}+a_{sv})(a_{tv}+a_{su}) \\ &+ \frac{1}{90}\sum_{\{st,uv\}\in Q}\left(\sum_{w_s\in\Gamma(s,-stuv)} a_{tw_s} + \sum_{w_u\in\Gamma(u,-stuv)} a_{vw_u}\right).\end{aligned} \tag{95}$$

In spite of our efforts we could not reduce the expressions any further for general graphs. For the case of trees, due to being acyclic, equation 95 is simplified to:

$$\begin{aligned}\mathbb{V}[C] = &\frac{2}{45}(m+2)|Q| - \frac{1}{180}n_G(\mathcal{L}_5) - \frac{2m+7}{180}n_G(\mathcal{L}_4) + \frac{1}{90}K_G \\ &- \frac{1}{60}\sum_{\{st,uv\}\in Q}(k_s(a_{tu}+a_{tv}) + k_t(a_{su}+a_{sv}) + k_u(a_{sv}+a_{tv}) + k_v(a_{su}+a_{tu})) \\ &+ \frac{1}{180}\sum_{\{st,uv\}\in Q}(a_{su}+a_{sv}+a_{tu}+a_{tv})(k_s+k_t+k_u+k_v) \\ &+ \frac{1}{180}\sum_{\{st,uv\}\in Q}(k_s+k_t)(k_u+k_v) - \frac{1}{90}\sum_{\{st,uv\}\in Q}(k_sk_t+k_uk_v).\end{aligned} \tag{96}$$

5. Theoretical examples

In the coming sections we derive compact formulae for different types of graphs.

5.1. 1-regular graphs

The general characterization of the f_i's in Table 4 based on equation 55 allows one to see that

$$f_{01}(\mathbf{1}_n) = f_{021}(\mathbf{1}_n) = f_{022}(\mathbf{1}_n) = f_{03}(\mathbf{1}_n) = f_{04}(\mathbf{1}_n) = f_{13}(\mathbf{1}_n) = 0 \tag{97}$$

and that

$$f_{00}(\mathbf{1}_n) = 6\binom{m}{4}, \quad f_{24}(\mathbf{1}_n) = \binom{m}{2}, \quad f_{12}(\mathbf{1}_n) = 6\binom{m}{3} \tag{98}$$

The applicationof the results above to equation 42 with $m = n/2$ yields, after some algebra,

$$\mathbb{V}\left[C(\mathbf{1}_n)\right] = \frac{1}{360}(n-2)n(n+6) \tag{99}$$

when $n \geq 2$ and and $\mathbb{V}\left[C(\mathbf{1}_n)\right] = 0$ otherwise.

5.2. Quasi-star trees

The general characterization of the f_i's in Table 4 based on equation 55 allows one to see that

$$f_{00}(\mathcal{Q}_n) = f_{12}(\mathcal{Q}_n) = f_{04}(\mathcal{Q}_n) = f_{03}(\mathcal{Q}_n) = f_{022}(\mathcal{Q}_n) = f_{021}(\mathcal{Q}_n) = f_{01}(\mathcal{Q}_n) = 0. \tag{100}$$

Recalling table 1,

$$f_{24}(\mathcal{Q}_n) = |Q(\mathcal{Q}_n)| = n - 3. \tag{101}$$

Since all f_i's have to add up to $|Q(\mathcal{Q}_n)|^2$ (equation 40), one has

$$f_{13}(\mathcal{Q}_n) = |Q(\mathcal{Q}_n)|^2 - f_{24}(\mathcal{Q}_n) = (n-3)(n-4). \tag{102}$$

Therefore,

$$\begin{aligned}\mathbb{V}\left[C(\mathcal{Q}_n)\right] &= \sum_{i=0}^{8} f_i(\mathcal{Q}_n)\mathbb{E}\left[\gamma_i\right] = f_{24}(\mathcal{Q}_n)\mathbb{E}\left[\gamma_{24}\right] + f_{13}(\mathcal{Q}_n)\mathbb{E}\left[\gamma_{13}\right] = \\ &= (n-3)\frac{2}{9} + (n-3)(n-4)\frac{1}{18} \\ &= \frac{1}{18}n(n-3)\end{aligned} \tag{103}$$

for $n \geq 3$ ($\mathbb{V}\left[C(\mathcal{Q}_n)\right] = 0$ otherwise).

5.3. Complete graphs

Although we know that $\mathbb{V}\left[C(\mathcal{K}_n)\right] = 0$, the case is important as a test of the soundness of the theory.

The derivation of many of the $f_i's$ requires calculating $n_{\mathcal{K}_n}(\mathcal{L}_{n'})$, the number of n'-paths in $\mathcal{K}_n$, for some $1 < n' \leq n$. Some $\mathcal{L}_{n'}$ is obtained with a path starting from a vertex and visiting new vertices $n'-1$ times. Then

$$n_{\mathcal{K}_n}(\mathcal{L}_{n'}) = \frac{1}{2}\prod_{i=0}^{n'-1}(n-i) = \frac{1}{2}\frac{n!}{(n-n')!} \tag{104}$$

for $1 < n' \leq n$. The 1/2 factor comes from the fact that the same $\mathcal{L}_n$ is obtained with a walk from the initial to the end vertex but also backwards. $n_{\mathcal{K}_n}(\mathcal{L}_2) = |E(\mathcal{K}_n)|$ and $n_{\mathcal{K}_n}(\mathcal{L}_n) = n!/2$ as expected. Furthermore,

$$n_{\mathcal{K}_n}(\mathcal{L}_{n_1} \oplus \mathcal{L}_{n_2}) = n_{\mathcal{K}_n}(\mathcal{L}_{n_1})n_{\mathcal{K}_{n-n_1}}(\mathcal{L}_{n_2}) = \frac{1}{4}\frac{n!}{(n-n_1-n_2)!} \tag{105}$$

for $1 < n_1, n_2 \leq n$ with $n_1 + n_2 \leq n$ and $n_1 \neq n_2$. In case $n_1 = n_2$ we have

$$n_{\mathcal{K}_n}(\mathcal{L}_{n_1} \oplus \mathcal{L}_{n_2}) = \frac{1}{2}n_{\mathcal{K}_n}(\mathcal{L}_{n_1})n_{\mathcal{K}_{n-n_1}}(\mathcal{L}_{n_2}) = \frac{1}{8}\frac{n!}{(n-n_1-n_2)!}. \tag{106}$$

$\tau = 0$, $\phi = 0$: $f_{00}(\mathcal{K}_n)$ is easy to calculate via equation 59. For any $\mathcal{K}_n$ it is easy to see that, for any $s \in V(G)$, G_{-s} is also a complete graph. This can also be generalized to G_{-L} with $L \subseteq V(G)$. Adapting the formula for $Q(\mathcal{K}_n)$ in table 1, we obtain $|Q((\mathcal{K}_n)_{-stuv})| = 3\binom{n-4}{4}$. Therefore,

$$f_{00}(\mathcal{K}_n) = \sum_{\{st,uv\}\in Q(\mathcal{K}_n)} |Q((\mathcal{K}_n)_{-stuv})| = \sum_{\{st,uv\}\in Q(\mathcal{K}_n)} 3\binom{n-4}{4} = 630\binom{n}{8}. \tag{107}$$

$\tau = 2$, $\phi = 4$ Combining equation 56 and table 1, we have that

$$f_{24}(\mathcal{K}_n) = |Q(\mathcal{K}_n)| = 3\binom{n}{4}. \tag{108}$$

$\tau = 1$, $\phi = 3$ We derive $f_{13}(\mathcal{K}_n)$ via equation 64, with the help of equation 105

$$f_{13}(\mathcal{K}_n) = 2n_{\mathcal{K}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2) = 2\frac{1}{4}\frac{n!}{(n-5)!} = 60\binom{n}{5} \tag{109}$$

$\tau = 1$, $\phi = 2$ We can easily obtain an expression for $f_{12}(\mathcal{K}_n)$ via equation 68, i.e., $6n_{\mathcal{K}_n}(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2)$. To do so we can first remove an edge of the graph and then count the amount of $\mathcal{L}_2 \oplus \mathcal{L}_2$, or the other way around. We use the first way, so

$$\begin{aligned} f_{04}(\mathcal{K}_n) &= 6n_{\mathcal{K}_n}(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) = \frac{1}{3}6n_{\mathcal{K}_n}(\mathcal{L}_2)n_{\mathcal{K}_{n-2}}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 2\binom{n}{2}3\binom{n-2}{4} = 90\binom{n}{6}. \end{aligned} \tag{110}$$

$\tau = 0$, $\phi = 4$ Via equation 69, one obtains

$$\begin{aligned} f_{04}(\mathcal{K}_n) &= \sum_{\{st,uv\}\in Q(\mathcal{K}_n)} (a_{su}a_{tv} + a_{sv}a_{tu}) = \sum_{\{st,uv\}\in Q(\mathcal{K}_n)} 2 \\ &= 2|Q(\mathcal{K}_n)| = 6\binom{n}{4}. \end{aligned} \tag{111}$$

$\tau = 0$, $\phi = 3$ Applying equation 104 to equation 75 one obtains

$$\begin{aligned} f_{03}(\mathcal{K}_n) &= n(n-1)(n-2)(n-3)(n-4) \\ &= 120\binom{n}{5}. \end{aligned} \tag{112}$$

$\tau = 0$, $\phi = 2$, type 1 Applying equation 105 to equation 87 one obtains

$$f_{021}(\mathcal{K}_n) = 2\frac{1}{4}\frac{n!}{(n-6)!} = 360\binom{n}{6} \tag{113}$$

$\tau = 0$, $\phi = 2$, type 2 Likewise, by applying 106 to equation 90 we immediately obtain

$$f_{022}(\mathcal{K}_n) = 4\frac{1}{8}\frac{n!}{(n-6)!} = 360\binom{n}{6} \tag{114}$$

Notice that $f_{021}(\mathcal{K}_n) = f_{022}(\mathcal{K}_n)$.

