Spatial Graph Embeddings REU 2020

Yasmin Aguillon   Xingyu "Tiger" Cheng   Spencer Eddins
Pedro Morales

June 2020

Contents

1 Introduction

The motivation behind the study of knot theory stems from the discipline's many applications to molecular biology, especially in the study of DNA. DNA lives within the nucleus of a cell. One way to visualize the DNA inside the nucleus is to imagine that if the nucleus of a cell is the size of a basketball the DNA within it is equivalent to 3 km of fishing line [8]. One can easily imagine how these large DNA molecules might inevitably become entangled when they are compressed within the relatively small confined space of the nucleus. An interesting theorem by Diao, Pippengers and Sumners states that as the length of a chain approaches infinity, the knotting probability goes to 1 [13][8]. This theorem supports the idea that the likelihood of DNA becoming entangled within the nucleus is very high.

However DNA entanglement can be a problem, as it makes essential process like DNA replication and transcription more difficult for the cell to perform. Nature solves this dilemma by using enzymes, call topoisomerases to cut through knots and reconnect stands in a more ordely fashion. This process topologically alters the knot and DNA [4] [1].

An important question many biochemist and molecular biologist have sought to answer is: how do enzymes act upon DNA? To study this biochemist and molecular biologist have been studying circular DNA. There are two reasons for this. First, knots on circular DNA will not slip off the ends as they could within a regular DNA strand. Thus, circular DNA helps researchers more easily and accurately quantify and analyze knots in DNA. Secondly, with circular DNA, researchers are now working with closed curves. Thus, this allows researchers to apply tools from knot theory and topology to analyze DNA.

So, knot theory and topology is invaluable to scientists and researchers, as it provides scientists with a quantitative and invariant way to measure the properties of DNA. A good example of such a tool is the linking number, which is a link invariant that describes how 2 curves are "linked" or "tangled" in 3-space. A link is defined as a set of closed curves or hoops. Biochemist have used this invariant to quantitatively study and analyze the entanglement of DNA.

The majority of the results presented in this document stem from two theorems from the papers by Conway-Gordon[6] and Arsuaga-Blackstone[3], respectively and the classification presented in the paper by Rowland[9]. The Conway-Gordon paper, titled "Knots and links in spatial graphs" deals with the properties of spatial embeddings of the complete graphs K_6 and K_7. We examined the first of the two main theorems presented in the paper which dealt with the case of K_6. The theorem concludes that every embedding of K_6 contains a nontrivial link.

The Arsuaga-Blackstone paper we examined was titled "Linking of uniform random polygons in confined spaces." This paper examined the entanglement properties of polygons generated by randomly generated vertices in a confined symmetric convex space. Our group in particular studied the results "Lemma 1" and "Theorem 1." These two results establish some technical bounds on the expected value of the linking number between two random polygons in a confined space (which is a symmetric convex volume).

The Rowland paper, titled "Classifications of book representations of K_6", classifies all possible book representations of K_6. The paper identifies the number and type of knots and linked cycles in each representation or embedding. We build on these results in section 5 to investigate how to characterize "random" book embeddings of K_6.

Furthermore, in section 4 of this document, we present results where we attempt to analytically calculate the values of 4 variables presented in lemma 1 and theorem 1 of the Arsuaga-Blackstone paper: p, u, v and q. We use the theorem presented in the Conway-Gordon paper and theorem 1 in the Arsuaga-Blackstone paper to present results on random variables and the linking number of linearly embedded K_6 graphs and randomly generated polygons using the URP model from the Arsuaga-Blackstone paper.

2 Background results

2.1 Links of complete graphs

The paper by Conway and Gordon [6], dealt with the properties of embedded graphs K_6 and K_7. By K_n we denote the complete graph of n-vertices. The definition for a complete graph is given below:

Definition 1. *We say that a graph of n vertices is a **complete graph** if each vertex of the graph is connected to every other vertex of the same graph by exactly one edge. We denote such graph as K_n.*

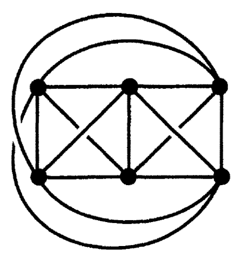


Figure 1: Complete graph of K_6

From this paper, we were mainly interested in the case of K_6, due to its accessibility, as the proof of the case K_7 is significantly more involved than the case of K_6. The result that comes from this investigation is given below:

Theorem 1. *Every spatial embedding of K_6 has a nontrivial link.*

We recall now the definition of a spatial graph embedding:

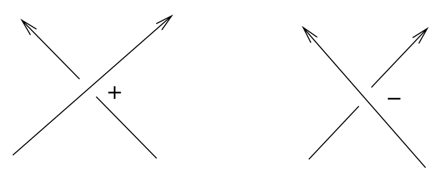
Definition 2. *A **spatial embedding** of a graph G is an embedding of the vertices of G in $\mathbb{R}^3$ along with the edges which connect these vertices.*

We then take an arbitrary projection of this spatial embedding of K_6 into $\mathbb{R}^2$, which would be our 2-dimensional picture. An important fact about this embedding is that if we choose any three vertices from this embedding then the edges connecting these three points would form what we call a cycle.

Since we have 6 nodes or vertices in K_6, the other 3 vertices also form a cycle. This means that there are always pairs of disjoint three-cycles in our embedding, so it is a natural question to ask if there are always cycles which admit nontrivial links. I.e. they are not (ambient) isotopic to the unlink.

Next, we define the linking number of two cycles C_1 and C_2:

Definition 3. *Given a set of orientations on C_1 and C_2, look at the crossing points between these two cycles, and for each of these crossings there will be an over strand and an under strand. We look at these crossings from the perspective of the over strand, and defined the sign of the crossing as such:*



That is, from the perspective of the overstrand, if the understrand is moving left, then we call that a positive crossing, while if the understand is oriented rightward, then we call that crossing negative. The ***linking number*** *is a number that describes how 2 components are linked in 3-space. It is simply the sum of all the positive and negative crossings divided by 2; we denote it as $lk(C_1, C_2)$, and it is given in equation form by*

$$lk(C_1, C_2) := \frac{(\#\text{ of positive crossings}) - (\#\text{ of negative crossings})}{2}.$$

Now that we have this setup, we are able to give a proof of Theorem 1.

Proof. Let $\lambda \in \mathbb{Z}_2$ be given as follows

$$\lambda = \sum \operatorname{lk}(C_1, C_2) \mod 2$$

where we are summing over all $10 = \frac{1}{2}\binom{6}{2}$ pairs of disjoint cycles (C_1, C_2).

We consider next what happens to our embedding under change of crossing. If the crossing we change is of an edge with itself, or crossings of adjacent edges, then nothing about λ would change because these crossings cannot happen between two triangles since two disjoint triangles cannot share the same edge.

The only crossing of significance is of non-adjacent edges. Notice that any crossing between non-adjacent edges E_1 and E_2 involves four points out of the six points in K_6. This means that there are two more points V_1 and V_2 that these two non-adjacent edges do not intersect.

If we connect V_1 to E_1 and V_2 to E_2 we get a pair of disjoint cycles. Similarly, we can connect V_1 to E_2 and V_2 to E_1 we get another pair of disjoint cycles. Figure 2

Then changing this crossing between E_1 and E_2 will change the linking number on two pairs of cycles by either $+1$ or -1 altering λ by either -2, 0 or $+2$

So in all cases, the linking numbers change by 0 mod 2, and hence λ remains unchanged. Finally we assert that the linking number is invariant under choice of projection and ambient isotopy. Thus λ will be as well. Moreover, it is well known that when allowing strands to pass

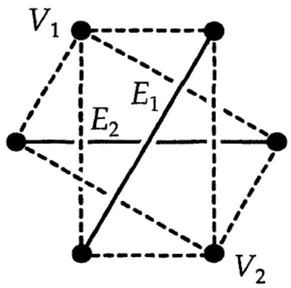


Figure 2: Non-adjacent edges of K_6

through each other, as in changing crossings, any embedding can be morphed into any other embedding. What we have shown above is that λ is thus invariant under such operations, so λ is the same for any two embeddings of K_6.

It now suffices to examine a particular embedding and check that indeed for this specific embedding, $\lambda = 1$. We can check that for the embedding in Figure 1, each pair of disjoint cycles forms a trivial link except one. That is, the sum of the linking numbers is not even, so it can't be 0, the unlink. Therefore for any embedding of K_6 there exists at least one nontrivial link. $\square$

2.2 Uniform Random Polygon

Let's consider a uniform random polygon, R_n, in a confined space, the unit cube. R_n is composed of n three-dimensional, random, independent, uniformly distributed points in the unit cube, which are denoted as U_i where $i = 1, 2, ...n$, and n segments connect these points such that e_i is the segment that joins the points U_i to U_{i+1}.

We will consider the projection diagram of R_n, where the projection is a regular projection. Consider the case where there are 2 independent edges, ℓ_1 and ℓ_2. Since the end points of these edges are independent and uniformly distributed, the probability that they intersect within the projection diagram is a positive value, which we will define as $2p$.

Let the orientation of an edge be determined by the order in which its vertices are chosen. That is, the edge containing v_i and v_{i+1} is directed from v_i to v_{i+1}. Let us define a random variable ε in the following ways: $\varepsilon = 0$ if the projection of ℓ_1 and ℓ_2 do not cross, $\varepsilon = 1$ if the projection of the crossing is a positive crossing and $\varepsilon = -1$ if the projection has a negative crossing. Figure 3.

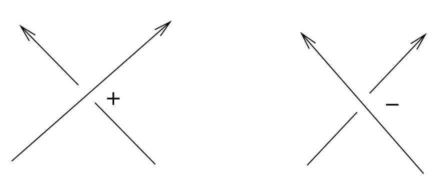


Figure 3: How we define positive and negative crossings

For simplicity, in this paper, we use a uniform distribution for the independent random points that form the uniform random polygons. It is worth saying that even if the distribution is changed, every argument in this section still holds. However, we cannot change the condition of independence.

Lemma 1. *Let K be a set of independently chosen random points. If V and V' are permutations of the same set of points, then $P(K = V) = P(K = V')$.*

Proof. Assume V and V' contain the same set of points as K, then we have that $K = V$ if and only if $v_i = k_i$ for all $i \in \{1, 2, \cdots, n\}$. Similarly for V'. The vertices of K were independently chosen, so the selection of v_i is independent of v_j for all $j < i$. That is, the event that $v_i = k_i$ is independent of the event that $v_j = k_j$. $\square$

Assuming that two edges cross, by symmetry, we see $P(\varepsilon = 1) = P(\varepsilon = -1) = p$ i.e. any crossing between ℓ_1 and ℓ_2 is no more likely to equal 1 as it is to equal -1. From this it follows that $E(\varepsilon) = 0$ because $E(\varepsilon) = 1 \cdot p - 1 \cdot p + 0 = 0$. We know the general formula for variance is $\mathrm{Var}(x) = E(x - E(x))^2$. Since $E(\varepsilon) = 0$, we have that $\mathrm{Var}(\varepsilon) = E(\varepsilon - E(\varepsilon))^2 = E(\varepsilon - 0)^2 = 2p$. Thus, $\mathrm{Var}(\varepsilon) = E(\varepsilon^2) = 2p$. (This observation will become useful in Lemma 1 part 3.)