$\tau = 0,\ \phi = 1$ We rely on equation 94 to derive an expression of this type in complete graphs. First, equation 105 tells us that in complete graphs we have

$$n_{\mathcal{K}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2) = \frac{1}{2}\binom{n}{2}3\binom{n-2}{3}.$$

as we have previously seen. Using this we obtain

$$n_{\mathcal{K}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) = \binom{n}{2} n_{\mathcal{K}_{n-2}}(\mathcal{L}_3 \oplus \mathcal{L}_2) = \frac{1}{16}\prod_{i=0}^{6}(n-i) = 315\binom{n}{7}.$$

Finally,

$$f_{01}(\mathcal{K}_n) = 4 n_{\mathcal{K}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) = 1260\binom{n}{7}. \tag{115}$$

5.3.1. Variance Now we are able to obtain a formula for the variance in complete graphs. As mentioned in previous sections we expect this variance to be 0. Indeed:

$$\begin{aligned}\mathbb{V}\left[C(\mathcal{K}_n)\right] = \sum_{i=0}^{8} f_i(\mathcal{K}_n)\mathbb{E}\left[\gamma_i\right] = 3\binom{n}{4}\frac{2}{9} + 60\binom{n}{5}\frac{1}{18} + 90\binom{n}{6}\frac{1}{45} + \\ + 6\binom{n}{4}\left(-\frac{1}{9}\right) + 120\binom{n}{5}\left(-\frac{1}{36}\right) + 360\binom{n}{6}\left(-\frac{1}{90}\right) + 360\binom{n}{6}\frac{1}{180} = \\ = \frac{2}{3}\binom{n}{4} + \frac{10}{3}\binom{n}{5} + 2\binom{n}{6} - \frac{2}{3}\binom{n}{4} - \frac{10}{3}\binom{n}{5} - 4\binom{n}{6} + 2\binom{n}{6} = 0.\end{aligned}$$

5.4. Complete bipartite graphs

We derive the f_i's for $\mathcal{K}_{n_1,n_2}$ by mere counting of subgraphs as in the previous section but with the support of a new figure (figure 18).

$\tau = 0,\ \phi = 0$ We follow the same technique used for complete graphs, i.e.

$$\begin{aligned} f_{00}(\mathcal{K}_{n_1,n_2}) &= \sum_{\{st,uv\}\in Q(\mathcal{K}_{n_1,n_2})} |Q((\mathcal{K}_{n_1,n_2})_{-stuv})| \\ &= \sum_{\{st,uv\}\in Q(\mathcal{K}_{n_1,n_2})} 2\binom{n_1-2}{2}\binom{n_2-2}{2} \\ &= |Q(\mathcal{K}_{n_1,n_2})| \cdot |Q(\mathcal{K}_{n_1-2,n_2-2})| \\ &= 2\binom{n_1-2}{2}\binom{n_2-2}{2}2\binom{n_1}{2}\binom{n_2}{2} \\ &= 144\binom{n_1}{4}\binom{n_2}{4}. \end{aligned} \tag{116}$$

$\tau = 2,\ \phi = 4$ Equation 56 gives

$$f_{24}(\mathcal{K}_{n_1,n_2}) = |Q(\mathcal{K}_{n_1,n_2})| = 2\binom{n_1}{2}\binom{n_2}{2}. \tag{117}$$

$\tau = 1,\ \phi = 3$ Via equation 64

$$f_{13}(\mathcal{K}_{n_1,n_2}) = 2 \sum_{st \in E(\mathcal{K}_{n_1,n_2})} n_{\mathcal{K}_{n_1-1,n_2-1}}(\mathcal{L}_3) = 2|E(\mathcal{K}_{n_1,n_2})| n_{\mathcal{K}_{n_1-1,n_2-1}}(\mathcal{L}_3).$$

Through figure 18(a) we can easily calculate the amount of $\mathcal{L}_3$

$$\begin{aligned} f_{13}(\mathcal{K}_{n_1,n_2}) &= 2n_1n_2\frac{1}{2}((n_1-1)(n_2-1)(n_1-2) + (n_2-1)(n_1-1)(n_2-2)) \\ &= 12\binom{n_1}{3}\binom{n_2}{2} + 12\binom{n_1}{2}\binom{n_2}{3}. \end{aligned} \tag{118}$$

$\tau = 1,\ \phi = 2$ Equation 68 gives

$$\begin{aligned} f_{12}(\mathcal{K}_{n_1,n_2}) &= 6n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 6\frac{1}{3}|E(\mathcal{K}_{n_1,n_2})| n_{\mathcal{K}_{n_1-1,n_2-1}}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 2|E(\mathcal{K}_{n_1,n_2})||Q(\mathcal{K}_{n_1-1,n_2-1})| \\ &= 36\binom{n_1}{3}\binom{n_2}{3}. \end{aligned} \tag{119}$$

$\tau = 0,\ \phi = 4$ Recall equation 70: $f_{04}(\mathcal{K}_{n_1,n_2}) = 2n_{\mathcal{K}_{n_1,n_2}}(\mathcal{C}_4)$. Looking at the way of building a path of 4 vertices in figure 18(b), and noting that a cycle of 4 vertices can be obtained in 4 different ways depending on the initial vertex, we obtain

$$n_{\mathcal{K}_{n_1,n_2}}(\mathcal{C}_4) = \frac{1}{4}n_1n_2(n_1-1)(n_2-1) = \binom{n_1}{2}\binom{n_2}{2},$$

hence

$$f_{04}(\mathcal{K}_{n_1,n_2}) = 2\binom{n_1}{2}\binom{n_2}{2}. \tag{120}$$

$\tau = 0,\ \phi = 3$ By recalling equation 75: $f_{03}(\mathcal{K}_{n_1,n_2}) = 2n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_5)$. we can see that the procedure all subgraphs isomorphic to $\mathcal{L}_5$ is similar to the procedure depicted in figure 18(a), though we need to continue with two more vertices (figure 18(c)). Then

$$\begin{aligned} n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_5) &= \frac{1}{2}(n_1n_2(n_1-1)(n_2-1)(n_1-2) + n_2n_1(n_2-1)(n_1-1)(n_2-2)) \\ &= n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_3 \oplus \mathcal{L}_2). \end{aligned}$$

Therefore,

$$f_{03}(\mathcal{K}_{n_1,n_2}) = f_{13}(\mathcal{K}_{n_1,n_2}). \tag{121}$$

$\tau = 0,\ \phi = 2$ *subtype 1* Figure 18(d) illustrates how to build pairs $\mathcal{L}_4 \oplus \mathcal{L}_2$. Therefore,

$$\begin{aligned} n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_4 \oplus \mathcal{L}_2) = n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_6) &= n_1n_2(n_1-1)(n_2-1)(n_1-2)(n_2-2) \\ &= 36\binom{n_1}{3}\binom{n_2}{3}, \end{aligned}$$

hence equation 87 becomes

$$f_{021}(\mathcal{K}_{n_1,n_2}) = 72\binom{n_1}{3}\binom{n_2}{3}. \tag{122}$$

$\tau = 0$, $\phi = 2$ *subtype 2* Equation 90 tells us that this type counts the pairs of independent 3-paths. Figure 18(e) shows the procedure followed to obtain such pairs when starting at the left subset of vertices. The procedure when starting at the right subset is completely symmetric.

Therefore, since $f_{022}(\mathcal{K}_{n_1,n_2}) = 4n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_3 \oplus \mathcal{L}_3)$, and

$$\begin{aligned} n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_3 \oplus \mathcal{L}_3) &= \frac{1}{2}n_1n_2(n_1-1)\left(\frac{1}{2}(n_1-2)(n_2-1)(n_1-3) + \frac{1}{2}(n_2-1)(n_1-2)(n_2-2)\right) + \\ &+ \frac{1}{2}n_2n_1(n_2-1)\left(\frac{1}{2}(n_2-2)(n_1-1)(n_2-3) + \frac{1}{2}(n_1-1)(n_2-2)(n_1-2)\right) = \\ &= 12\binom{n_1}{2}\binom{n_2}{4} + 12\binom{n_1}{4}\binom{n_2}{2} + 18\binom{n_1}{3}\binom{n_2}{3} \end{aligned}$$

we have that

$$f_{022}(\mathcal{K}_{n_1,n_2}) = 24\binom{n_1}{2}\binom{n_2}{4} + 24\binom{n_1}{4}\binom{n_2}{2} + 36\binom{n_1}{3}\binom{n_2}{3}. \tag{123}$$

$\tau = 0$, $\phi = 1$ The paths depicted in figure 18(f) allow one to see that after removing a $\mathcal{L}_3$ we only have to count all the $\mathcal{L}_2 \oplus \mathcal{L}_2$, i.e. the size of Q of the resulting graph, yielding

$$\begin{aligned} n_{\mathcal{K}_{n_1,n_2}}(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) &= \frac{1}{2}n_1n_2(n_1-1)n_{\mathcal{K}_{n_1-2,n_2-1}}(\mathcal{L}_2 \oplus \mathcal{L}_2) + \frac{1}{2}n_2n_1(n_2-1)n_{\mathcal{K}_{n_1-1,n_2-2}}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= \frac{1}{2}n_1n_2(n_1-1)|Q(\mathcal{K}_{n_1-2,n_2-1})| + \frac{1}{2}n_2n_1(n_2-1)|Q(\mathcal{K}_{n_1-1,n_2-2})| \\ &= 36\binom{n_1}{4}\binom{n_2}{3} + 36\binom{n_1}{3}\binom{n_2}{4}. \end{aligned}$$

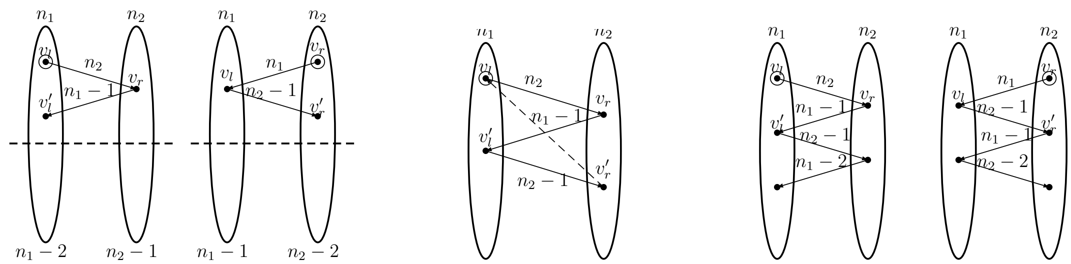
Then, equation 94 becomes

$$f_{01}(\mathcal{K}_{n_1,n_2}) = 144\binom{n_1}{4}\binom{n_2}{3} + 144\binom{n_1}{3}\binom{n_2}{4}. \tag{124}$$