Lemma 2. *(Modified version of Lemma 1 in [3]) Consider 4 edges: ℓ_1, ℓ_2, ℓ_1', and ℓ_2' with randomly and independently chosen endpoints. Some of these edges may be identical or share a common end point. Let's define ε_1 as the number ε between ℓ_1 and ℓ_1', and let ε_2 be the number ε between edges ℓ_2 and ℓ_2'.*

1. *If the end points of ℓ_1, ℓ_2, ℓ_1' and ℓ_2' are distinct, then $E(\varepsilon_1\varepsilon_2) = 0$ (This is the case when there are eight random points involved);*
2. *If $\ell_1 = \ell_2$, and the end points of ℓ_1' and ℓ_2' are distinct (this is the case when there are 6 independent points involved with 3 distinct edges), then $E(\varepsilon_1\varepsilon_2) = 0$.*
3. *Extra case: If l_1 and l_2 are adjacent and l_1' and l_2' are distinct (7 independent random points) then $E(\varepsilon_1\varepsilon_2) = 0$.*
4. *In the case that $\ell_1 = \ell_2$ and ℓ_1' and ℓ_2 are distinct (so there are only five independent random points involved), let $u = E(\varepsilon_1\varepsilon_2)$. In the case where ℓ_1 and ℓ_2 share a common point, ℓ_1' and ℓ_2' also share a common point (so there are four edges with six independent points involved in this case), let $E(\varepsilon_1\varepsilon_2) = v$. Lastly, let $p = P(\varepsilon = 1)$ for ε being the crossing sign of any two edges without restriction. Then we have that $q = p + 2(u+v) > 0$.*

Proof. 1. Since the vertices in ε_1 and ε_2 are chosen without dependence on each other, we know ε_1 and ε_2 must be independent variables. Since ε_1 and ε_2 are independent random variables, we know $E(\varepsilon_1\varepsilon_2) = E(\varepsilon_1)E(\varepsilon_2)$. Assuming, ℓ_1 and ℓ_1' cross and ℓ_2 and ℓ_2' cross, we know by symmetry that $P(\varepsilon_1 = 1) = P(\varepsilon_1 = -1)$ and $P(\varepsilon_2 = 1) = P(\varepsilon_2 = -1)$i.e. for any crossing between ℓ_1 and ℓ_1' or ℓ_2 and ℓ_2', ε_1 or ε_2, depending on the crossing, is no more likely to equal 1 as it is to equal -1. So, it follows that $E(\varepsilon) = 0$. So, $E(\varepsilon_1) = 0$ and $E(\varepsilon_2) = 0$. Since $E(\varepsilon_1\varepsilon_2) = E(\varepsilon_1)E(\varepsilon_2)$. We know $E(\varepsilon_1\varepsilon_2) = 0$. Thus, the claim holds.

2. For each configuration where both ℓ'_1 and ℓ'_2 cross and $\ell_1 = \ell_2$, i.e. the configurations where $\varepsilon_1\varepsilon_2 \neq 0$, there are a total of 8 different ways of assigning orientations. (See Figure 2 below from Arsuaga's paper.) For 4 of these 8 orientations $\varepsilon_1\varepsilon_2 = 1$, and for the remaining 4 of the 8, $\varepsilon_1\varepsilon_2 = -1$. So, $E(\varepsilon_1\varepsilon_2) = 1(\frac{4}{8}) - 1(\frac{4}{8}) = 0$. Thus, by symmetry $E(\varepsilon_1\varepsilon_2) = 0$, and Arsuaga's claim holds.

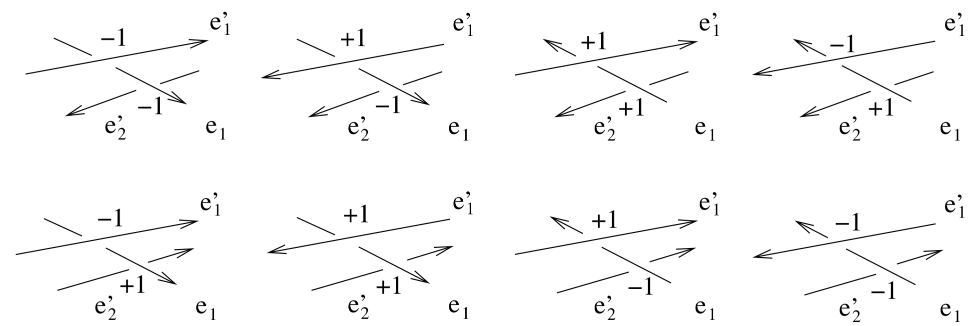


Figure 4: For each configuration described in lemma 1 part 2 where $\varepsilon_1\varepsilon_2 \neq 0$, there are 8 symmetric ways of assigning the orientations.

3. Extra case: For the scenario when l_1, l_2 are adjacent and l'_1, l'_2 are distinct, we can use the symmetry argument by noticing that it is equally probable, by independence of vertices, for l'_1 to have either orientation. Then for any product $\varepsilon_1\varepsilon_2$, there is an equally probable crossing where ε_1 is negative, and thus the product is negative. That is, the expected value for a nonzero product of crossings is also 0.

4. Let's consider two random triangles. The first triangle will consist of sides ℓ_1, ℓ_2, ℓ_3 while the second triangle will consist of sides $\ell'_1, \ell'_2, \ell'_3$. Let ε_{ij} be the number ε between edges ℓ_i and ℓ'_j. Let us consider the variance of the summation

$$\sum_{i,j=1}^{3} \varepsilon_{ij} = (\varepsilon_{1,1} + \varepsilon_{1,2} + \varepsilon_{1,3} + \varepsilon_{2,1} + \varepsilon_{2,2} + \varepsilon_{2,3} + \varepsilon_{3,1} + \varepsilon_{3,2} + \varepsilon_{3,3}).$$

Notice that $\mathrm{Var}(\varepsilon) = E(\varepsilon^2) - [E(\varepsilon)]^2$. However, given that $P(\varepsilon = 1) = P(\varepsilon = -1)$, it follows that $E(\varepsilon) = 0$. Therefore, we get that $\mathrm{Var}(\varepsilon) = E(\varepsilon^2)$.
Now, consider the variance of the summation $\sum_{i,j=1}^{3} \varepsilon_{ij}$, where indices are taken using *mod* 3, then we have

$$\mathrm{Var}\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right) = E\left(\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right)^2\right).$$

This equation can be simplified into the 3 following sums:

$$\mathrm{Var}\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right) = E\left(\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right)^2\right)$$

$$= \sum_{i,j=1}^{3} E(\varepsilon_{ij}^2) + 2 \sum_{i,j=1}^{3} [E(\varepsilon_{ij}\varepsilon_{i(j-1)}) + E(\varepsilon_{ij}\varepsilon_{i(j+1)})]$$
$$+ 2 \sum_{i,j=1}^{3} [E(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + E(\varepsilon_{ij}\varepsilon_{i-1,j+1})].$$

These 3 sums can be simplified even further. Each term in the first summation, $\sum_{i,j=1}^{3} E(\varepsilon_{ij})^2$, yields $2p$. When we distribute the terms in the sum we know that each term will be multiplied by itself once resulting in the terms of the first sum (so, this sum will have 9 terms). Since ε in this scenario can only equal 1 or -1, we know $\varepsilon_{i,j}^2$ will always equal 1. Since we previously defined the probability of two edges intersecting within the projection diagram as $2p$, for each term in the sum $E(\varepsilon_{ij})^2 = 1 \cdot 2p = 2p$. So, the first sum

$$\sum_{i,j=1}^{3} E(\varepsilon_{ij})^2 = 9 \cdot 2p.$$

Each term of the second sum, $2\sum_{i,j=1}^{3} E(\varepsilon_{ij}\varepsilon_{i(j-1)}) + E(\varepsilon_{ij}\varepsilon_{i(j+1)})$, yields u, where $u = E(\varepsilon_1\varepsilon_2)$ in the case that the edges $\ell_1 = \ell_2$ and the edges ℓ_1' and ℓ_2' and share a common point (so, there are five independent random points and 3 distinct edges). Visually, each term of the sum can be seen in terms of two triangles as shown in Figure 5.

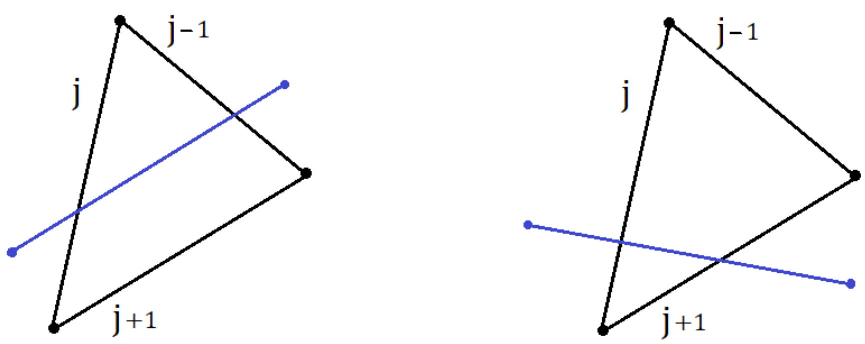


Figure 5: Let the blue edge be $\ell_1 = \ell_2$. In the left triangle, edge j and $j-1$ are the adjacent edges that share a point i.e. ℓ_1' and ℓ_2'. For the right triangle edge j and $j+1$ are the adjacent edges that share a point i.e. ℓ_1' and ℓ_2'.

Thus, each term in the sum will yield u. The sum is multiplied by 2 to account for the other set of terms that results when the order of ε's is switched. This is analogous to when we simplify the expression $(a+b)^2$. So, $(a+b)^2 = a^2 + 2ab + b^2$. Our middle term has a factor of 2 to account for both the ab and ba term. We multiply the sum by 2 to account for both of these terms. The sum will yield a total of 18 terms. Thus,

$$2\sum_{i,j=1}^{3} E(\varepsilon_{ij}\varepsilon_{i(j-1)}) + E(\varepsilon_{ij}\varepsilon_{i(j+1)}) = 2 \cdot 18u.$$

Each term of the third sum, $2\sum_{i,j=1}^{3}(E(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + E(\varepsilon_{ij}\varepsilon_{i-1,j+1})$, yields v, where $v = E(\varepsilon_1\varepsilon_2)$ in the case where ℓ_1 and ℓ_2 share a common point and ℓ_1' and ℓ_2' also share a common point (thus, there are 4 distinct edges and six independent points in this case). Visually, each term of the sum can be seen as parts of two triangles as shown in Figure 6.

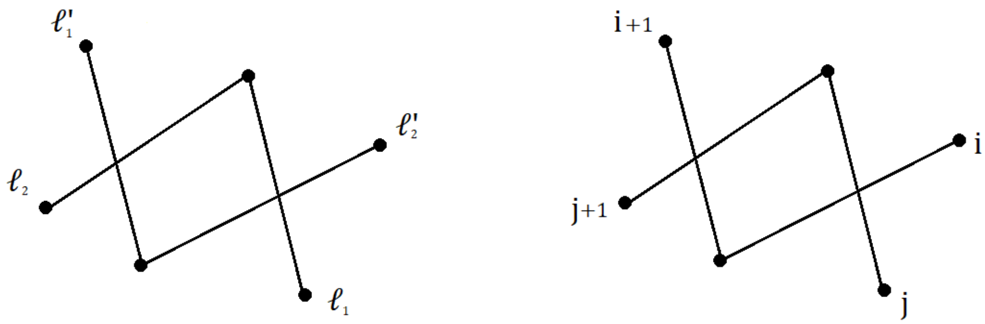


Figure 6: Illustration of the v case.

Thus, each term in the sum will yield v. Like the second sum, this sum is also multiplied by 2 to account for the other set of terms that results when the order of each ε is switched. This is analogous to when we simplify the expression $(a+b)^2$. So, $(a+b)^2 = (a+b)(a+b) = a^2 + ab + ba + b^2 = a^2 + 2ab + b^2$. Our middle term has a factor of 2 to account for both the ab and ba term. This sum also results in a total of 18 different terms. So,

$$2\sum_{i,j=1}^{3}(E(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + E(\varepsilon_{ij}\varepsilon_{i-1,j+1}) = 2 \cdot 18v.$$

Thus,

$$Var\left(\sum_{i,j=1}^{3}\varepsilon_{ij}\right) = E\left(\sum_{i,j=1}^{3}\varepsilon_{ij}\right)^2 = 18[p + 2(u+v)].$$

Since Var $\left(\sum_{i,j=1}^{3}\varepsilon_{ij}\right) > 0$, this implies that $q = 18[p+2(u+v)] > 0$. Thus, Arsuaga's claim holds.

□

Theorem 2. *(Theorem 1 of [3])*

The mean squared linking number between two uniform random polygons R_1 and R_2 of n edges each (in the confined space C^3) is $\frac{1}{2}n^2q$ where $q = p + 2(u+v)$ as defined in lemma 1.

Proof. Recall that the linking number can be written as the sum of each signed crossing. Then we have that

$$E\left(\mathrm{lk}(C_1, C_2)^2\right) = E\left(\left(\frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}\varepsilon_{i,j}\right)^2\right) = \frac{1}{4}\sum_{i,j,i',j'=1}^{n} E(\varepsilon_{i,j}\varepsilon_{i',j'}).$$

This gives

$$E\left(\mathrm{lk}(C_1,C_2)^2\right) = \frac{1}{4}\Bigg(\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}^2)\right) + 2\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}\varepsilon_{i(j-1)}) + E(\varepsilon_{ij}\varepsilon_{i(j+1)})\right)$$
$$+\, 2\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + E(\varepsilon_{ij}\varepsilon_{i-1,j+1})\right)\Bigg).$$

Note that these sums simplify via lemma 1. Firstly,

$$\begin{aligned}\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}^2)\right) &= \sum_{i,j=1}^{n}\left((\pm 1)^2 P(\varepsilon_{ij}=\pm 1) + 0\cdot P(\varepsilon_{ij}=0)\right)\\ &= \sum_{i,j=1}^{n}\left(P(\varepsilon_{ij}=1) + P(\varepsilon_{ij}=-1)\right)\\ &= \sum_{i,j=1}^{n} 2p = 2p\cdot\sum_{i,j=1}^{n} 1 = 2n^2p.\end{aligned}$$

Secondly,

$$2\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}\varepsilon_{i(j-1)}) + E(\varepsilon_{ij}\varepsilon_{i(j+1)})\right) = 2\sum_{i,j=1}^{n}\left(u+u\right) = 4u\sum_{i,j=1}^{n}1 = 4n^2u.$$

Similarly,

$$2\sum_{i,j=1}^{n}\left(E(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + E(\varepsilon_{ij}\varepsilon_{i-1,j+1})\right) = 2\sum_{i,j=1}^{n}\left(v+v\right) = 4n^2v.$$

Putting these sums together we get

$$E\left(\mathrm{lk}(C_1,C_2)^2\right) = \frac{1}{4}(2n^2p + 4n^2u + 2n^2v) = \frac{1}{2}n^2(p+2u+2v) = \frac{1}{2}n^2q.$$

□

3 Random variables and the linking number

3.1 Number of segments as a random variable

Before introducing the topics of this section, it is important to present some probability background.

Theorem 3 (Law of Total Expectation)**.** *Let X, Y be random variables. Then*

$$E(X) = E\left(E\left(X|Y\right)\right)$$

where $X|Y$ is the conditional probability distribution of X given Y.

If Y takes the outcomes $A_1, A_2, \ldots, A_n$, then the law can be rewritten as

$$E(X) = \sum_{i=1}^{n} E(X|A_i)\, P(A_i).$$

Below is a result about the linking number given the number of edges is now a random variable.

Lemma 3. *Let N be a random variable and let R_1 and R_2 be two uniform random polygons in a convex confined space C with N vertices each. Then we have that the mean squared linking number*

$$E\left(\left(\frac{1}{2}\sum_{i,j}^{N} \varepsilon_{ij}\right)^2\right) = \frac{1}{2} q E(N^2),$$

where $q = p + 2(u+v)$ as in Theorem 2.

Proof. This is a result of a straightforward calculation involving conditional expectation. Note that by Theorem 3 we have that

$$\begin{aligned} E\left(\left(\frac{1}{2}\sum_{i,j}^{N} \varepsilon_{ij}\right)^2\right) &= \sum_{k\geq 3} E\left(\left(\frac{1}{2}\sum_{i,j}^{N} \varepsilon_{ij}\right)^2 \middle| N = k\right) P(N=k) \\ &= \sum_{k\geq 3} \frac{1}{2} k^2 q P(N=k) \\ &= \frac{1}{2} q \sum_{k\geq 3} k^2 P(N=k) \\ &= \frac{1}{2} q E(N^2). \end{aligned}$$

The last equality comes from the fact that $E(g(X)) = \sum_x g(x) P(X=x)$. ∎

3.2 Mean squared linking number of linear embeddings of K_6

For a random embedding of K_6 whose projection is connected by straight edges the expected value of the sum of the linking number squared is given by the following theorem

Theorem 4. *The mean squared linking number of two cycles in a random linear embedding of K_6 is*

$$E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i,j}^{3} \varepsilon_{ij}\right)^2\right) \approx 1.521$$

where $\sum_{C_1,C_2}$ means that we are summing over all $10 = \frac{1}{2}\binom{6}{2}$ pairs of cycles $\{C_1, C_2\}$ in K_6.

Proof. Given that we are connecting vertices of K_6 with straight edges we get that each cycle in any embedding of K_6 will be a triangle, that is they have 3 edges. Now, using the expression previously stated we have the mean of the sum of the linking number squared for an embedding connected by straight edges is

$$E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i,j}^{3}\varepsilon_{ij}\right)^2\right) = 5q(3)^2 = 45q \tag{1}$$

Using the value of $q \approx 0.0338$ provided in [3] we can approximate (1) to 1.521, value that says it is more likely to get a link with squared linking number equal to 1 than squared linking number equal to 3,5,7, or 9 (since by [6], the sum of the linking number must be odd). □

Given the nature of a random embedding of K_6, a good question to ask is what happens when we increase the number of edges connecting the vertices of each cycle. If we do so we can rearrange (1) to express the expected value of the sum of the linking number squared in terms of the number of segments connecting each vertex in K_6.

Lemma 4. *Consider a random embedding of K_6 whose projection is connected by straight edges. Let m denote the number of segments connecting each vertex in a cycle of K_6, so each edge has the same number of segments m. Then we have that the average of the sum of the squared linking number over 10 different unordered pairs of cycles $\{C_1, C_2\}$ is given by*

$$E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i,j}^{3m}\varepsilon_{ij}\right)^2\right) = 45m^2q$$

Proof. If we increase the number of segments connecting two vertices of a cycle by the same number m then we end up with $3m$ edges in the cycle. Thus, the linking number is taken over $3m$. Now, the expected value of the sum of the linking number squared can we expressed as following

$$\begin{aligned}
E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i,j}^{3m}\varepsilon_{ij}\right)^2\right) &= \sum_{C_1,C_2}\left(E\left(\frac{1}{2}\sum_{i,j}^{3m}\varepsilon_{ij}\right)^2\right) \\
&= \sum_{C_1,C_2}\left(\frac{1}{2}(3m)^2q\right) \\
&= 10\left(\frac{9}{2}m^2q\right) \\
&= 45m^2q
\end{aligned}$$

The first equality comes from the fact that the average of a sum is equal to the sum of the averages. The second equality is given by Theorem 1 in [3] with $n = 3m$, and the third equality comes from the sum being over the 10 different cycles $\{C_1, C_2\}$ in the embedding. □

Another natural question to ask is what happens when we choose m to be a random variable. If we take the same approach used in lemma 2 together with our result in lemma 3 we get the following expression for the expected value of the sum of the linking number squared

Theorem 5. *Let M denote the random variable representing the number of segments connecting any two vertices in a random embedding of K_6. Then we have that*

$$E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i,j}^{3M}\varepsilon_{ij}\right)^2\right) = 45qE(M^2)$$

Proof. The proof for this Lemma is a direct application of Lemma 2 to the result obtained in Lemma 3. □

3.3 Uniform random polygons with different number of segments

The next step in the study of uniform random polygons and their linking number is a generalization of Theorem 1 in [3] to polygons with different length.

If we consider this new condition we get the following theorem

Theorem 6. *Let C_1 and C_2 be two randomly chosen polygons formed by uniformly and independently chosen vertices in a convex space, and connected by straight edges. If C_1 and C_2 have length m and n respectively , then the mean of the squared linking number between these two random polygons is given by*

$$E\left(\left(\frac{1}{2}\sum_{i}^{m}\sum_{j}^{n}\varepsilon_{i,j}\right)^2\right) = \frac{1}{2}mnq$$

where q is as defined in [3].

Proof. The proof of this result is exactly analogous to the $\frac{1}{2}qn^2$ result. Note first that

$$\begin{aligned} E\left(\left(\sum_{i}^{m}\sum_{j}^{n}\varepsilon_{ij}\right)^2\right) =& \sum_{i}^{m}\sum_{j}^{n}E(\varepsilon_{ij}^2) + 2\sum_{i}^{m}\sum_{j}^{n}E(\varepsilon_{ij}\varepsilon_{(i+1)j}) + E(\varepsilon_{ij}\varepsilon_{(i-1)j}) \\ &+ 2\sum_{i}^{m}\sum_{j}^{n}E(\varepsilon_{ij}\varepsilon_{(i+1)(j+1)}) + E(\varepsilon_{ij}\varepsilon_{(i-1)(j+1)}) + R \end{aligned}$$

Where R constitutes the sum of terms where the index of ε_{ij} differs by more than 1. Note now that R must be zero, since each of the expected values summed in R must be zero (cf. Lemma 2).

Now, similarly to Theorem 2, we have that each of the terms in the first first sum equals $2p$, each of the terms in the second sum equals u and each of the terms in the third sum equals v. This means then that

$$E\left(\left(\sum_{i}^{m}\sum_{j}^{n}\varepsilon_{ij}\right)^2\right) = \sum_{i}^{m}\sum_{j}^{n}2p + 2\sum_{i}^{m}\sum_{j}^{n}2u + 2\sum_{i}^{m}\sum_{j}^{n}2v.$$

Note now that there are mn terms in each double sum, so we get that

$$E\left(\left(\sum_i^m \sum_j^n \varepsilon_{ij}\right)^2\right) = mn(2p + 4(u+v)).$$

Multiplying back the 1/4 gives us that the expected value equals $\frac{1}{2}mnq$ as desired. □

If M and N are random variables, then the mean of the linking number squared is given by the following theorem.

Theorem 7. *Let C_1 and C_2 be two uniform random polygons (i.e. polygons defined by vertices chosen uniformly and independently) with lengths M and N respectively, with M and N independent random variables. Then, the mean of the squared linking number between these two polygons is given by*

$$E\left(\left(\frac{1}{2}\sum_i^M \sum_j^N \varepsilon_{ij}\right)^2\right) = \frac{1}{2}\, E(N)\, E(M)\, q,$$

where q is as in [3].

Proof. This result is achieved by a straight forward but tedious application of conditional expectation.

$$\begin{aligned}
E\left(\left(\frac{1}{2}\sum_i^N \sum_j^M \varepsilon_{ij}\right)\right) &= E\left(E\left(\left(\frac{1}{2}\sum_i^M \sum_j^N \varepsilon_{ij}\right) \Big| N\right)^2\right) \\
&= E\left(E\left(E\left(\left(\frac{1}{2}\sum_i^M \sum_j^N \varepsilon_{ij}\right) \Big| N\right)^2 \Big| M\right)\right) \\
&= \sum_{m\geq 3} E\left(E\left(\left(\frac{1}{2}\sum_i^M \sum_j^N \varepsilon_{ij}\right)^2 \Big| N\right) \Big| M = m\right) P(M=m) \\
&= \sum_{m\geq 3} E\left(\left(\frac{1}{2}\sum_i^m \sum_j^N \varepsilon_{ij}\right)^2 \Big| N\right) P(M=m) \\
&= \sum_{m\geq 3}\left[\sum_{n\geq 3} E\left(\left(\frac{1}{2}\sum_i^m \sum_j^N \varepsilon_{ij}\right)^2 \Big| N=n\right) P(N=n)\right] P(M=m) \\
&= \sum_{m\geq 3}\left[\sum_{n\geq 3} E\left(\left(\frac{1}{2}\sum_i^m \sum_j^n \varepsilon_{ij}\right)^2\right) P(N=n)\right] P(M=m) \\
&= \sum_n \sum_m \frac{1}{2} mnq P(N=n)P(M=m)
\end{aligned}$$

$$
\begin{aligned}
&= \frac{1}{2} q \sum_n \left[nP(N=n) \sum_m mP(M=m) \right] \\
&= \frac{1}{2} q \sum_n nP(N=n) E(M) \\
&= \frac{1}{2} E(M)\, E(N)\, q.
\end{aligned}
$$

Line 8 is achieved from Theorem 6, while line 9 is achieved by factoring out terms not dependent on m. □

3.4 Mean squared linking number in linear embeddings of K_6 with number of segments as a random variable

In this section we look at the mean squared linking number of linear embeddings of K_6 when take the number of segments, in cycles as random variables. The following results are applications of the previous sections.