5.4.1. Variance Now we are able to obtain a formula for the variance in complete bipartite graphs of n and m vertices. After several algebraic manipulations we obtain

$$\begin{aligned} \mathbb{V}\left[C(\mathcal{K}_{n_1,n_2})\right] &= \sum_{i=0}^{8} f_i(\mathcal{K}_{n_1,n_2})\mathbb{E}\left[\gamma_i\right] = \\ &= \frac{1}{90}\binom{n_1}{2}\binom{n_2}{2}((n_1+n_2)^2 + n_1 + n_2). \end{aligned} \tag{125}$$

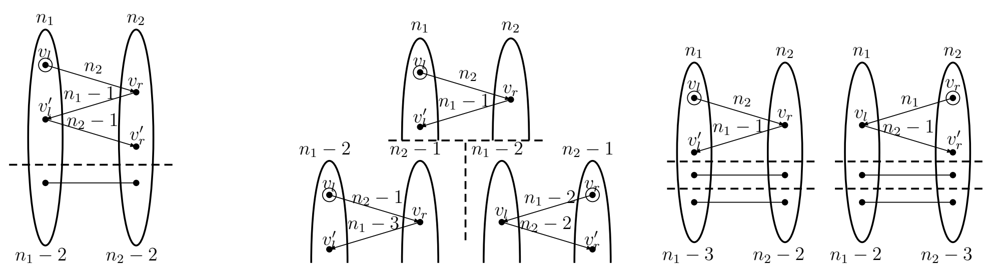
In a star tree, $\mathbb{V}[C] = C = 0$. Given that $\mathcal{S}_n = \mathcal{K}_{1,n-1}$, equation 125 gives $\mathbb{V}[C(\mathcal{S}_n)] = 0$ as expected.



(a) Pairs of independent 3-paths and 2-paths.

(b) Cycles of 4 vertices.

(c) Paths of 5 vertices.



(d) Pairs of independent 4-paths and 2-paths.

(e) Half of the pairs of independent 3-paths.

(f) Tuples of independent 3-, 2-, 2-paths.

Figure 18: Paths in $\mathcal{K}_{n_1,n_2}$. v_l, v_l', v_r and v_r' are vertices, drawn using dots. Circled dots are is used to indicate the vertex that is the start of a path.

5.5. Cycle graphs

In this section we assume a labeling of the vertices of $\mathcal{C}_n$ from 1 to n and that its set of edges is then formed by consecutively labeled vertices:

$$E(\mathcal{C}_n) = \{\{s,t\} \mid 1 \leq s,t \leq n,\ s-t \equiv 1 \mod n\}.$$

In cycle graphs $|E(\mathcal{C}_n)| = |V(\mathcal{C}_n)| = n$. The vertex labels define a circular arrangement of the vertices, namely the placement of the vertices on a circle [35].

In this section, let $e_1, e_2 \in \mathbb{N}$ be edge indices, $e_1 \neq e_2$, and $\{e_1, e_2\} \in Q$ be formed by the e_1-th and e_2-th edges.

Recall the value of $|Q(C_n)|$ in table 1. Before we start calculating the $f_i's$, we note some useful general properties. First,

$$|Q| = n_G(\mathcal{L}_2 \oplus \mathcal{L}_2) \tag{126}$$

Second,

$$n_{\mathcal{C}_{n_1}}(\mathcal{L}_{n_2}) = n_1 \tag{127}$$

for $n_2 \leq n_1$. Third,

$$n_{\mathcal{L}_{n_1}}(\mathcal{L}_{n_2}) = n_1 - n_2 + 1 \tag{128}$$

for $n_2 \leq n_1$. Finally, we define $\mathcal{C}_{n_1} \setminus \mathcal{L}_{n_2}$ as the outcome of removing some $\mathcal{L}_{n_2}$ from $\mathcal{C}_{n_1}$ with $n_2 \leq n_1$, which is a $\mathcal{L}_{n_1-n_2}$.

$\tau = 0$, $\phi = 0$ The rationale behind equation 126 allows one to express equation 60 as

$$\begin{aligned} f_{00}(\mathcal{C}_n) &= 6n_{\mathcal{C}_n}(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 6 \sum_{\{st,uv\} \in Q} n_{(\mathcal{C}_n)_{-stuv}}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 6 \sum_{\{st,uv\} \in Q} |Q((\mathcal{C}_n)_{-stuv})| \end{aligned} \quad (129)$$

Crucially, $(\mathcal{C}_n)_{-stuv}$ can only be of 3 mutually exclusive types depending on the "distance" between the vertices s, t, u, and v in the circular arrangement, i.e.

(i) $\mathcal{L}_{n-4}$, when the edges st and uv are separated by just one edge in the circular arrangement (figure 19(a)). This type needs $n \geq 5$. There are n pairs of independent edges that are separated by one edge in the circular arrangement. These are the pairs $\{e_1, e_2\}$ such that $1 \leq e_1, e_2 \leq n$, $e_2 - e_1 \equiv 2 \mod n$, i.e.

$$\{1,3\}, \{2,4\}, \{3,5\}, \cdots, \{n-3, n-1\}, \{n-2, n\}, \{n-1, 1\}, \{n, 2\}.$$

We have

$$|Q((\mathcal{C}_n)_{-stuv})| = |Q(\mathcal{L}_{n-4})| = \binom{n-6}{2}.$$

(ii) $\mathbf{0}_1 \oplus \mathcal{L}_{n-5}$, when the edges st and uv are separated by two edges in the circular arrangement (figure 19(b)). This type needs $n \geq 6$. There are n pairs of independent edges that are separated by two edges in the circular arrangement. These pairs are the $\{e_1, e_2\}$ such that $1 \leq e_1, e_2 \leq n$, $e_2 - e_1 \equiv 3 \mod n$, i.e.

$$\{1,4\}, \{2,5\}, \{3,6\}, \cdots, \{n-4, n-1\}, \{n-3, n\}, \{n-2, 1\}, \{n-1, 2\}, \{n, 3\}.$$

We have

$$|Q((\mathcal{C}_n)_{-stuv})| = |Q(\mathbf{0}_1 \oplus \mathcal{L}_{n-5})| = |Q(\mathcal{L}_{n-5})| = \binom{n-7}{2}.$$

(iii) $T_1 \oplus T_2$, where T_1 and T_2 are a couple of linear trees, $\mathcal{L}_{n_1}$ and $\mathcal{L}_{n_2}$, such that $1 \leq n_1, n_2$ and $n_1 + n_2 = n - 4$ (figure 19(c)). This type needs $n \geq 7$. The amount of pairs of edges leading to this forest of two trees is just all the elements of $Q(\mathcal{C}_n)$ except for those leading to the previous two scenarios, i.e.

$$|Q(\mathcal{C}_n)| - 2n = \frac{n(n-5)}{2}.$$

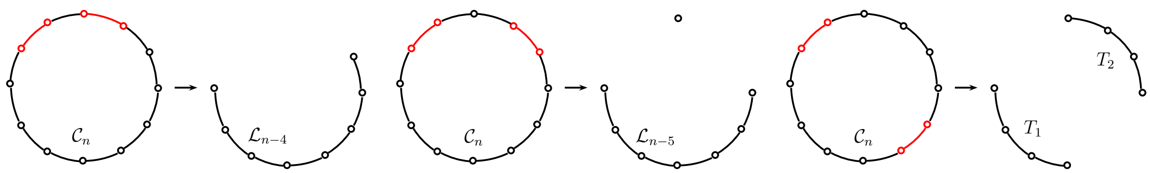
The size of $|Q((\mathcal{C}_n)_{-stuv})|$ is the same for all pairs of linear trees. As T_1 and T_2 do not share edges,

$$\begin{aligned} |Q(T_1 \oplus T_2)| &= |Q(T_1)| + |Q(T_2)| + |E(T_1)| \cdot |E(T_2)| \\ &= |Q(\mathcal{L}_{n_1})| + |Q(\mathcal{L}_{n_2})| + |E(\mathcal{L}_{n_1})| \cdot |E(\mathcal{L}_{n_2})| \\ &= \binom{n_1-2}{2} + \binom{n_2-2}{2} + (n_1-1)(n_2-1). \end{aligned}$$

The substitution $n_2 = n - 4 - n_1$, leads to

$$|Q(T_1 \oplus T_2)| = \frac{1}{2}(n-15)n + 29 \tag{130}$$

after some algebra.



(a) Case (i). (b) Case (ii). (c) Case (iii).

Figure 19: Different results obtained by removing pairs of edges from $\mathcal{C}_n$. Figure 19(a) illustrates the scenario described in item (i), figure 19(b) the scenario of item (ii), and figure 19(c) the scenario described in (iii).

As a result of the case analysis above, equation 129 can be written as

$$f_{00}(\mathcal{C}_n) = 6\left[n|Q(\mathcal{L}_{n-4})| + n|Q(\mathcal{L}_{n-5})| + \frac{n(n-5)}{2}|Q(T_1 \oplus T_2)|\right],$$

which gives

$$f_{00}(\mathcal{C}_n) = \frac{3}{2}n\binom{n-5}{3}. \tag{131}$$

after some algebra.