Theorem 8. *Let C_1 and C_2 be two uniform random cycles in K_6. Let v_1, v_2 and v_3 be vertices of C_1 and v_4, v_5 and v_6 be vertices of C_2. Define X_{ij} as the random variable for the number of segments in the edge from v_i to v_j, and further suppose that each X_{ij} is iid. Then, the mean of the squared linking number is given by*

$$
E\left(\sum_{C_1, C_2} \left(\frac{1}{2} \sum_i^M \sum_j^N \varepsilon_{ij} \right)^2 \right) = 45\, q\, E(X_{ij}^2)
$$

where $M = X_{12} + X_{13} + X_{23}$ and $N = X_{45} + X_{46} + X_{56}$.

Proof. This proof is a direct application of Theorem 4.

$$
\begin{aligned}
E\left(\sum_{C_1, C_2} \left(\frac{1}{2} \sum_i^M \sum_j^N \varepsilon_{ij} \right)^2 \right) &= \sum_{C_1, C_2} \left(E\left(\frac{1}{2} \sum_i^M \sum_j^N \varepsilon_{ij} \right)^2 \right) && (2) \\
&= \sum_{C_1, C_2} \left(\frac{1}{2}\, q\, E(N)\, E(M) \right) && (3) \\
&= \sum_{C_1, C_2} \left(\frac{1}{2} q\, E(X_{12} + X_{13} + X_{23})\, E(X_{45} + X_{46} + X_{56}) \right) && (4)
\end{aligned}
$$

We know by properties of the expected value that the expected value of a sum of random variables is equal to the sum of the expected values of the same random variables, so we have that

$$
E(X_{12} + X_{13} + X_{23}) = E(X_{12}) + E(X_{13}) + E(X_{23}).
$$

A similar decomposition happens with $E(X_{45} + X_{46} + X_{56})$. Also, considering that every X_{ij} is chosen the same, this means that

$$
E(X_{12} + X_{13} + X_{23}) = 3E(X_{ij}),
$$

and similarly $E(X_{45} + X_{46} + X_{56}) = 3E(X_{ij})$. Hence, we can rewrite (4) as

$$
\begin{aligned}
E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{i}^{M}\sum_{j}^{N}\varepsilon_{ij}\right)^2\right) &= \sum_{C_1,C_2}\frac{1}{2}q\Big(3E(X_{ij})\Big)\Big(3E(X_{ij})\Big) \\
&= \sum_{C_1,C_2}\left(\frac{9}{2}\,q\,E(X_{ij}^2)\right) \\
&=10\left(\frac{9}{2}\,q\,E(X_{ij}^2)\right) \\
&=45\,q\,E(X_{ij}^2)
\end{aligned}
$$

where the last sum is over the 10 different pairs of cycles $\{C_1, C_2\}$ in a linear embedding of K_6. □

4 Attempts to analytically calculate q, p, u and v

Our investigation begins by thinking about the easiest cases of this problem. We narrowed down that good ways to analyze this problem begins with thinking about spaces such as $K \times I$, where $K \subset \mathbb{R}^2$ a convex, and I is the interval. The simplest of these cases for now is the unit cube. In this section, we will be using techniques and results from probabilistic geometry to attempt to calculate the values of p, u, v and q. Throughout our calculations, we find that the analytic calculation of p is relatively simple, while the analytical calculation of u and v present some difficulty.

Calculating p for the unit cube is equivalent to calculating the probability of two random lines intersecting in the unit square in $\mathbb{R}^2$. This is because we can simply take the projection to be $I^2 \times I \twoheadrightarrow I^2$, i.e. the map $(x, y, z) \mapsto (x, y)$. This projection will give us an uniform distribution on the square I^2, and we need not worry about the case when our projection is not regular because that happens with probability zero. This type of argument holds true for u and v as well.

4.1 Calculating p

Now if we view p as the probability of two random line segments intersecting in $R \in \mathbb{R}^2$, where R is our chosen region, then p has a known value. This is an application of Sylvester's four point problem. Note that the end points of two crossing line segments form a convex quadrilateral; these two conditions are equivalent, since any convex quadrilateral has intersecting diagonals. Note now that for any set of four points, there are $\frac{1}{2}\binom{4}{2} = 3$ ways to pick the two line segments given four points in the plane. By Lemma 1, these are all equally likely. Now out of all those configurations, only one is a pair of intersecting line segments. It is now prudent to state Sylvester's four point problem.

Theorem 9 (Sylvester, [10])**.** *Let $K \subset \mathbb{R}^2$ be an open convex set of finite area. Pick four points uniformly and independently in the region. Then the probability of those four points* *forming a convex quadrilateral is given by*

$$P(K) = 1 - 4\frac{A_K}{A(K)}$$

where A_K is the expected area of random triangles in K, while $A(K)$ is the area of the region K.

Now in the case of the square $K = [0,1]^2$, A_K is a known value. [11] gives the expected area as $\frac{11}{144}$, which is also consistent with other sources ([2], [10]). $P(K)$ then is a known value. Plugging into the formula we have that

$$\begin{aligned} P(K) &= 1 - 4\frac{A_K}{A(K)} \\ &= 1 - 4\left(\frac{11}{144}\right) \\ &= \frac{25}{36}. \end{aligned}$$

This means that the probability of two random line segments intersecting is then

$$\frac{1}{3} \cdot \frac{25}{36} = \frac{25}{108},$$

since we have $\frac{1}{3} = \frac{1}{2}\binom{4}{2}$ chance of picking the intersecting line segments from the convex polygon formed by our four points, and we have a $\frac{25}{36}$ chance of those four points making a convex quadrilateral. To conclude the calculation for p, we notice that

$$E(\varepsilon^2) = \frac{25}{108},$$

where ε is the value of the crossing between two randomly chosen line segments in the cube $[0,1]^3$. Now, if $\varepsilon = 1$ or $\varepsilon = -1$, then that means that in the projection, the two line segments intersect. Hence, $E(\varepsilon^2)$ is the probability that two random line segments intersect in the cube. Now, since we defined $P(\varepsilon = 1) = P(\varepsilon = -1) = p$, this means that $E(\varepsilon^2) = 2p$, so we have that $2p = \frac{25}{108}$, and hence

$$p = \frac{25}{216}.$$

4.2 Calculating u

Next, we want to consider how we can calculate u. This is the case when we have three edges define by five independently chosen points: one independent edge, and two adjacent edges.

Definition 4 (Convex Hull)**.** *Given a set of points S in the plane, we define as a convex hull the smallest convex set that contains S.*

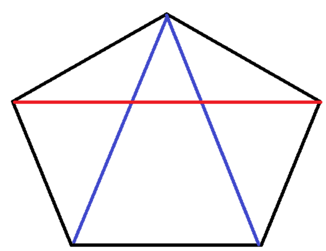
To think about this problem, we can break it down into cases where we can apply methods similar to the Sylvester problem. Because we are considering five points, the convex hull of these points can form either a triangle, quadrilateral, or pentagon. We'll abbreviate the event in which the hull is a convex pentagon, quadrilateral, and triangle as "5-gon," "4-gon," and "3-gon" respectively. Now, u is defined to be

$$u = E(\varepsilon_1 \varepsilon_2),$$

where ε_1 is defined to be crossing between ℓ_1 and ℓ_1' and ε_2 is defined to be ℓ_1 and ℓ_2', where we denote ℓ_1 as the independent edge, and ℓ_1' and ℓ_2' are the two adjacent edges. Now, breaking u up, we write

$$\begin{aligned} u &= P(\text{5-gon})E(\varepsilon_1\varepsilon_2|\text{5-gon}) + P(\text{4-gon})E(\varepsilon_1\varepsilon_2|\text{4-gon}) + P(\text{3-gon})E(\varepsilon_1\varepsilon_2|\text{3-gon}) \\ &= u_{3-gon} + u_{4-gon} + u_{5-gon}. \end{aligned}$$

Now, in a set of 5 points forming a convex pentagon, there are $3\binom{5}{3} = 30$ combinations of line segments, each equally likely by Lemma 1. Note that no edge that is part of the boundary of the convex hull can intersect any other edges, so the product of their signed crossings is 0. There are only $\frac{1}{2}\binom{5}{3} = 5$ combinations that only include interior edges, and they are the 5 rotations of following image.



This gives a 1/6 chance of picking 3 line segments with 2 crossings, given these points form a convex pentagon. Next, we need the probability of five uniformly chosen points forming a convex pentagon. This is a known result by Valtr [11], and we state it as the theorem below.

Theorem 10 (Valtr)**.** *The set A of n random points chosen independently and uniformly from a parallelogram S is convex with probability*

$$P_n = \left(\frac{1}{n!}\binom{2n-2}{n-1}\right)^2.$$

Plugging in for $n = 5$ then we get that

$$\begin{aligned} P_5 &= \left(\frac{1}{5!}\binom{2(5)-2}{5-1}\right)^2 \\ &= \left(\frac{1}{5!}\binom{8}{4}\right)^2 \\ &= \left(\frac{7}{12}\right)^2 \end{aligned}$$

$$= \frac{49}{144}.$$

Now, multiply with the $\frac{1}{6}$ chance of picking intersecting line-segments, and we get that there is a

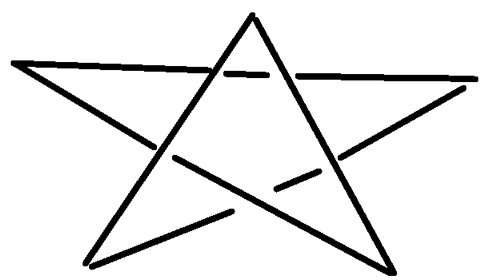
$$\frac{1}{6} \cdot \frac{49}{144} = \frac{49}{864}$$

chance for $\varepsilon_1\varepsilon_2 \neq 0$. So in the convex pentagon case, we have

$$u_{5-gon} = \big(P(\varepsilon_1\varepsilon_2 = +1) - P(\varepsilon_1\varepsilon_2 = -1)\big)\frac{49}{864},$$

where we are taking the probability to be implicitly taken given that we have non-zero crossings, and that the points form in the projection a convex pentagon. Let c be this difference of probabilities $P(\varepsilon_1\varepsilon_2 = +1) - P(\varepsilon_1\varepsilon_2 = -1)$. Then c can be calculated using combinatorics, and we will state the result as a lemma below:

Lemma 5. *Pick five points in 3-space. If those five points form a convex pentagon in the 2-dimensional projection, then there is a unique crossing diagram up to planar isotopy of a linear embedding of the pentagram formed by those five points, pictured below. Furthermore, from all the crossings in the pentagram, there is exactly one crossing configuration such that $\ell_1 = \ell_2$ goes in between the adjacent edges ℓ_1' and ℓ_2'.*



Proof. Call the configuration where we have two adjacent edges and an independent (third) edge passing between the two adjacent edges $\cancel{V}$ crossing. We first show that every pentagram knot diagram must contain at most one $\cancel{V}$ crossing configuration.

First consider a $\cancel{V}$ crossing, and label the points of this crossing v_1, v_2, v_3, v_4, v_5 as in Figure 7, and likewise label ℓ_1, ℓ_1' and ℓ_2' as in the figure.