$\tau = 2$, $\phi = 4$ This type is trivial. By recalling equation 56

$$f_{24}(\mathcal{C}_n) = |Q(\mathcal{C}_n)| = \frac{1}{2}n(n-3). \tag{132}$$

$\tau = 1$, $\phi = 3$ Since

$$n_{\mathcal{C}_{n_1} \setminus \mathcal{L}_{n_2}}(\mathcal{L}_{n_3}) = n_{\mathcal{L}_{n-k}}(\mathcal{L}_i) = n_1 - n_2 - n_3 + 1 \tag{133}$$

then, by applying equations 133 and 127 to equation 64 we get

$$\begin{aligned} f_{13}(\mathcal{C}_n) &= 2n_{\mathcal{C}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2) = 2n_{\mathcal{C}_n}(\mathcal{L}_3)n_{\mathcal{C}_n \setminus \mathcal{L}_3}(\mathcal{L}_2) \\ &= 2n_{\mathcal{C}_n}(\mathcal{L}_3)n_{\mathcal{L}_{n-3}}(\mathcal{L}_2) = 2n(n-3-2+1) = 2n(n-4). \end{aligned}$$

Recall equation 64

$$f_{13}(\mathcal{C}_n) = 2n_{\mathcal{C}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2).$$

When a $\mathcal{L}_3$ is removed from $\mathcal{C}_n$ we obtain $\mathcal{L}_{n-3}$. Due to this, we have that:

$$f_{13}(\mathcal{C}_n) = 2n_{\mathcal{C}_n}(\mathcal{L}_3)n_{\mathcal{C}_n \setminus \mathcal{L}_3}(\mathcal{L}_2).$$

It is easy to see that $n_{\mathcal{C}_n}(\mathcal{L}_3) = n$ since the $\mathcal{L}_3$ in $\mathcal{C}_n$ are of the form

$$(s, (s+1) \bmod n, (s+2) \bmod n), \forall s \in \{1, \cdots, n\}.$$

Although $n_{\mathcal{C}_n}(\mathcal{L}_2) = n$, when removing a connected subgraph from $\mathcal{C}_n$ (any $\mathcal{L}_k$ with $k \leq n$), we have that

$$n_{\mathcal{C}_n \setminus \mathcal{L}_k}(\mathcal{L}_i) = n_{\mathcal{L}_{n-k}}(\mathcal{L}_i) = n - k - i + 1$$

for all $i \leq n - k$. Therefore, $n_{\mathcal{C}_n \setminus \mathcal{L}_3}(\mathcal{L}_2) = n - 4$, and

$$f_{13}(\mathcal{C}_n) = 2n(n-4). \tag{134}$$

$\tau = 1$, $\phi = 2$ Thanks to equation 68, we have that

$$\begin{aligned} f_{12}(\mathcal{C}_n) &= 6 n_{\mathcal{C}_n}(\mathcal{L}_2 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= \frac{6}{3} \sum_{st \in E} n_{(\mathcal{C}_n)-st}(\mathcal{L}_2 \oplus \mathcal{L}_2) \end{aligned} \tag{135}$$

As removing an edge from $\mathcal{C}_n$ produces $\mathcal{L}_{n-2}$, the rationale behind equation 126 leads to

$$\begin{aligned} f_{12}(\mathcal{C}_n) &= 2n \cdot n_{\mathcal{L}_{n-2}}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 2n|Q(\mathcal{L}_{n-2})| \\ &= n(n-4)(n-5) \end{aligned} \tag{136}$$

after some algebra.

$\tau = 0$, $\phi = 4$ Applying the fact that $n_{\mathcal{C}_n}(\mathcal{C}_4)$ is 1 if $n = 4$ and 0 otherwise to equation 70, produces

$$f_{04}(\mathcal{C}_n) = \begin{cases} 2 & \text{if } n = 4 \\ 0 & \text{otherwise.} \end{cases} \tag{137}$$

$\tau = 0$, $\phi = 3$ Applying equation 127) with $n_2 = 5$ to equation 75, we obtain

$$f_{03}(\mathcal{C}_n) = 2n. \tag{138}$$

$\tau = 0$, $\phi = 2$, *subtype 1* Thanks to equation 87,

$$f_{021}(\mathcal{C}_n) = 2n_{\mathcal{C}_n}\mathcal{L}_4 \oplus \mathcal{L}_2 = 2n_{\mathcal{C}_n}(\mathcal{L}_4) n_{\mathcal{L}_{n-4}}(\mathcal{L}_2).$$

As $n_{\mathcal{C}_n}(\mathcal{L}_4) = n$ (equation 127) and $n_{\mathcal{L}_{n-4}}(\mathcal{L}_2) = n - 5$ (equation 128), we finally obtain

$$f_{021}(\mathcal{C}_n) = 2n(n-5). \tag{139}$$

$\tau = 0$, $\phi = 2$, *subtype 2* Following the procedure for type 021, equation 90 leads to

$$f_{022}(\mathcal{C}_n) = 4n_{\mathcal{C}_n}(\mathcal{L}_3 \oplus \mathcal{L}_3) = 4 \cdot \frac{1}{2} n_{\mathcal{C}_n}(\mathcal{L}_3) n_{\mathcal{L}_{n-3}}(\mathcal{L}_3) = 2n(n-5). \tag{140}$$

Notice that $f_{021}(\mathcal{C}_n) = f_{022}(\mathcal{C}_n)$.

$\tau = 0,\ \phi = 1$ Equation 94, leads to

$$\begin{aligned} f_{01}(\mathcal{C}_n) &= 4n_{\mathcal{C}_n}(\mathcal{L}_3 \oplus \mathcal{L}_2 \oplus \mathcal{L}_2) = 4\sum_{\mathcal{L}_3} n_{\mathcal{C}_n \setminus \mathcal{L}_3}(\mathcal{L}_2 \oplus \mathcal{L}_2) \\ &= 4n_{\mathcal{C}_n}(\mathcal{L}_3)|Q(\mathcal{L}_{n-3})| = 4n\binom{n-5}{2} \\ &= 2n(n-5)(n-6). \end{aligned}$$

5.5.1. Variance Now we are able to obtain a formula for the variance in cycle graphs:

$$\mathbb{V}\left[C(\mathcal{C}_n)\right] = \sum_{i=0}^{8} f_i \mathbb{E}\left[\gamma_i\right] = \frac{1}{45}n^3 + \frac{1}{90}n^2 - \frac{1}{3}n.$$

5.5.2. Notes on the occurrences of edges In this section we study the amount of times an edge e is involved in the elements of $Q(G) \times Q(G)$ of a certain type i. With this information, obtaining the $f_i(\mathcal{L}_n)$ is done in a much more systematic way, with the help of the fact that a linear tree is obtained by removing one edge from a cycle graph. First, for a simple graph G we define the set of elements of $Q(G) \times Q(G)$ of a certain type i, $\rho_i(G) \subseteq Q(G) \times Q(G)$, and the set of elements of $Q(G) \times Q(G)$ of a certain type i that contain a certain edge $e \in E$, $\rho_i(G, e) \subseteq Q(G) \times Q(G)$. Then

$$\begin{aligned} \rho_i(G) &= \{(q_1, q_2) \in Q(G) \times Q(G) \mid \mathcal{T}(q_1, q_2) = i\}, \\ \rho_i(G, e) &= \{(q_1, q_2) \in \rho_i(G) \mid e \in q_1 \cup q_2\}. \end{aligned} \tag{141}$$

Notice

$$|\rho_i(G, e)| = \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q\,: \\ e \in q_1 \cup q_2 \\ \mathcal{T}(q_1, q_2) = i}} 1.$$

Interestingly,

$$\sum_{e \in E} |\rho_i(G, e)| = c_i f_i(G). \tag{142}$$

Now we prove this claim. Recall the definition of the f_i in equation 39. Let us change it slightly

$$f_i = \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q\,: \\ \mathcal{T}(q_1, q_2) = i}} \frac{1}{c_i(q_1, q_2)} \sum_{\substack{e \in E\,: \\ e \in q_1 \cup q_2}} 1,$$

where

$$c_i(q_1, q_2) = \sum_{\substack{e \in E\,: \\ e \in q_1 \cup q_2 \\ \mathcal{T}(q_1 \cup q_2) = i}} 1.$$

Therefore, $c_i = c_i(q_1, q_2) = |q_1 \cup q_2| = 4 - \tau_i$ namely the number of distinct edges of any element of $Q \times Q$ of type i. Simply,

$$c_{00} = c_{01} = c_{021} = c_{022} = c_{03} = c_{04} = 4, \quad c_{12} = c_{13} = 3, \quad c_{24} = 2. \tag{143}$$

By floating the inner-most summation we obtain

$$f_i = \frac{1}{c_i} \sum_{e \in E} \sum_{q_1 \in Q} \sum_{\substack{q_2 \in Q: \\ e \in q_1 \cup q_2 \\ \mathcal{T}(q_1,q_2)=i}} 1$$
$$= \frac{1}{c_i} \sum_{e \in E} |\rho_i(G,e)|,$$

hence equation 142.

Whereas equation 142 is true for general graphs, in a cycle graph one has

$$|\rho_i(\mathcal{C}_n, e_1)| = |\rho_i(\mathcal{C}_n, e_2)| = \cdots = |\rho_i(\mathcal{C}_n, e_n)|. \tag{144}$$

In words, the amount of occurrences of an edge of $\mathcal{C}_n$ in the elements of $Q(\mathcal{C}_n) \times Q(\mathcal{C}_n)$ is the same independently of the edge. This is true due to the graph's structure, i.e., the properties of one edge are the same as any other edges' properties. We use $|\rho_i(\mathcal{C}_n, e_*)|$ to denote any of the $|\rho_i(\mathcal{C}_n, e_j)|$, where e_* denotes any edge and $1 \leq j \leq n$. Thus equation 142 becomes

$$\sum_{e \in E(\mathcal{C}_n)} |\rho_i(\mathcal{C}_n, e)| = n|\rho_i(\mathcal{C}_n, e_*)| = c_i f_i(G).$$

Finally,

$$|\rho_i(\mathcal{C}_n, e_*)| = \frac{c_i}{n} f_i(\mathcal{C}_n). \tag{145}$$

Equation 144 may also hold for other types of graphs, e.g. regular graphs, but such analysis is beyond the scope of the present article.

5.6. Linear trees

In order to calculate the $f_i(\mathcal{L}_n)$ for every i we apply the following strategy, whenever necessary:

(i) Convert $\mathcal{L}_n$ into a cycle graph $\mathcal{C}_n$ by joining the two leaves of $\mathcal{L}_n$. Let e_n be the new edge joining the two leaves.

(ii) Compute the value of $f_i(\mathcal{C}_n)$.

(iii) Substract from $f_i(\mathcal{C}_n)$ the amount of elements of $Q(\mathcal{C}_n) \times Q(\mathcal{C}_n)$ classified as type i in which e_n appears, namely $|\rho_i(\mathcal{C}_n, e_n)|$. Thus, using equation 145, the case of linear trees is simple to solve. We get:

$$f_i(\mathcal{L}_n) = f_i(\mathcal{C}_n) - |\rho_i(\mathcal{C}_n, e_n)| = f_i(\mathcal{C}_n) - \frac{c_i}{n} f_i(\mathcal{C}_n). \tag{146}$$

where the values of c_i are given in equation 143.

$\tau = 0$, $\phi = 0$

$$f_{00}(\mathcal{L}_n) = f_{00}(\mathcal{C}_n) - \frac{4}{n} f_{00}(\mathcal{C}_n) = 6\binom{n-4}{4}. \tag{147}$$

$\tau = 2,\ \phi = 4$ This is trivially obtained via equation 56

$$f_{24}(\mathcal{L}_n) = |Q(\mathcal{L}_n)| = \binom{n-2}{2}. \tag{148}$$

$\tau = 1,\ \phi = 3$

$$f_{13}(\mathcal{L}_n) = f_{13}(\mathcal{C}_n) - \frac{3}{n} f_{13}(\mathcal{C}_n) = 4\binom{n-3}{2}. \tag{149}$$

$\tau = 1,\ \phi = 2$

$$f_{12}(\mathcal{L}_n) = f_{12}(\mathcal{C}_n) - \frac{3}{n} f_{12}(\mathcal{C}_n) = 6\binom{n-3}{3}. \tag{150}$$

$\tau = 0,\ \phi = 4$ This is trivially 0 because trees are not cyclic

$$f_{04}(\mathcal{L}_n) = 0. \tag{151}$$

$\tau = 0,\ \phi = 3$

$$f_{03}(\mathcal{L}_n) = f_{03}(\mathcal{C}_n) - \frac{4}{n} f_{03}(\mathcal{C}_n) = 2n - 8. \tag{152}$$

$\tau = 0,\ \phi = 2$

$$f_{021}(\mathcal{L}_n) = f_{021}(\mathcal{C}_n) - \frac{4}{n} f_{021}(\mathcal{C}_n) = 4\binom{n-4}{2}. \tag{153}$$

Since $f_{021}(\mathcal{C}_n) = f_{022}(\mathcal{C}_n)$

$$f_{022}(\mathcal{L}_n) = f_{021}(\mathcal{L}_n) = 4\binom{n-4}{2}. \tag{154}$$

$\tau = 0,\ \phi = 1$

$$f_{01}(\mathcal{L}_n) = f_{01}(\mathcal{C}_n) - \frac{4}{n} f_{01}(\mathcal{C}_n) = 12\binom{n-4}{3}. \tag{155}$$

5.6.1. Variance The variance in linear trees is

$$\mathbb{V}\left[C(\mathcal{L}_n)\right] = \sum_{i=0}^{8} f_i(\mathcal{L}_n)\mathbb{E}\left[\gamma_i\right] = \frac{1}{45}n^3 - \frac{1}{18}n^2 - \frac{11}{45}n + \frac{2}{3}.$$

5.7. The scaling of $\mathbb{V}[C]$ as function of n.

Figure 20 shows $\mathbb{V}[C]$ as a function of n for the special graphs where $\mathbb{V}[C]$ depends only on the numbre of vertices of the graph. According to Table 1, $\mathbb{V}[C]$ is expected to scale as $\sim n^{\gamma}$, with $\gamma = 2$ for quasi-star trees, $\gamma = 3$ for 1-regular graphs and cycles, and $\gamma = 4$ for linear trees. Figure 20 validates the theoretical $\mathbb{V}[C]$ by means of numerical estimates. The testing protocol for $\mathbb{V}[C]$ is explained in Appendix C.

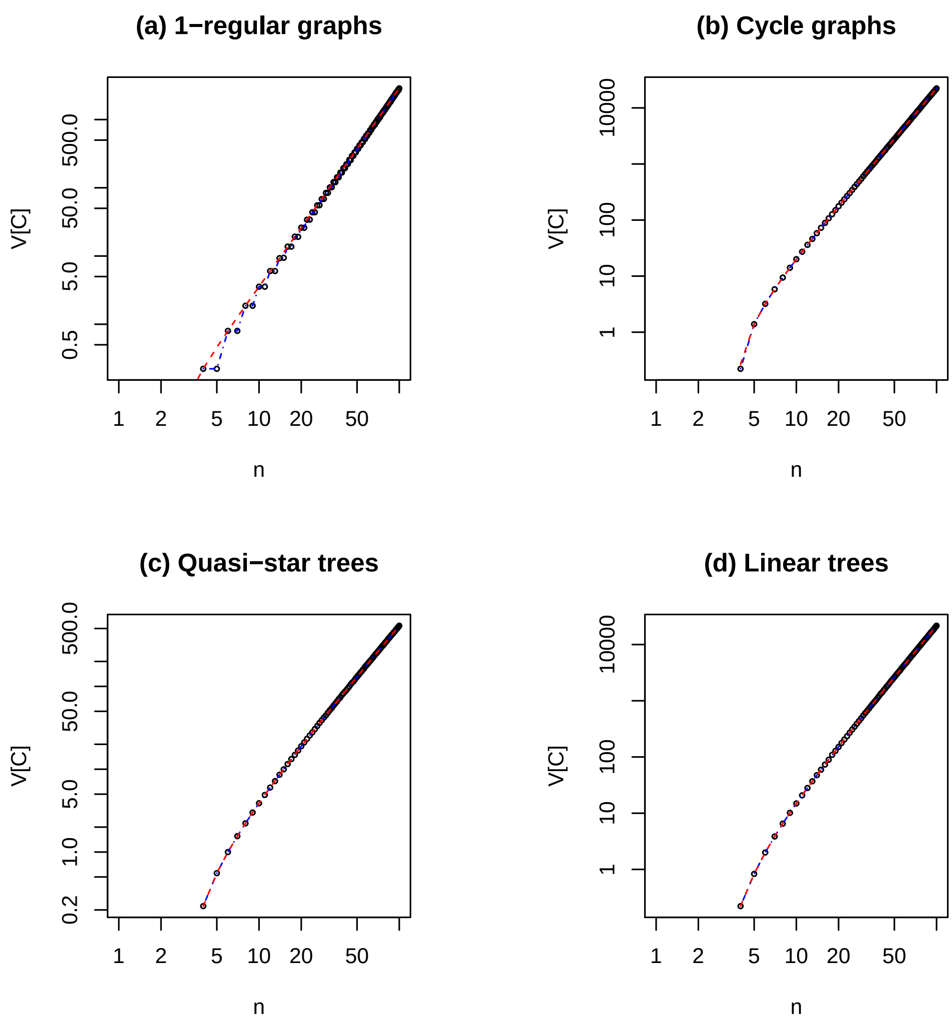


Figure 20: The variance of C, the number of crossings, as a function of n, the number of vertices of the graph. For every value n, the variance of C is estimated over T different random linear arrangements ($T = n!$ for $n \leq 10$, then $T = 10^5$ for $n \geq 11$). When $T = n!$, the variance matches $\mathbb{V}[C]$. $\mathbb{V}[C]$, the theoretical variance is calculated in two ways: Table 1 (red dashed line) and the general formula for $\mathbb{V}[C]$ where the value of every f_i is calculated with a brute force algorithm (blue dashed line). As the scale is log-log, values with $n < 4$ are not shown because $\mathbb{V}[C] = 0$.

5.8. Summary

Table 6 summarizes the formulae for the number of products of each type as a function of n that have been obtained in the preceding sections for particular graphs. See table 2 to see the valid values of the parameter n for each f_i. In particular, $n \geq |\nu|$ is needed. All variances are 0 for $n \leq 3$ and all the expressions are valid for $n \geq 4$, with the exception of $\mathbb{V}\left[\mathcal{C}_4\right] = 2/9$, as detailed in the table. In the case of $\mathcal{K}_{n_1,n_2}$, the expression for the variance is valid for values $n_1, n_2 \geq 2$ and for $n_1, n_2 < 2$ the variance is 0.

Table 6: Expressions for the f_i's and variance on each type of graph as a function of n, the number of vertices. The expressions of the f_i's are valid for values of $n \geq |\nu|$ (see table 2 for details on the values of $|\nu|$). All variances are 0 for $n \leq 3$. In the case of $\mathcal{K}_{n_1,n_2}$ the variance is 0 when $n_1, n_2 < 2$.

f_i	$\mathbf{1}_n$	$\mathcal{Q}_n$	$\mathcal{K}_n$	$\mathcal{K}_{n_1,n_2}$	$\mathcal{C}_n$	$\mathcal{L}_n$
f_{00}	$6\binom{n/2}{4}$	0	$630\binom{n}{8}$	$144\binom{n_1}{4}\binom{n_2}{4}$	$\frac{3}{2}n\binom{n-5}{3}$	$6\binom{n-4}{4}$
f_{24}	$\binom{n/2}{2}$	$n-3$	$3\binom{n}{4}$	$2\binom{n_1}{2}\binom{n_2}{2}$	$\frac{1}{2}n(n-3)$	$\binom{n-2}{2}$
f_{13}	0	$(n-3)(n-4)$	$60\binom{n}{5}$	$12\binom{n_1}{3}\binom{n_2}{2} + 12\binom{n_1}{2}\binom{n_2}{3}$	$2n(n-4)$	$4\binom{n-3}{2}$
f_{12}	$6\binom{n/2}{3}$	0	$90\binom{n}{6}$	$36\binom{n_1}{3}\binom{n_2}{3}$	$n(n-4)(n-5)$	$6\binom{n-3}{3}$
f_{04}	0	0	$6\binom{n}{4}$	$2\binom{n_1}{2}\binom{n_2}{2}$	0 (except 2 for $n=4$)	0
f_{03}	0	0	$120\binom{n}{5}$	$12\binom{n_1}{3}\binom{n_2}{2} + 12\binom{n_1}{2}\binom{n_2}{3}$	$2n$	$2n-8$
f_{021}	0	0	$360\binom{n}{6}$	$72\binom{n_1}{3}\binom{n_2}{3}$	$2n(n-5)$	$4\binom{n-4}{2}$
f_{022}	0	0	$360\binom{n}{6}$	$24\binom{n_1}{2}\binom{n_2}{4} + 24\binom{n_1}{4}\binom{n_2}{2} + 36\binom{n_1}{3}\binom{n_2}{3}$	$2n(n-5)$	$4\binom{n-4}{2}$
f_{01}	0	0	$1260\binom{n}{7}$	$144\binom{n_1}{4}\binom{n_2}{3} + 144\binom{n_1}{3}\binom{n_2}{4}$	$2n(n-5)(n-6)$	$12\binom{n-4}{3}$
$\mathbb{V}[C]$	$\frac{1}{360}(n-2)n(n+6)$	$\frac{1}{18}n(n-3)$	0	$\frac{1}{90}\binom{n_1}{2}\binom{n_2}{2}((n_1+n_2)^2+n_1+n_2)$	$\mathbb{V}[\mathcal{C}_n] = 2/9$ (for $n=4$) $\frac{1}{45}n^3 + \frac{1}{90}n^2 - \frac{1}{3}n$ (for $n>4$)	$\frac{1}{45}n^3 - \frac{1}{18}n^2 - \frac{11}{45}n + \frac{2}{3}$

6. Discussion

In the preceding sections we have investigated the expectation and the variance of C in random linear arrangements in general graphs, given the structure of concrete graph and we have provided compact formulae for specific graphs (Table 1). The scaling of the expectation and the variance of C as function of the number of vertices is asymptotically power-law-like (figure 4 and 20 and also Table 1).