Our goal is to show that such a configuration generates a unique pentagram. To see that, we attempt to connect vertices v_4 and v_3 and v_1 and v_5. Call the edge connecting v_3 and v_4, l_{34}, and the edge connecting v_1 and v_5, l_{15}. Now, consider the affine plane formed by v_4, v_2, and v_5, call it P. Note now that v_1 and v_3 must be on opposite sides of the plane P (in Figure 8, v_1 is above, and v_3 is below). Without loss of generality, suppose v_1 is above and v_3 is below, then this means that l_{15} must be above the plane P and l_{34} must be below the plane, and further ℓ_1' is also above P, while ℓ_2' is below P. This implies then that ℓ_1' is the overstrand with respect to l_{34}, and l_{15} is the overstrand with respect to ℓ_2' and l_{34}, since lines above the plane must be the overstrand with respect to lines below.

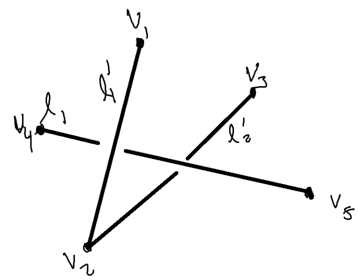


Figure 7: $\not V$

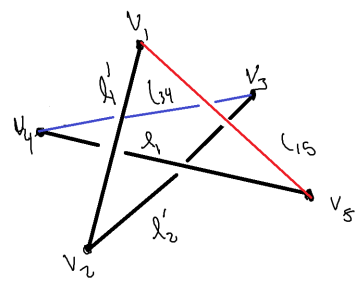


Figure 8: Completing the pentagram from the $\not V$ configuration

Look at all possible combination of edges from the pentagram where two are adjacent, and one is the unique disjoint edge. We find that there is only one combination where the third disjoint edge passes through the two adjacent edges. This is our original $\not V$, which means that this configuration generates the pentagram, and hence there are at most one combination in any pentagram where there is a disjoint edge passing through the two adjacent edges.

Next, we show that for all pentagrams, there must be at least one combination of edges where two are adjacent, and the third disjoint edge passes through the two adjacent edges, that is, for all pentagrams we can get from five (convex) points, there is at least one $\not V$ configuration. To show this, we first consider five points whose projection forms a convex pentagon, and call those points v_1, v_2, v_3, v_4 and v_5 like in the picture below, that is, we label the points in counter clockwise order by adding two to the next index until we reach 4; furthermore, the addition is done modulo 5, but we label $\hat{0} = 5$ (cf. Figure 9).

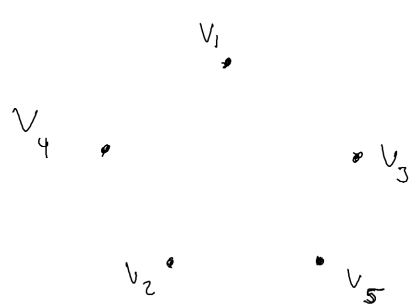


Figure 9: Convex configuration of five points

Assume further that for any pair of adjacent edges the disjoint edge always passes over both adjacent edges or always passes under both adjacent edges. That is, in the pentagram diagram, there are no $\not{V}$ configuration. We then connect the edges v_1 and v_2 and v_2 and v_3, and call those lines l_{12} and l_{23}. This is illustrated in Figure 10.

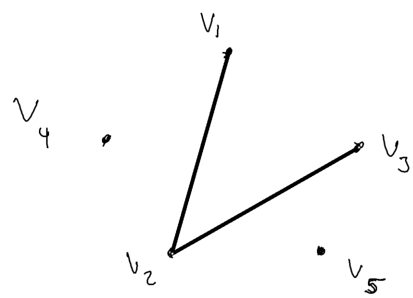


Figure 10: Constructing lines l_{12} and l_{23}

Next, we connect v_4 and v_5, and again, call that line l_{45}. Without loss of generality, assume that this edge is underneath l_{12} in the knot diagram. Consequently, by our assumption, l_{45} must also pass underneath l_{23} in the knot diagram as well. This step is illustrated in Figure 11.

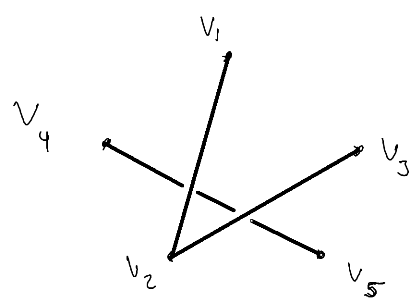


Figure 11: Constructing l_{45}

We then connect the points v_4 and v_3 and call that edge l_{43}. Since l_{45} is passing beneath l_{12} this means that l_{43} must also pass underneath l_{12}. This is as in Figure 12.

Finally, we connect vertices v_1 and v_5 with edge l_{15}. Since l_{45} passes underneath l_{23}, this means that l_{15} must also pass underneath edge l_{23}, and consequently, it must also pass underneath edge l_{43}. However, since edge l_{12} passes over l_{43} this means that l_{15} must also pass over l_{43}, since we assumed that the adjacent edges must both either pass over or pass under the disjoint edge, but this is a contradiction since then l_{15} must cross both over and underneath l_{43}, impossible!

We must conclude then that the only such pentagram that we can draw is the one which contains exactly one $\not{V}$ configuration. $\square$

We note next that in a fixed pentagram the $\not{V}$ configuration of adjacent edges in the pentagram with the disjoint edge always has positive product of crossings while the other

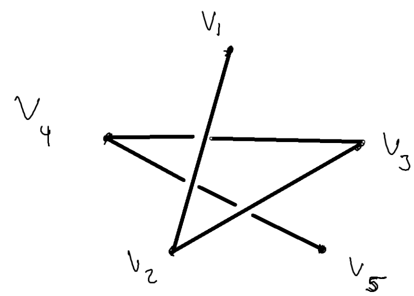


Figure 12: Constructing l_{43}

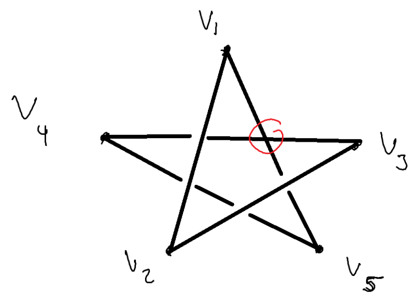


Figure 13: Contradictory crossing between l_{43} and l_{15}

configurations always have negative product of crossings. Since given a fixed set of five points, whose projection forms a pentagram, each configuration of two adjacent edges and a disjoint edge whose end points coincide with five points forming the pentagram are all equally likely to occur (they are simply permutations of the order in which we pick the points), this means that given the five points form in the project a convex pentagon, positive product crossings happen with probability 1/5, while negative product of crossings happen with probability 4/5. This means then that our difference of the probability of positive crossings to the probability of negative crossings, c, is

$$\frac{1}{5} - \frac{4}{5} = -\frac{3}{5}.$$

Taking all the products together then, we have that

$$u_{5-gon} = \left(-\frac{3}{5}\right) \cdot \frac{49}{864}.$$

Simplifying, we get the following partial result for u:

Theorem 11 (u_{5-gon})**.** *In the case that the five independent points in the u case forms (in the projection) a convex pentagon, then the expected value of the product of the two linking numbers is given by*

$$u_{5-gon} = -\frac{49}{1440}.$$

$u_{5-gon} \approx -0.034027$, which is consistent with the average product of the crossing numbers when the projection forms a convex 5-gon from numerical simulations.

Next, before we move on to the 4-gon case, we wish to consider the case where the convex hull is a triangle. We, in fact, do not have to worry about this case, since the product of crossings will be identically 0. We present the result as a lemma below.

Lemma 6 (Triangle Lemma)**.** *Whenever there is a triangle convex hull, we have two cases:*

1. *If the convex hull of 4 points is a triangle, then no two linear edges cross.*

2. *If the convex hull of 5 points is a triangle, then exactly two linear edges cross.*

Proof. 1. If the convex hull of 4 points is a triangle, then 3 of the 4 points will form the vertices of the triangle while the remaining vertex will lie inside the triangle. Using this configuration construct K_4, the complete graph on 4 vertices. This creates a total of 6 edges, 3 exterior edges plus 3 interior edges, and will result in an image like the one below in Figure 14.

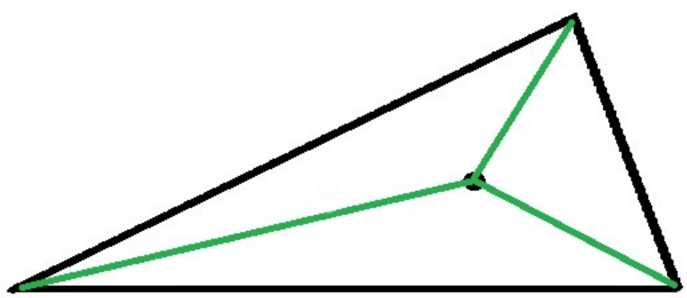


Figure 14

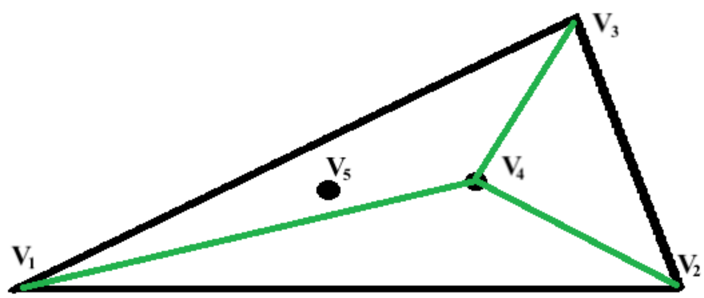
Exterior and interior edges in a convex space will never intersect. So, our 3 interior and 3 exterior edges will never cross. Furthermore, this claim can be proven by examining 2 cases. In the first case, consider 1 interior and 1 exterior edge such that the 2 edges are adjacent and there are 3 distinct points. For the purposes of this argument, a shared vertex between two or more adjacent edges will not be considered an intersection between those edges. Since infinite lines either cross only once or are parallel to each other, it follows that any two linear and adjacent edges will never cross. So, our edges in case 1 will never intersect.

In the second case, consider 1 interior and 1 exterior edge such that the 2 edges are not adjacent and there are 4 distinct endpoints. Two non-adjacent edges will only cross if the convex hull of their points forms a quadrilateral. Since we assumed the convex hull of our 4 points was a triangle, our convex hull is definitely not a quadrilateral. So, our edges in case 2 will never intersect. Thus, we have shown that the exterior and interior edges of a convex space never intersect.

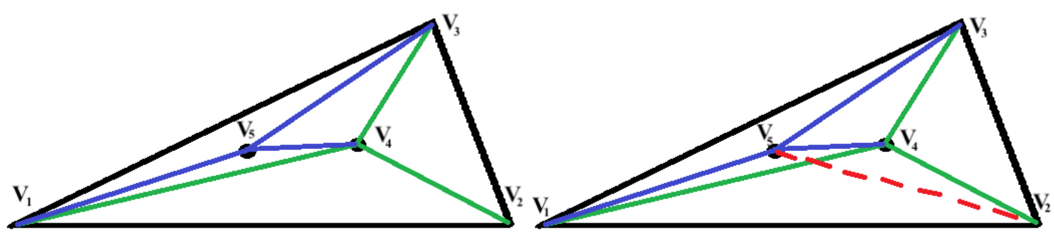
Now, let's consider our three exterior edges. Since we have already asserted that any 2 linear and adjacent edges will never cross. We know all 3 of our exterior edges will never intersect amongst themselves since they are all linear and adjacent. The same is true for our 3 interior edges. Since all 3 are linear and adjacent these 3 edges will never intersect themselves.

Since we have shown that none of the edges of the K_4 graph intersect when the graph is constructed such that its 4 vertices have a convex hull of a triangle, our claim holds.