$\mathbb{V}[C]$ turns out to be a weighted sum of the number of products of each type and their probabilities. We have shown that the number of products of a given type is proportional to the number subgraphs of a certain kind and obtained general alternative general formulae 4. We have also provided simple exact formulae for these numbers in particular graphs (table 6). As a side-effect of such characterization, we have contributed with expressions for the number of paths of length 5 of a graph (section 4.5.6) that are more compact than the one obtained in previous work [34]. Future work should investigate simple formulae for the number of products of a type, specially types that are relevant for the calculation of variance but for whom a simple alternative formula is not forthcoming yet, i.e. types 03, 021 i 022 (Table 4). Our exploration has focused the mathematical aspects. Fast algorithms for the calculation of the number of products of each type or the whole variance should be investigated.

Our analysis on individual graphs has paved the way to study the distribution of crossings for classic ensembles of graphs, e.g., Erdős-Rényi graphs [36] and uniformly random trees [37, 38], as well as other classes of random networks with more realistic characteristics [39, 3].

In previous work on syntactic dependency trees, C has been shown to be significantly low with respect to random linear arrangements with the help of Monte Carlo statistical tests [3]. Thanks to our novel knowledge about the expectation and the variance of C in these arrangements, fast statistical tests of the number of crossings of graphs could be derived with the help of Chebychev-like inequalities to linguistic and biological networks were vertices form a 1-dimensional lattice [5, 3]. Such a procedure has already been outlined to check if D, the sum of edge lengths in a linear arrangement, is significantly low with respect to random linear arrangements [8].

Our algebraic characterization of the types of products in $\mathbb{V}[C]$ (figure 5 and table 2) was motivated by spatial networks in 1-dimensional lattices but it has been derived independently from that layout based on purely algebraic criteria. Therefore we have have established the foundations to study $\mathbb{V}[C]$ in other layouts or embeddings, e.g., lattices of 2 or 3 dimensions. In addition, we have also established the foundations to revise previous work in other scenarios. In section 4 we adapted a derivation of $\mathbb{V}[C]$ by Moon [20] for the particular case of a complete graph whose vertices are arranged at random on the surface of a sphere. In the process, we discovered that some types of products that we have identified were missing in the original derivation. Indeed, Moon's derivation for the spherical case is inaccurate and should be corrected [40].

Acknowledgements

We are grateful to Kosmas Palios for helpful discussions and Antoni Lozano for advice on notation. We also thank Merc Mora for helpful suggestions to improve the quality of the manuscript. This research was supported by the grant TIN2017-89244-R from MINECO (Ministerio de Economía y Competitividad) and the recognition 2017SGR-856 (MACDA) from AGAUR (Generalitat de Catalunya).

Appendix A. Potential number of crossings of an edge

The potential number of crossings involving the edge formed by u and v can be defined as

$$q(u,v) = \frac{1}{2} \sum_{s \neq u,v} \sum_{t \neq u,v} a_{st}.$$

Noting that

$$\sum_{s \neq u,v} k_s = 2m - k_u - k_v$$
$$a_{uu}, a_{vv} = 0$$
$$a_{uv} = 1$$

and applying iteratively

$$\sum_{s \neq u,v} a_{st} = k_t - a_{tu} - a_{tv},$$

it is easy to see that $q(u,v)$ can be expressed as

$$\begin{aligned} q(u,v) &= \frac{1}{2} \sum_{s \neq u,v} (k_s - a_{su} - a_{sv}) \\ &= \frac{1}{2} \left[\sum_{s \neq u,v} k_s - \sum_{s \neq u,v} a_{su} - \sum_{s \neq u,v} a_{sv} \right] \\ &= \frac{1}{2} \left[2m - k_u - k_v - (k_u - a_{uu} - a_{uv}) - (k_v - a_{vv} - a_{uv}) \right] \\ &= m - k_u - k_v + 1. \end{aligned}$$

Appendix B. Toy examples

First we consider all trees up to $n = 5$. Suppose that linear trees are labeled with integers starting at one in one leaf and increasing one unit per vertex till the other leaf (figure B1). When $n < 4$, $\mathbb{V}[C] = 0$ trivially because $|Q| = 0$ and then $C = 0$. The same happens to all star trees with $n > 4$. When $n = 4$ the only trees are the star tree and the linear tree.

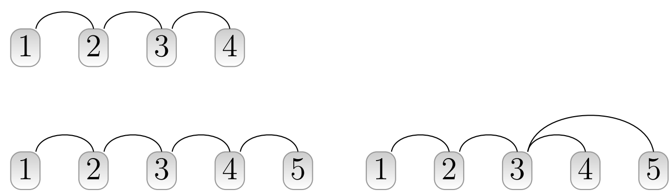


Figure B1: Some labeled trees.

4 vertices

Here, $\langle k^2 \rangle = 5/2$ and then $|Q| = 1$ (recall equation 1). For the linear tree with $n = 4$ in figure B1, we have

$$Q = \{\{12, 34\}\}.$$

Since the graphs used as examples are small we denote edges as pairs of integers. Then, the pair $\{12, 34\}$ represents the pair of edges $\{1, 2\}$ and $\{3, 4\}$. As $Q \times Q = \{\{\{12, 34\}, \{12, 34\}\}\}$, $f_{24} = 1$, and $f_i = 0$ for the other types, and then

$$\mathbb{V}[C] = \sum_{i=0}^{8} \mathbb{E}[\gamma_i] = \mathbb{E}[\gamma_{24}] = \frac{2}{9}.$$

$\mathbb{V}[C]$ could have been derived also as the variance of C, a Bernoulli variable because $C \in \{0, 1\}$ due to $|Q| = 1$. As $p(C = 1) = 1/3$, namely the probability that two independent edges cross,

$$\mathbb{V}[C] = p(C = 1)(1 - p(C = 1)) = \frac{2}{9}.$$

When $n = 4$, we conclude that $\mathbb{V}[C] = 0$ if the tree is a star or $\mathbb{V}[C] = 2/9$ if the tree is linear.

Table B1: Types of products in $Q \times Q$ for a linear tree with $n = 5$.

	$\{12, 34\}$	$\{12, 45\}$	$\{23, 45\}$
$\{12, 34\}$	24	13	03
$\{12, 45\}$	13	24	13
$\{23, 45\}$	03	13	24

Table B2: Types of products in $Q \times Q$ for a quasi-star tree with $n = 5$.

	$\{12, 34\}$	$\{12, 35\}$
$\{12, 34\}$	24	13
$\{12, 35\}$	13	24

5 vertices

When $n = 5$, the only possible trees are linear, quasi-star [15] and star (figure B1). As for the linear tree, applying $\langle k^2 \rangle = 14/5$ (equation 26) to equation 1 gives $|Q| = 3$. Following the labeling in figure B1, it is easy to see that

$$Q = \{\{12, 34\}, \{12, 45\}, \{23, 45\}\}.$$

The types of products in $Q \times Q$ in Table B1 give $f_{24} = 3$, $f_{13} = 4$, $f_{03} = 2$, $f_i = 0$ for the other types, and then

$$\mathbb{V}[C] = 3\mathbb{E}[\gamma_{24}] + 4\mathbb{E}[\gamma_{13}] + 2\mathbb{E}[\gamma_{03}] = 3\frac{2}{9} + 4\frac{1}{18} - 2\frac{1}{36} = \frac{5}{6}.$$

As for the quasi-star tree, we have [15]

$$\langle k^2 \rangle = n - 3 + 6/n.$$

Applying $\langle k^2 \rangle = 16/5$ to equation 1 gives $|Q| = 2$. Following the labeling in figure B1, it is easy to see that

$$Q = \{\{12, 34\}, \{12, 35\}\}.$$

The summary of types of products in $Q \times Q$ in Table B2 gives $f_{24} = f_{13} = 2$ and then

$$\mathbb{V}[C] = 2\mathbb{E}[\gamma_{24}] + 2\mathbb{E}[\gamma_{13}] = 2\left(\frac{2}{9} + \frac{1}{18}\right) = \frac{5}{9}.$$

When $n = 5$, we conclude that:

- $\mathbb{V}[C] = 0$ if the tree is a star
- $\mathbb{V}[C] = 5/9$ if the tree is a quasi-star
- $\mathbb{V}[C] = 5/6$ if the tree is linear

6 vertices

Here we only consider linear trees with $n = 6$.

The second moment degree of linear trees for $n = 6$ gives $\langle k^2 \rangle = 3$ (equation 26). Therefore $|Q| = 6$ (equation 1). Indeed, by following the same labeling method, we see that:

$$Q = \{\{12, 34\}, \{12, 45\}, \{12, 56\}, \{23, 45\}, \{23, 56\}, \{34, 56\}\}.$$

The summary of types of products in $Q \times Q$ is shown in table B3.

Table B3: Types of products in $Q \times Q$ for a quasi-star tree with $n = 6$.

	{12, 34}	{12, 45}	{12, 56}	{23, 45}	{23, 56}	{34, 56}
{12, 34}	24	13	12	03	021	12
{12, 45}	13	24	13	13	022	022
{12, 56}	12	13	24	022	13	12
{23, 45}	03	13	022	24	13	03
{23, 56}	021	022	13	13	24	13
{34, 56}	12	021	12	03	13	24

Table B4: Types of products in $Q \times Q$ for a linear tree with $n = 7$.

	$\{12,34\}$	$\{12,45\}$	$\{12,56\}$	$\{12,67\}$	$\{23,45\}$	$\{23,56\}$	$\{23,67\}$	$\{34,56\}$	$\{34,67\}$	$\{45,67\}$
$\{12,34\}$	24	13	12	12	03	021	021	12	12	01
$\{12,45\}$	13	24	13	12	13	022	01	021	01	12
$\{12,56\}$	12	13	24	13	022	13	022	12	01	021
$\{12,67\}$	12	12	13	24	01	022	13	01	12	12
$\{23,45\}$	03	13	022	01	24	13	12	03	021	12
$\{23,56\}$	021	022	13	022	13	24	13	13	022	021
$\{23,67\}$	021	01	022	13	12	13	24	022	13	12
$\{34,56\}$	12	021	12	01	03	03	022	24	13	03
$\{34,67\}$	12	01	01	12	021	022	13	13	24	13
$\{45,67\}$	01	12	021	12	12	021	12	03	13	24

The summary of types of products in $Q \times Q$ in Table B3 gives $f_{24} = f_{12} = 6$, $f_{13} = 12$, $f_{03} = f_{021} = f_{022} = 4$ and then

$$\begin{aligned}\mathbb{V}[C] &= 6(\mathbb{E}[\gamma_{24}] + \mathbb{E}[\gamma_{12}]) + 12\mathbb{E}[\gamma_{13}] + 4(\mathbb{E}[\gamma_{03}] + \mathbb{E}[\gamma_{021}] + \mathbb{E}[\gamma_{022}]) \\ &= 6\left(\frac{2}{9} + \frac{1}{45}\right) + 12\frac{1}{18} + 4\left(-\frac{1}{36} - \frac{1}{90} + \frac{1}{180}\right) = 2.\end{aligned}$$

7 vertices

Here we only consider linear trees with $n = 7$.

The second moment of degree of linear trees for $n = 7$ gives $\langle k^2 \rangle = 22/7$. Therefore $|Q| = 10$ (see equation 1). Indeed, by following the same labeling method, we see that:

$$\begin{aligned}Q = \{&\{12,34\}, \{12,45\}, \{12,56\}, \{12,67\}, \{23,45\}, \\ &\{23,56\}, \{23,67\}, \{34,56\}, \{34,67\}, \{45,67\}\}.\end{aligned}$$

The summary of types of products in $Q \times Q$ is shown in table B4, and gives $f_{24} = 10$, $f_{13} = f_{12} = 24$, $f_{03} = 6$, $f_{021} = f_{022} = f_{01} = 12$, and $f_{00} = f_{24} = 0$, and then

$$\begin{aligned}\mathbb{V}[C] &= 10\mathbb{E}[\gamma_{24}] + 24(\mathbb{E}[\gamma_{13}] + \mathbb{E}[\gamma_{12}]) + 6\mathbb{E}[\gamma_{03}] + 12(\mathbb{E}[\gamma_{021}] + \mathbb{E}[\gamma_{022}] + \mathbb{E}[\gamma_{01}]) \\ &= 10\frac{2}{9} + 24\left(\frac{1}{18} + \frac{1}{45}\right) + 6\left(-\frac{1}{36}\right) + 12\left(-\frac{1}{90} + \frac{1}{180}\right) = \frac{347}{90}\end{aligned}$$

Complete graphs

Finally, we consider complete graphs with $n = 4$ and $n = 5$. We know that $\mathbb{V}[C] = 0$. The case is not only interesting due to the cancellation of the $f_i\mathbb{E}[\gamma_i]$'s but also for showing products of type 4 in $Q \times Q$, which cannot be found in trees.

When $n = 4$, we have $|Q| = 3$ (recall equation 15) with

$$Q = \{\{12,34\}, \{13,24\}, \{14,23\}\}.$$

The types of products in Table B5 give

$$\mathbb{V}[C] = 3\mathbb{E}[\gamma_{24}] + 6\mathbb{E}[\gamma_{04}] = 3\frac{2}{9} - 6\frac{1}{9} = 0$$

as expected.

Table B5: Types of products in $Q \times Q$ for a complete graph with $n = 4$.

	$\{12, 34\}$	$\{13, 24\}$	$\{14, 23\}$
$\{12, 34\}$	24	04	04
$\{12, 24\}$	04	24	04
$\{14, 23\}$	04	04	24

When $n = 5$, we have $|Q| = 15$ (recall equation 15) with

$$Q = \{\{12, 34\}, \{12, 35\}, \{12, 45\}, \{13, 24\}, \{13, 25\}, \{13, 45\}, \{14, 23\}, \{14, 25\}, \{14, 35\}, \{15, 23\}, \{15, 24\}, \{15, 34\}, \{23, 45\}, \{24, 35\}, \{25, 34\}\}.$$

The types of products in Table B6 give $f_{24} = 15$, $f_{13} = 60$, $f_{04} = 30$, $f_{03} = 120$, $f_{00} = f_{12} = f_{021} = f_{022} = f_{01} = 0$, and then

$$\begin{aligned} \mathbb{V}[C] &= 15\mathbb{E}[\gamma_{24}] + 60\mathbb{E}[\gamma_{13}] + 30\mathbb{E}[\gamma_{04}] + 120\mathbb{E}[\gamma_{03}] \\ &= 15\frac{2}{9} + 60\frac{1}{18} - 30\frac{1}{9} - 120\frac{1}{36} = 0 \end{aligned}$$

as expected.

In all the examples above we have obtained all the types of products that can appear when $n < 6$, namely types 1,2,4 and 5. $n = 6$ is needed to obtain a combination of type 2. $n = 8$ is needed to obtain a combination of type 0.

Appendix C. Protocol for testing

The formulae for $\mathbb{E}[C]$ are mathematically trivial and easy to compute. In contrast, the formulae and the algorithms for calculating $\mathbb{V}[C]$ (equation 42) are complex and require a validation protocol. That protocol is inspired by that of [10] and consists of two types of tests: computational tests and manual mathematical tests. Computational tests take a certain class of graphs and, for each graph in the class, they calculate $\mathbb{V}[C]$ following two independent procedures. First, the f_i's are calculated by brute force with a general algorithm and then plugged into equation 42 to produce $\mathbb{V}[C]'$. Second, $\mathbb{V}[C]$ is estimated over the $n!$ permutations of the vertices of the graph producing $\mathbb{V}[C]''$. A key point here is that $\mathbb{V}[C]$ has to be calculated via the biased estimator as no sampling bias is possible (the customary biased estimator yields wrong results). $\mathbb{V}[C]'$ and $\mathbb{V}[C]''$ are rational numbers that are simplified so that they can be compared easily. The test is successful if $\mathbb{V}[C]' = \mathbb{V}[C]''$. The rational numbers were represented, simplified and compared using the GMP Library (version 6.1.2, see `https://gmplib.org/`). Alternatively, real numbers for $\mathbb{V}[C]'$ and $\mathbb{V}[C]''$ could have been used but that would have required defining an error threshold to decide if the two independent calculations yield the the same result or not.

Table B6: Types of products in $Q \times Q$ for a complete graph with $n = 5$.

	{12, 34}	{12, 35}	{12, 45}	{13, 24}	{13, 25}	{13, 45}	{14, 23}	{14, 25}
{12, 34}	24	13	13	04	03	03	04	03
{12, 35}	13	24	13	03	04	03	03	03
{12, 45}	13	13	24	03	03	13	03	04
{13, 24}	04	03	03	24	13	13	04	03
{13, 25}	03	04	03	13	24	13	03	13
{13, 45}	03	03	13	13	13	24	03	03
{14, 23}	04	03	03	04	03	03	24	13
{14, 25}	03	03	04	03	13	03	13	24
{14, 35}	03	13	03	03	03	04	13	13
{15, 23}	03	04	03	03	04	03	13	03
{15, 24}	03	03	04	13	03	03	03	04
{15, 34}	13	03	03	03	03	04	03	03
{23, 45}	03	03	13	03	03	13	13	03
{24, 35}	03	13	03	13	03	03	03	03
{25, 34}	13	03	03	03	13	03	03	13

	{14, 35}	{15, 23}	{15, 24}	{15, 34}	{23, 45}	{24, 35}	{25, 34}
{12, 34}	03	03	03	13	03	03	13
{12, 35}	13	04	03	03	03	13	03
{12, 45}	03	03	04	03	13	03	03
{13, 24}	03	03	13	03	03	13	03
{13, 25}	03	04	03	03	03	03	13
{13, 45}	04	03	03	04	13	03	03
{14, 23}	13	13	03	03	13	03	03
{14, 25}	13	03	04	03	03	03	13
{14, 35}	24	03	03	04	03	13	03
{15, 23}	03	24	13	13	13	03	03
{15, 24}	03	13	24	13	03	13	03
{15, 34}	04	13	13	24	03	03	13
{23, 45}	03	13	03	03	24	04	04
{24, 35}	13	03	13	03	04	24	04
{25, 34}	03	03	03	13	04	04	24

The computational tests were applied to the following classes of graphs taking every value of n within the interval $[2, n_{max}]$:

- All labeled undirected trees of n vertices. The n^{n-2} labeled undirected trees of n vertices [41] were generated using of Prüfer codes [42], with the algorithm described in [43] (Chapter 3, section 3.3.4). $n_{max} = 9$ was used.
- Representatives of isomorphic classes of graphs [44]. An undirected graph of n vertices without loops is defined simply by a triangle of the $n \times n$ adjacency matrix that has $\binom{n}{2}$ cells. Therefore, the space of potential graphs of n vertices is $2^{\binom{n}{2}}$. To reduce the cost of testing, we consider a smaller space defined by representatives of isomorphic classes of [44]. These representatives were downloaded from a database

https://users.cecs.anu.edu.au/~bdm/data/graphs.html. $n_{max} = 9$ was used.

- The specific kinds of graphs listed introduced in section 2.5: 1-regular graphs, cycle graphs, quasi-star trees and linear trees. Their simplicity allows one to test larger values of n compared to the preceding classes of graphs.

For the special kinds of graphs, the computational testing procedure was extended. First, $\mathbb{V}[C]''$ was not calculated exactly when $n \geq n_{MC} = 11$, but estimated using a Monte Carlo procedure over 10^5 random permutations. Figure 20 shows the great accuracy of the Monte Carlo estimates for $n \geq n_{MC}$. Second, we calculated $\mathbb{V}[C]$ using the simple formulae in Table 1 producing $\mathbb{V}[C]'''$. We checked that $\mathbb{V}[C]' = \mathbb{V}[C]'' = \mathbb{V}[C]'''$ for $n_{min} \leq n \leq n_{MC}$. For simplicity, the Monte Carlo estimation was not applied to bipartite graphs, that were tested for $1 \leq n_1, n_2 \leq 32$.

In all the tests described above, some elementary properties where checked:

- Any time C is calculated, the condition $C \leq |Q|$ is checked.
- Any time that $\mathbb{V}[C]$ is calculated theoretically, namely via the amount of occurrences of the 9 different types, two conditions are tested. First,
$$\sum_i f_i = |Q|^2$$
Second, f_i is even for any i excluding $i = 24$.