2. If the convex hull of 5 points is a triangle, then there are two interior points. Call the exterior vertices v_1, v_2, v_3 and the interior vertices v_4 and v_5. By the first part of the lemma, v_4 in the triangle formed by v_1, v_2, v_3 has no edge crossings. Note that the edges $\{v_1, v_4\}, \{v_2, v_4\}, \{v_3, v_4\}$ partition the triangle into three smaller triangles, so v_5 must lie in one such region, as seen in the figure.

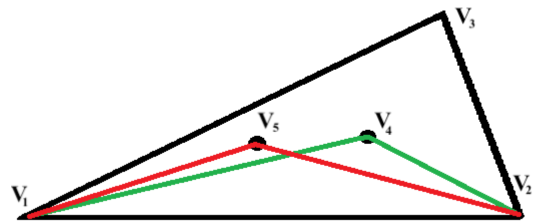


Without loss of generality, let v_5 lie in the triangle formed by v_1, v_2, v_4. Then again by the first part of the lemma, no edge from these four points have a crossing. Also, the edges formed by v_1, v_2, v_3 are the boundary of the convex hull, so they cannot be crossed. Notice then that we have described a set of edges that have no crossings with each other and that this set contains every edge except $\{v_3, v_5\}$. That is, if there are any crossings, then $\{v_3, v_5\}$ must be one of the edges involved.



Alternatively, we could have chosen to partition the outermost triangle using v_5, found which region v_4 lies in, and have come to a similar conclusion. Recall that we said that v_5 was in the triangle formed by v_1, v_2, v_4. Then it certainly isn't the case that v_4 is in the triangle formed by v_1, v_2, v_5. Therefore v_4 lies in either v_1, v_3, v_5 or v_2, v_3, v_5. Without loss of generality, say that v_4 lies in the triangle v_1, v_3, v_5. If we follow through like before, we'll notice that the set of all edges excluding $\{v_2, v_4\}$ has no crossings, so if the complete set of edges has a crossing, $\{v_2, v_4\}$ must be one of the edges in the crossing.

Since the set of all edges excluding $\{v_3, v_5\}$ has no crossings and similarly for the set excluding $\{v_2, v_4\}$, then if there is a crossing it must be between those two edges. In fact, these two edges will always cross.



Consider the convex hull of these four vertices. We will show that none of these four vertices lie in the triangle formed by the other three. Recall that v_4 and v_5 partition the outer triangle such that v_4 is in v_1, v_3, v_5, which implies that it's not in v_2, v_3, v_5, and v_5 is in v_1, v_2, v_4, which implies that v_5 isn't in v_2, v_3, v_4. Lastly, the convex hull of these four points is a subset of the convex hull of all five points. Since v_2 and v_3 are exterior vertices, they are not in the open set of the convex hull of all five points which implies they are not in any subsets of that set. That is, v_2 is not in the triangle formed by v_3, v_4, v_5 and v_3 is not in the triangle formed by v_2, v_4, v_5. Therefore none of the four vertices are in the convex hull of the other three, so the convex hull of all four is a quadrilateral.

Recall in our discussion for the value of p that four points form a convex quadrilateral if and only if the "diagonals" intersect. Then because these four points form a convex quadrilateral, its diagonals $\{v_2, v_4\}$ and $\{v_3, v_5\}$ intersect.

□

Now, the only other case we need consider is when the five points' convex hull is a quadrilateral. A natural first step then is to calculate the probability that five random chosen points in the square ($\subset \mathbb{R}^2$) has a convex hull of four points (quadrilateral case). We present the calculation below:

Lemma 7 ([7])**.** *The probability that five points picked uniformly from the unit square forms a convex hull of four points is equal to $x_4 = 5/9$.*

Proof. Denote the probability for the convex hull of five points to consist of k points by x_k, and denote the probability for the convex hull of k points to consist of k points by p_k. The convex hull has five points if and only if the five points form a convex pentagon, so we have that $x_5 = p_5$.

Now, there are $\binom{5}{4} = 5$ subsets of 4 points in our collection of five points. Each of these subsets have p_4 probability of forming a convex quadrilateral, so the expected number of subsets of four points forming a convex quadrilateral is given by $5p_4$ by linearity of expected value.

Note also, that if the convex hull of five points is five points, then all 5 subsets of four points form a convex quadrilateral. If the convex hull is four points, then the hull itself and two of the other four quadrilaterals are convex, for a total of three convex quadrilaterals.

If the convex hull is three points, then exactly one of the five quadrilaterals is convex (cf. Lemma 6).

Thus, we have that

$$5p_4 = 5x_5 + 3x_4 + x_3.$$

Note next that $x_5 = p_5$, and

$$x_3 + x_4 + x_5 = 1,$$

which gives us a system of three linear equations with three unknowns. Solving this system of equations gives us

$$\begin{aligned}
x_3 &= \frac{3}{2} - \frac{5}{2}p_4 + p_5, \\
x_4 &= -\frac{1}{2} + \frac{5}{2}p_4 - 2p_5, \\
x_5 &= p_5.
\end{aligned}$$

Now, $p_4 = 25/36$ and $p_5 = 49/144$ as previously calculated, which means that

$$\begin{aligned}
x_4 &= -\frac{1}{2} + \frac{5}{2}\left(\frac{25}{36}\right) - 2\left(\frac{49}{144}\right), \\
&= \frac{5}{9}.
\end{aligned}$$

□

Next, note that whenever we have a configuration of five points where the convex hull is a quadrilateral, we have that there are only two configuration of edges out of the $30 = 3\binom{5}{3}$ total combination of edges that have non-zero product of crossings. This gives us a chance of $2/30 = 1/15$ of forming a configuration where $\varepsilon_1\varepsilon_2 \neq 0$ whenever we have a configuration of points with their convex hull being a quadrilateral. Putting everything together, we get that

$$u_{4-gon} = \frac{1}{15} \cdot \frac{5}{9} \cdot d,$$

where d denotes difference between the probability of having $\varepsilon_1\varepsilon_2$ being positive versus $\varepsilon_1\varepsilon_2$ being negative, whenever we are in this configuration, where the convex hull in the projection is a quadrilateral. Although at first, it may seem that we can use the same type of combinatorics here as in the 5-gon case, however, we soon find that when we try to complete all the possible crossing configurations, we find two distinct pictures. In one, we have that we get a positive crossing and a negative crossing, and in the other, we find that we have two negative crossings. Numerical simulations suggest that $d \approx -0.506$, a value that is very close to 1/2, but differs by just that slight value. This gives us that

$$u_{4-gon} \approx \frac{1}{15} \cdot \frac{5}{9} \cdot (-0.506) \approx 0.0187$$

Putting everything together, we get that

$$u = u_{5-gon} + u_{4-gon}$$

$$= -\frac{49}{1440} + \frac{1}{15} \cdot \frac{5}{9} \cdot d$$
$$\approx -0.0527.$$

This u value is consistent with our numerical simulations, which also suggests a value of -0.0527, and it is also consistent with the value given by [3], within margin of error. The exact value of d, if it can be calculated, remains a mystery.

4.3 The values of q and v

The calculation for v might proceed in the same way as the calculation for u. Similar to the calculation for u, we will this time need to consider when the convex hull of 6 points is a 6-gon, 5-gon, 4-gon and 3-gon. In this decomposition, we have then that

$$v = v_6 + v_5 + v_4 + v_3,$$

where v_n is the value of v for the case where the convex hull is an n-gon. However, a full investigation of this value could be conducted in the future to make exact the nature of this decomposition.

The value of q may be calculated from the probability that two random triangles embedded in 3-space are linked. This discussion from [12] gives an explicit formula for finding the linkage of two random triangles. We will give a brief overview of the discussion leading to the formula for the linking number. Define triangles T_1 and T_2 by the points $\{a, b, c\}$ and $\{d, e, f\}$ respectively in an embedded simplex Δ' in general position (no four points coplanar). The two triangles are linked if and only if exactly one edge of T_2 pierces the triangle T_1. Let uv be an arbitrary edge of T and xy an arbitrary edge of T_2. Define

$$D(u, v, x, y) = \text{sgn}(\det(y - u, y - v, y - x)),$$

where $\text{sgn}(x) = \frac{x}{|x|}$. The function $D(u, v, x, y)$ equals 1 if the the crossing uv has positive sign and -1 if the crossing has negative sign. $D(u, v, x, y) \neq 0$ since Δ' is in general position.

The signed intersection number of T_1 and an edge xy of T_2 is the same regardless of the chosen projection. So we have two conditions that must be met in order to conclude that xy pierces T_1:

1. The sign of the crossing between uv and xy is the same for all edges in T_1 and T_2.

2. The edge xy intersects the plane containing T_1.

The first condition is satisfied if

$$D(a, b, x, y) = D(b, c, x, y) = D(c, a, x, y) \tag{5}$$

and the second condition is satisfied if

$$D(a, b, c, x) = -D(a, b, c, y). \tag{6}$$

We can combine those two conditions into a formula for the intersection number xy with T_1. Since $D(u,v,x,y)$ can only take on the values 1 and -1, the product

$$(D(a,b,x,y) - D(b,c,x,y))(D(a,b,x,y) - D(b,c,x,y))(D(b,c,x,y) - D(c,a,x,y)) \quad (7)$$

can only take on the values -8, 8 and 0, and furthermore, if the product is 0, then that means one of the differences in the product is 0, and hence equation 5 is violated. Additionally, the product

$$\frac{1}{4}(D(a,b,c,x) - D(a,b,c,y))^2 = \frac{1}{2}(1 - D(a,b,c,x)D(a,b,c,y)) \quad (8)$$

is non-zero if and only if equation 6 is satisfied. Now that we have these expressions, we can take the product of (7) and (8) to get

$$\begin{aligned} n_\cap(T_1,x,y) := \frac{1}{8}&(D(a,b,x,y) + D(b,c,x,y))(D(a,b,x,y) + D(c,a,x,y)) \\ &(D(b,c,x,y) + D(c,a,x,y)) \cdot \frac{1}{2}(1 - D(a,b,c,x)D(a,b,c,y)). \end{aligned}$$

The function $n_\cap(T_1,x,y)$ is equal to the signed intersection number of the triangle T_1 with the edge xy of T_2. Expanding the product and using the fact that $D(u,v,x,y)^2 = 1$, we can simplify the expression to be

$$\begin{aligned} n_\cap(T_1,x,y) = \frac{1}{8}&(D(a,b,x,y) + D(b,c,x,y) + D(c,a,x,y) \\ &+D(a,b,x,y)D(b,c,x,y)D(c,a,x,y)) \\ &(1 - D(a,b,c,x)D(a,b,c,y)). \end{aligned}$$

Now, for the linking number of T_1 and T_2, we observe that only one of the following may occur:

1. No edges of T_2 pierces T_1.
2. Exactly one edge of T_2 pierces T_1.
3. Exactly two edges of T_2 pierces T_1, and their intersection numbers have opposite sign.

From these conditions, we get a formula for the squared linking number of the triangles T_1 and T_2:

$$\mathrm{lk}(T_1,T_2)^2 = [n_\cap(T_1,d,e) + n_\cap(T_1,e,f) + n_\cap(T_1,f,d)]^2. \quad (9)$$

Now that we have an explicit formula for lk^2, we now recall Theorem 2 which states that the mean squared linking number between two uniform random n-gons is given by

$$E\left(\left(\frac{1}{2}\sum_{i,j}\varepsilon_i\varepsilon_j\right)^2\right) = \frac{1}{2}n^2 q.$$

Now since we are in the case of triangles, this means that we take $n = 3$ to get that

$$E(\mathrm{lk}(T_1,T_2)^2) = \frac{9}{2}q.$$

So we have that $q = \frac{2}{9}E(\mathrm{lk}^2)$. Using equation 9 we get that

$$E(\mathrm{lk}^2) = \int_{\Omega^3}\int_{\Omega^3} \mathrm{lk}(T_1, T_2)^2,$$

where Ω is the space where we pick our points which generate the triangles T_1 and T_2, and we integrate over the points $\{a, b, c\}$ and $\{d, e, f\}$ (hence the Ω^3). In the case of the unit cube, we have that $\Omega = [0, 1]^3$.