The manual mathematical tests consisted on checking that $\mathbb{V}[C] = 0$ in complete graphs (section 5.3.1) and star trees as expected. The case of star tree is trivial because $|Q| = 0$. The values of $\mathbb{V}[C]$ and the $f_i's$ for the toy graphs in Appendix B were calculated by hand. These independent results provide test cases for all the computer algorithms on small graphs.

Appendix D. Alternative proof for $f_{021} = 2n_G(\mathcal{L}_4 \oplus \mathcal{L}_2)$.

We can see that type 021 counts over the $\mathcal{L}_4 \oplus \mathcal{L}_2$. This can be readily seen by noting that for a fixed $\{st, uv\} \in Q$ the graphs counted by φ_{st}, φ_{uv} and the $\varepsilon_{..}$ are of the form

$$\begin{aligned} \varphi_{st} &: \{(z_s, s, t, z_t), (u, v)\}, & \varphi_{uv} &: \{(s, t), (z_u, u, v, z_v)\}, \\ \varepsilon^*_{su} &: \{(t, s, u, v), (z_1, z_2)\}, & \varepsilon_{sv} &: \{(t, s, v, u), (z_1, z_2)\}, \\ \varepsilon_{tu} &: \{(s, t, u, v), (z_1, z_2)\}, & \varepsilon_{tv} &: \{(s, t, v, u), (z_1, z_2)\}. \end{aligned}$$

We marked ε_{su} with * for later reference. We now show that each $\mathcal{L}_4 \oplus \mathcal{L}_2$ is counted twice in equation 77. Consider a fixed $\mathcal{L}_4 \oplus \mathcal{L}_2$ of the form $H = \{(s, t, u, v), (w, x)\}$. We can make four elements of Q using its edges, i.e.

$$q_1 = \{st, uv\},\ q_2 = \{st, wx\},\ q_3 = \{tu, wx\},\ q_4 = \{uv, wx\},$$

only two of which make equation 77 count over H, once for each of the two. This can be seen by noting that the graphs counted by the equation have to preserve the order of the vertices of the $\mathcal{L}_4$, achieved only by ε_{su} when the equation is evaluated with q_1 (see the description above), and by φ_{tu} when evaluated with q_3 (both marked with * above and below). For the sake of brevity we only show the form of the graphs counted by equation 77 for q_2 and q_3. The rest (for q_4) are derived similarly. Notice that for q_2 none of the graphs have the form of H and that the same happens with q_4.

$$\begin{aligned}
q_2 = \{st, wx\} \qquad & \varphi_{st} : \{(z_s, s, t, z_t), (w, x)\}, & \varphi_{wx} : \{(s, t), (z_w, w, x, z_x)\}, \\
& \varepsilon_{sw} : \{(t, s, w, x), (z_1, z_2)\}, & \varepsilon_{sx} : \{(t, s, x, w), (z_1, z_2)\}, \\
& \varepsilon_{tw} : \{(s, t, w, x), (z_1, z_2)\}, & \varepsilon_{tx} : \{(s, t, x, w), (z_1, z_2)\}, \\
q_3 = \{tu, wx\} \qquad & \varphi^*_{tu} : \{(z_t, t, u, z_u), (w, x)\}, & \varphi_{wx} : \{(t, u), (z_w, w, x, z_x)\}, \\
& \varepsilon_{tw} : \{(u, t, w, x), (z_1, z_2)\}, & \varepsilon_{tx} : \{(u, t, w, x), (z_1, z_2)\}, \\
& \varepsilon_{uw} : \{(t, u, w, x), (z_1, z_2)\}, & \varepsilon_{ux} : \{(t, u, x, w), (z_1, z_2)\}.
\end{aligned}$$

Therefore, each $\mathcal{L}_4 \oplus \mathcal{L}_2$ is only counted once for only two of the q_i's that can be made with its edges, namely

$$f_{021} = 2n_G(\mathcal{L}_4 \oplus \mathcal{L}_2).$$

Interestingly, the analysis above shows that a $\mathcal{L}_4 \oplus \mathcal{L}_2$ is only counted in exactly one of the two $\varphi_{..}$ and in exactly one of the four $\varepsilon_{..}$, hence equations 85 and 86.

References

[1] Marc Barthélemy. Spatial networks. *Physics Reports*, 499(1):1 – 101, 2011.

[2] David Eppstein and Siddharth Gupta. Crossing patterns in nonplanar road networks. In *Proceedings of the 25th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, SIGSPATIAL'17, pages 40:1–40:9, New York, NY, USA, 2017. ACM.

[3] 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.

[4] R. Ferrer-i-Cancho and C. Gómez-Rodríguez. Liberating language research from dogmas of the 20th century. *Glottometrics*, 33:33–34, 2016.

[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] R. Ferrer-i-Cancho. Hubiness, length, crossings and their relationships in dependency trees. *Glottometrics*, 25:1–21, 2013.

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

[8] R. Ferrer-i-Cancho. The sum of edge lengths in random linear arrangements. *Journal of Statistical Mechanics*, page 053401, 2019.

[9] R. Ferrer-i-Cancho. Euclidean distance between syntactically linked words. *Physical Review E*, 70:056135, 2004.

[10] J. L. Esteban, R. Ferrer-i-Cancho, and C. Gómez-Rodríguez. The scaling of the minimum sum of edge lengths in uniformly random trees. *Journal of Statistical Mechanics*, page 063401, 2016.

[11] Ramon Ferrer-i-Cancho. Why do syntactic links not cross? *Europhysics Letters*, 76(6):1228–1235, 2006.

[12] Ramon Ferrer-i-Cancho. A stronger null hypothesis for crossing dependencies. *Europhysics Letters*, 108:58003, 2014.

[13] Ramon Ferrer-i-Cancho and Carlos Gómez-Rodríguez. Crossings as a side effect of dependency lengths. *Complexity*, 21:320–328, 2016.

[14] C. Gómez-Rodríguez and R. Ferrer-i-Cancho. Scarcity of crossing dependencies: a direct outcome of a specific constraint? *Physical Review E*, 96:062304, 2017.

[15] R. Ferrer-i-Cancho. Non-crossing dependencies: least effort, not grammar. In A. Mehler, A. Lücking, S. Banisch, P. Blanchard, and B. Job, editors, *Towards a theoretical framework for analyzing complex linguistic networks*, pages 203–234. Springer, Berlin, 2016.

[16] M. Chimani, S. Felsner, S. Kobourov, T. Ueckerdt, P. Valtr, and A. Wolff. On the maximum crossing number. *https://arxiv.org/abs/1705.05176*, 2017.

[17] 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 Com- bin. Appl.*, volume 2, pages 975–989, New York, 1991. Wiley.

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

[19] R. J. Wilson. *Introduction to graph theory*. Longman, Harlow, England, 4th edition, 1996.

[20] J. W. Moon. On the distribution of crossings in random complete graphs. *Journal of the Society for Industrial and Applied Mathematics*, 13, 1965.

[21] K.H. Rosen, D. R. Shier, and W. Goddard. *Hanbook of discrete and combinatorial mathematics*. CRC Press, Boca Raton, FL, 2017.

[22] F. Bernhart and P. C. Kainen. The book thickness of a graph. *Journal of Combinatorial Theory, Series B*, 27(3):320 – 331, 1979.

[23] F. Harary. *Graph Theory*. Addison-Wesley, Reading, MA, 1969.

[24] H. N. de Ridder et al. Information System on Graph Classes and their Inclusions (ISGCI). http://www.graphclasses.org.

[25] D. Berend, S. Dolev, and A. Hanemann. Graph degree sequence solely determines the expected Hopfield network pattern stability. *Neural Computation*, 27(1):202–210, 2015.

[26] C. Castellano, S. Fortunato, and V. Loreto. Statistical physics of social dynamics. *Rev. Mod. Phys.*, 81:591–646, 2009.

[27] J. Smit. *Introduction to Quantum Fields on a Lattice*. Cambridge University Press, Cambridge, 2002.

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

[29] 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.

[30] N. Movarraei and M. Shikare. On the number of paths of lengths 3 and 4 in a graph. *International Journal of Applied Mathematical Research*, 3(2), 2014.

[31] N. Alon, R. Yuster, and U. Zwick. Finding and counting given length cycles. *Algorithmica*, 17(3):209–223, Mar 1997.

[32] F. Harary and B. Manvel. On the number of cycles in a graph. *Matematický časopis*, 21(1):55–63, 1971.

[33] M. A. Serrano, M. Boguñá, R. Pastor-Satorras, and A. Vespignani. Correlations in complex networks. In G. Caldarelli and A. Vespignani, editors, *Structure and Dynamics of Complex Networks, From Information Technology to Finance and Natural Science*, chapter 1. World Scientific, 2006.

[34] N. Movarraei and S. A. Boxwala. On the number of paths of length 5 in a graph. *International Journal of Applied Mathematical Research*, 4(1):30–51, 2015.

[35] E. Deutsch and M. Noy. Statistics on non-crossing trees. *Discrete Mathematics*, 254(1):75 – 87, 2002.

[36] B. Bollobás and O. Riordan. Mathematical results on scale-free random graphs. In S. Bornholdt and H. Schuster, editors, *Handbook of graphs and networks: from the genome to the Internet*, pages 1–34. Wiley-VCH, Berlin, 2003.

[37] D. Aldous. The random walk construction of uniform spanning trees and uniform labelled trees. *SIAM J. Disc. Math.*, 3:450–465, 1990.

[38] A. Broder. Generating random spanning trees. In *Symp. Foundations of Computer Sci., IEEE,*

pages 442–447, New York, 1989.

[39] M. E. J. Newman. *Networks. An introduction.* Oxford University Press, Oxford, 2010.

[40] L. Alemany-Puig and R. Ferrer i Cancho. Reappraising the distribution of the number of edge crossings of graphs on a sphere. *in prep*, 2018.

[41] A. Cayley. A theorem on trees. *Quart. J. Math*, 23:376–378, 1889.

[42] H. Prüfer. Neuer Beweis eines Satzes über Permutationen. *Arch. Math. Phys*, 27:742–744, 1918.

[43] L. Alonso and R. Schott. *Random generation of trees. Random generators in computer science.* Springer, Dordrecht, 1995.

[44] B. D. McKay and A. Piperno. Practical graph isomorphism, ii. *Journal of Symbolic Computation*, 60:94–112, 2013.