The challenge in this integral revolves around the division by absolute values required in the definition of $D(u, v, x, y)$; a closed form solution to this integral seems unlikely, and symbolic integration by Mathematica was not able to lead to any results. Mathematica was able to calculate $E(\mathrm{lk}^2)$ to around 0.152. However, numerical integration was not able to return a value with precision beyond three decimal places; numerical simulations of $E(\mathrm{lk}^2)$ directly from random crossings also give a value of approximately 0.1521. This seems to yield a value consistent with [3] (cf. Theorem 4).

4.4 p for other spaces

Now, our calculation for p is valid for a whole host of spaces. Note that the Sylvester four point problem is valid for all convex subsets of $\mathbb{R}^2$, and hence that means that if we take K to be a convex set, and pick points uniformly in $K \times I$, where I is the interval, then we have again that

$$2p = P(K) = 1 - 4\frac{A_K}{A(K)},$$

where again, A_K is the expected area of the with vertices picked uniformly and independently over K, while $A(K)$ is the area of the convex K.

For a cylinder, then the equivalent plane problem would be to find the expected area of a random triangle in a disk. The formula of Alikoski ([2], [10]) says that the expected area of a randomly picked triangle in an n-gon is given by

$$\frac{A_n}{A} = \frac{9\cos^2\omega_n + 52\cos\omega_n + 44}{36n^2\sin^2\omega_n},$$

where $\omega_n = \frac{2\pi}{n}$, and A is the area of the $n-$gon. Taking the limit as $n \to \infty$ gives us the expected area of a randomly picked triangle in a disk or ellipse. This limit goes to $\frac{35}{48\pi^2}$, so we have that for a disk

$$\begin{aligned} 2p_{disk} &= \frac{1}{3}\left(1 - 4 \cdot \frac{35}{48\pi^2}\right) \\ &= \frac{1}{3} - \frac{35}{36\pi^2}, \end{aligned}$$

hence

$$p_{disk} = \frac{1}{6} - \frac{35}{72\pi^2}.$$

For a triangular prism, i.e. the space defined by $K \times I$, where K is a triangle, the value of p can also be determined. Call this value $p_{triangle}$. From Alikoski's formula above, we find that in this case

$$\frac{A_n}{A} = \frac{1}{12},$$

so that we have

$$2p_{triangle} = \frac{1}{3}\left(1 - 4 \cdot \frac{1}{12}\right)$$
$$= \frac{2}{9}.$$

This means then that

$$p_{triangle} = \frac{1}{9}.$$

Now, Blaschke [5] proved that

$$\frac{2}{3} \leq P(K) \leq \frac{1}{6} - \frac{35}{12\pi^2},$$

where the lower bound is the value of $P(K)$ for a triangle, and the upper bound is the value of $P(K)$ for an ellipse. These bounds lead to the following theorem.

Theorem 12. *In the case where the confined space is of form $\Omega = K \times I$, where K is a convex region in $\mathbb{R}^2$, then the value of p is bounded by*

$$\frac{1}{9} \leq p_\Omega \leq \frac{1}{3} - \frac{35}{72\pi^2}.$$

4.5 Relationship between u, v and p

Numerical simulations suggests a linear relationship between u, v and p. By picking points from different spaces (even spaces not of form $K \times I$), and with different distributions, then the value of p would vary. Plotting the graphs of those values, we find that u and p and v and p follow an approximately linear relation, as can be see in Figure 15. Doing linear regression on this system we find that

$$u(p) \approx 0.0311 \cdot p - 0.723,$$
$$v(p) \approx -0.0258 \cdot p + 0.324.$$

Furthermore, we find that

$$\mathrm{Corr}(u, p) \approx -0.988$$

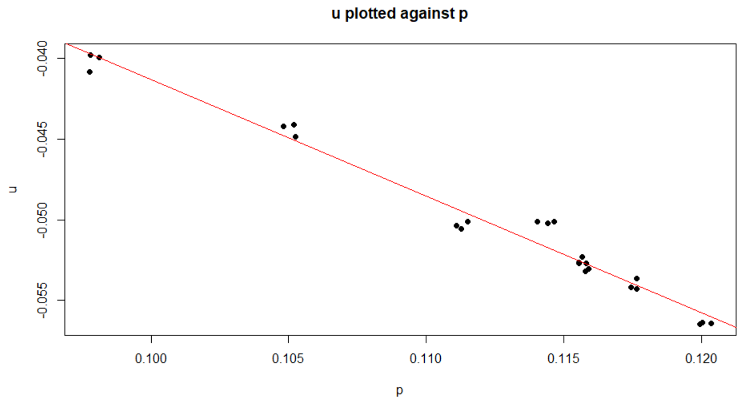
and

$$\mathrm{Corr}(v, p) \approx 0.974.$$

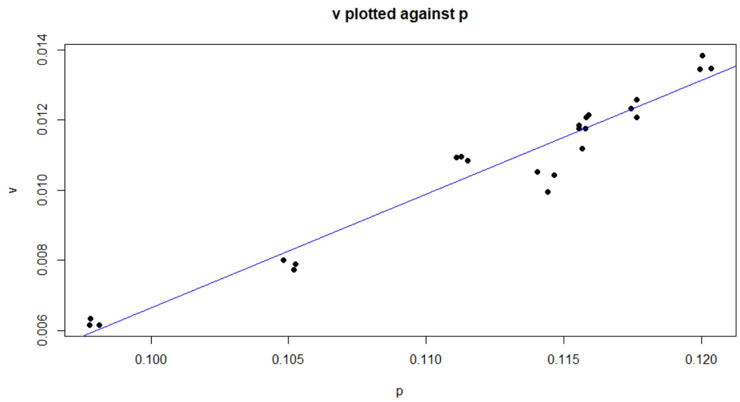
This suggests that the values of u and v are more dependent on two dimensional information that is encoded in p rather than the three dimensional information carried by the difference between the positive and negative crossings.

We can describe this relationship through the definitions of u and v. We show the case of u, since the case of v is completely analogous. Note first that

$$u = E(\varepsilon_1 \varepsilon_2).$$



(a) u plotted against p



(b) v plotted against p

Figure 15: Plots of different u and v for different values of p

Now, by definition, this means that

$$u = P(\varepsilon_1\varepsilon_2 = 1)(1) + P(\varepsilon_1\varepsilon_2 = -1)(-1).$$

Notice, then that

$$P(\varepsilon_1\varepsilon_2 = 1) = P(\varepsilon_1 = 1, \varepsilon_2 = 1) + P(\varepsilon_1 = -1, \varepsilon_2 = -1),$$

and

$$P(\varepsilon_1\varepsilon_2 = -1) = P(\varepsilon_1 = 1, \varepsilon_2 = -1) + P(\varepsilon_1 = -1, \varepsilon_2 = 1).$$

Note now that $P(\varepsilon_1 = 1, \varepsilon_2 = 1) = P(\varepsilon_1 = -1, \varepsilon_2 = -1)$ by symmetry, and similarly $P(\varepsilon_1 = 1, \varepsilon_2 = -1) = P(\varepsilon_1 = -1, \varepsilon_2 = 1)$. So we can simplify u to be

$$u = 2P(\varepsilon_1 = 1, \varepsilon_2 = 1) - 2P(\varepsilon_1 = 1, \varepsilon_2 = -1).$$

Note now, that $P(\varepsilon_1 = 1, \varepsilon_2 = 1) = P(\varepsilon_1 = 1|\varepsilon_2 = 1)P(\varepsilon_2 = 1)$, and similarly $P(\varepsilon_1 = 1, \varepsilon_2 = -1) = P(\varepsilon_1 = 1|\varepsilon_2 = -1)P(\varepsilon_2 = -1)$. By definition of p, we have that $P(\varepsilon_1 =$

$1) = P(\varepsilon_1 = -1) = p$, so we get that $P(\varepsilon_1 = 1, \varepsilon_2 = 1) = pP(\varepsilon_1 = 1|\varepsilon_2 = 1)$ and $P(\varepsilon_1 = 1, \varepsilon_2 = -1) = pP(\varepsilon_1 = 1|\varepsilon_2 = -1)$. Hence we get that

$$u = 2p[P(\varepsilon_1 = 1|\varepsilon_2 = 1) - P(\varepsilon_1 = 1|\varepsilon_2 = -1)].$$

Now, the difference of conditional probabilities is dependent on the space we choose. Call this difference of probabilities $c_u(\Omega)$, and similarly for v, call the difference of probabilities that we get in that case $c_v(\Omega)$, where Ω denotes the space where we pick our points. Hence, we can describe the relationship between u, v and p as

$$u = 2pc_u(\Omega)$$

and

$$v = 2pc_v(\Omega).$$

5 Random book embeddings

Definition 5. *According to Rowland [9], a book embedding of a graph G is an embedding satisfying*

1. *The vertices of G lie on a circle C in a plane.*
2. *The edges of G are distributed among disjoint "sheets," topological disks with boundary C.*
3. *The projection of edges of G to the plane containing C are chords of C.*

Furthermore, Rowland in [9] classifies all possible book representations of K_6, showing that the set of nontrivial links and knots are limited to the Hopf link, Solomon's knot, trefoil knots, and figure-eight knot. This opened up the question of how to characterize "random" book embeddings of K_6 and the distribution of these links and knots.

Any algorithm to generate random book embeddings is based on how the edges of K_6 are placed on disjoint sheets. Note that the exterior edges, as in the edges between two vertices neighbor on C's boundary, never intersect other edges. Therefore one needs only to consider the placement of the 9 interior edges. The classification from [9] establishes that there are Hopf links, Solomon links, and trivial links within book embeddings of K_6. Below are two ways to generate random embeddings.

5.1 Greedy Embedding Algorithm

We will label our first method of generating random embeddings as the "greedy" algorithm. This algorithm is called so because it chooses edges in a greedy manner. We will describe the method below:

Definition 6 (Greedy random embedding)**.** *The **greedy random embedding** (for K_6) algorithm is done in the following steps:*

1. *Pick a random edge of K_6.*

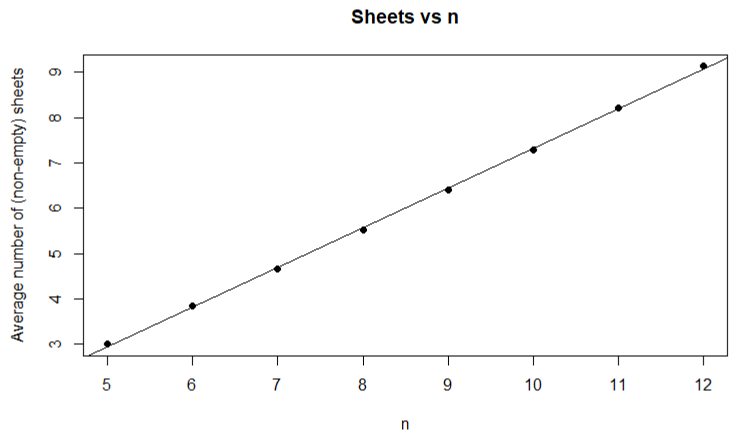


Figure 16: The average number of sheets in the greedy embedding appears to grow in $O(n)$ 'time.'

2. *Place the edge in the first sheet. If there is an intersection in the first sheet, then put the edge in the second sheet. If the edge still has an intersection in the second sheet, place on third...*

   Repeat until the edge is placed on the first sheet where there are no intersections with any other edges.

3. *Repeat steps 1. and 2. until no edges of K_6 remain.*

We can apply a similar algorithm as above to randomly embed edges of K_n, and when we numerically simulate the average number of sheets required to achieve random book embeddings of K_n, for $n \geq 3$, we find that the number sheets required appears to be $O(n)$ (cf. Figure 16). Let S represent the number of sheets a random greedy embedding takes and n be as in K_n. From our data, we have that

$$\text{Corr}(S, n) = 0.9997962.$$

This hints at a very strong, and perhaps interesting relationship between the two, however a rigorous proof eludes us for now. This increase in the number of sheets tells us that the complexity of embeddings of K_n increases as n increases, however the exact nature of this complexity is not yet known.

For the case of K_6, we were able to bound the number of sheets required for a random greedy embedding. We state the result below:

Proposition 1. *The number of sheets required for a greedy embedding of K_6 is between* 3 *and* 5.

Proof. We first label the points of K_6 in clockwise order and enumerate them from 0 to 5. We will also call the edges between points a and b as the unordered pair (a, b). Note now that all the exterior edges $(0, 1)$, $(1, 2)$, $(2, 3)$, $(3, 4)$, $(4, 5)$, and $(5, 0)$ will not intersect any other edge, so we will place them on the first sheet. This means that we need only worry about the 9 left over edges. For the rest of the proof, we will ignore the exterior edges and handle only the 9 other edges.

Note now that for any edge of K_6, there will always be 6 other edges which does not intersect it. Of those 6 edges, one can pick two edges with does not intersect each other; all other left over 4 edges will now have an intersection in the first sheet. This means that the first sheet must have three edges. By a similar argument, the second sheet must also have three edges. Then this means that the rest of the three edges must be collected in sheets 3, 4, and 5. Hence the minimum number of sheets required is 3, and the maximum number of sheets is 5. □

5.2 Non-greedy embeddings

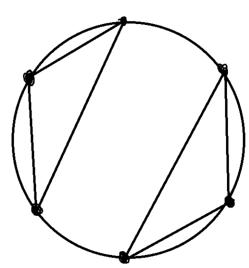
The second of our methods of randomly generating book embeddings of K_6, we will call the non-greedy method. This algorithm will randomly shuffle the order of the edges of K_6, and place them on discrete sheets. These discrete sheets will then generate a knot diagram. This algorithm is described step-by-step below:

Definition 7 (Non-greedy random book embeddings)**.** *The* ***Non-greedy random embedding*** *(for K_6) algorithm is done in the following steps:*

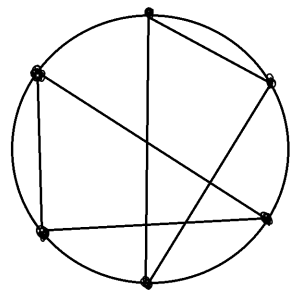
1. *We order the edges of K_6 and put them into a list.*
2. *Apply a random shuffle on that list to randomly order the edges.*
3. *The edges are then placed on discrete sheets to generate the book embedding.*

Using this method of generating embeddings, we can ask for the probability of getting Hopf links, Solomon links, and Trivial links. We analyze the types of links which appear in non-greedy embeddings of K_6. We first break down the possible triangles into three cases. We present those cases along with illustrations of the cases below.

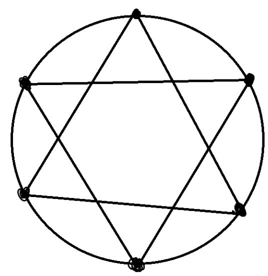
Case 1: This is the case when we have two non-intersecting interior edges along with four exterior edges which when connected with the two interior edges form triangles. There are three such cases.



Case 2: This is the case when we have two pairs of intersecting interior edges along with two exterior edges which completes the pair of triangles. In any book embedding of K_6, there are six of these cases.



Case 3: This is the case when we have all interior edges. There is only one of this case for any book embedding of K_6.



Now we go on to analyze the links which appear in each of the cases. For case 1, it is clear that no matter what order we pick the edges, there will never be any intersection, hence the linking number will always be zero, so the only possible link is the trivial link. For case 2, each interior edge is in two crossings, which means that the only possible link is the Hopf link. For case 3, we have six intersections, which means that the crossing number is either 0, 2, or 4, so the Solomon link is possible in this case (indeed, [9] details book embeddings with Solomon links).

Combinatorics is able to tell us about the probability of each link appearing for each case. For case 1, this is easy, since the trivial link is the only one which appear, this link happens with probability 1. For the cases 2 and 3, we need to consider the order in which the edges are picked from each triangle. Since shifting the sheets, that is changing the last sheet to be the first sheet, amounts to a Reidemeister move (cf. [9]), this means that, without loss of generality, we can pick a triangle to be the "A" triangle, and the other as the "B" triangle. Furthermore, we can also choose which edge we would like to be on top. We shall consider case 2 and case 3 separately.

In case 2, there are four edges to consider. The order we can pick our edges can be either AABB, ABAB, or ABBA. Without loss of generality, choose the triangle $(0, 1, 3)$ to be our "A" triangle, and choose our first edge to be $(0, 3)$. In this case, our "B" triangle will be $(2, 4, 5)$. Note next that by shifting, the order ABBA is the same as AABB. However we note that the only crossing diagram that AABB can generate is the one where the edge $(0, 3)$ is the overstrand to both $(2, 4)$ and $(2, 5)$, and $(1, 3)$ is the overstrand to be both $(2, 4)$ and $(2, 5)$. In this case, the crossings between $(0, 3)$ and $(2, 4)$ must be the negative of $(0, 3)$ and $(2, 5)$, and the crossing between $(1, 3)$ and $(2, 4)$ is the negative of the crossing between $(1, 3)$

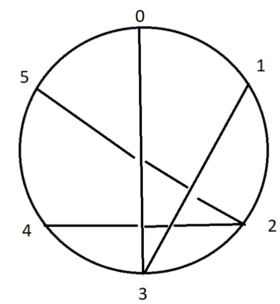


Figure 17: Case 2: AABB. The two triangles form an unlink.

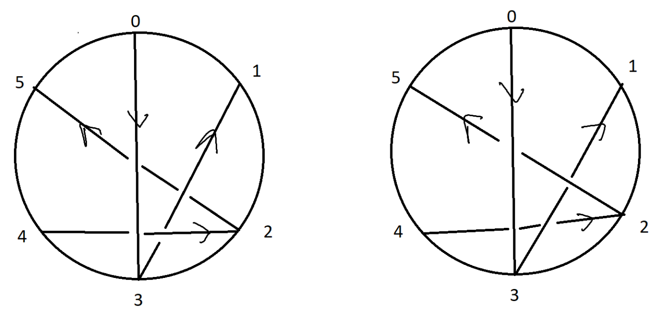


Figure 18: Case 2: triangles formed by ABAB (with orientation).

and $(2,5)$. This means that in this case, the only link possible is the trivial link (cf. Figure 17).

So any nontrivial links in the case 2, if they occur, must happen when we choose edges in the order ABAB. We will detail this case next.

1. Again, the first edge we choose is the edge $(0,3)$.
2. The next edge we choose can be either $(2,4)$ or $(2,5)$.
3. The third edge we choose must be $(1,3)$.
4. Depending on which edge we chose as the second edge, we must choose either $(2,5)$ or $(2,4)$, i.e. we choose the edge from triangle B that was not picked in step 2.

If we look at the orientation given in 18, we see that in the left picture we have that:

- The crossing between $(0,3)$ and $(2,4)$ is +1.
- The crossing between $(0,3)$ and $(2,5)$ is -1.
- The crossing between $(2,4)$ and $(1,3)$ is +1.
- The crossing between $(1,3)$ and $(2,5)$ is +1.

So in total we have for the left triangle in 18 that the total linking number is 1, hence we get a Hopf link. In the right triangle we have that

- The crossing between $(0, 3)$ and $(2, 4)$ is +1.
- The crossing between $(0, 3)$ and $(2, 5)$ is -1.
- The crossing between $(1, 3)$ and $(2, 4)$ is -1.
- The crossing between $(2, 5)$ and $(1, 3)$ is -1.

Hence the total linking number for the right triangle in 18 is -1, and hence we have a Hopf link in this case as well. So in either case of the ABAB order, we have that we achieve a Hopf link. Now, since each order is equally likely, this means that the total probability of achieving a Hopf link in the case 2 is 1/3.

We now move on the case 3. Again, we can fix which edge we choose first, and hence, without loss of generality, we can choose A to be the first in our ordering. The possible orders for our edges are now

1. AAABBB
2. AABABB
3. AABBAB
4. AABBBA
5. ABAABB
6. ABABAB
7. ABABBA
8. ABBAAB
9. ABBABA
10. ABBBAA.

Note that orders 1) and 4) are the same since we can shift 4) to become 1). Next, notice that orders 2), 3), 5), 7), 8), 9), and 10) are also all the same by shifting and renaming A to B. Order 6) is unique. This means that the only unique orders possible are AAABBB, AABBAB and ABABAB.

We look at the AAABBB case first. In this case we can always collapse the two triangles into two discrete sheets, and since we can do this, this means that the two triangles will not be linked (cf. Figure 19).

Next we move on the AABBAB case. In this case we can assume without loss of generality that $(0, 4)$ is the first edge that we pick. Now, since combinatorially, picking $(0, 2)$ and $(2, 4)$ will achieve the same result, we can assume that we picked $(0, 2)$ after $(0, 4)$ for the sake of simplicity. If we assume that, then the ways we can pick the next edges will split into three

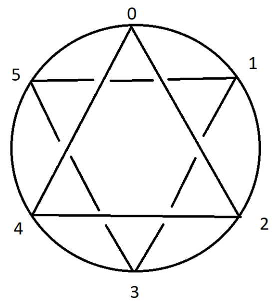


Figure 19: AAABBB picture. We see that we have two non-linked triangles.

paths as in Figure 20. Each of the three paths has a 1/3 probability of occurring, given that we pick $(0,4)$ and $(0,2)$ as our first two edges. We pick $(0,2)$ after $(0,4)$ with probability 1/2, hence each of the end result has a 1/6 chance of happening. On the left path in Figure 20 we have a Hopf link, in the middle path we have a trivial link while on the last path we have a Hopf link as well. Hence we get that in the AABBAB case, we achieve a Hopf link with 2/3 probability, while we achieve a trivial link in this case with a probability of 1/3 (we get 2/3 and 1/3 because we also have to add the cases when we pick $(2,4)$ as our second edge).

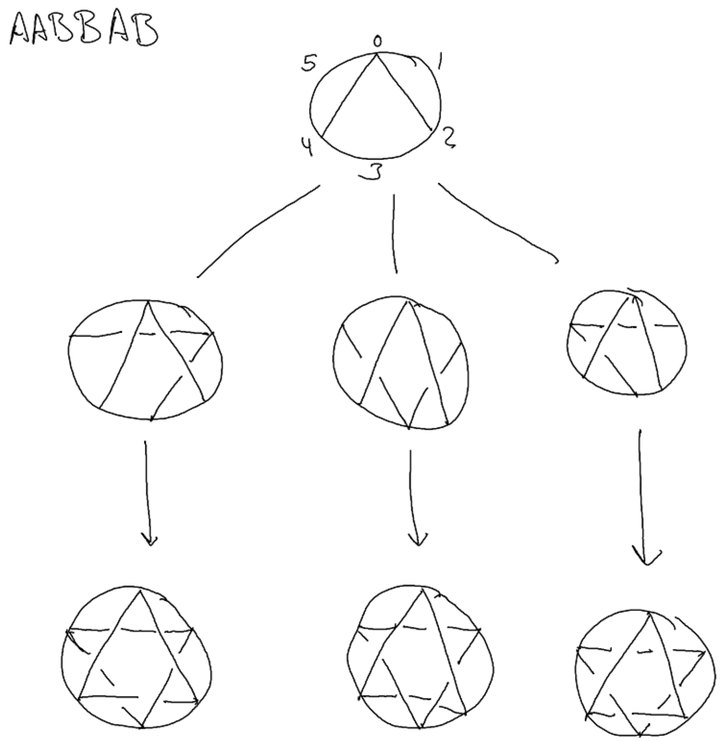


Figure 20: The AABBAB case. Each branch has a 1/3 probability of occurring.

Finally, we deal with the case of ABABAB. In this case, we can assume without loss of generality that $(5,1)$ is the first edge that we choose. Next, we choose our second edge from triangle B. This choice will create 2 paths. The case where the edge from triangle B intersects the first edge we chose creates the left path, and the case where the edge from triangle B does not intersect the first edge we chose creates the right path. See Figure 21 below.

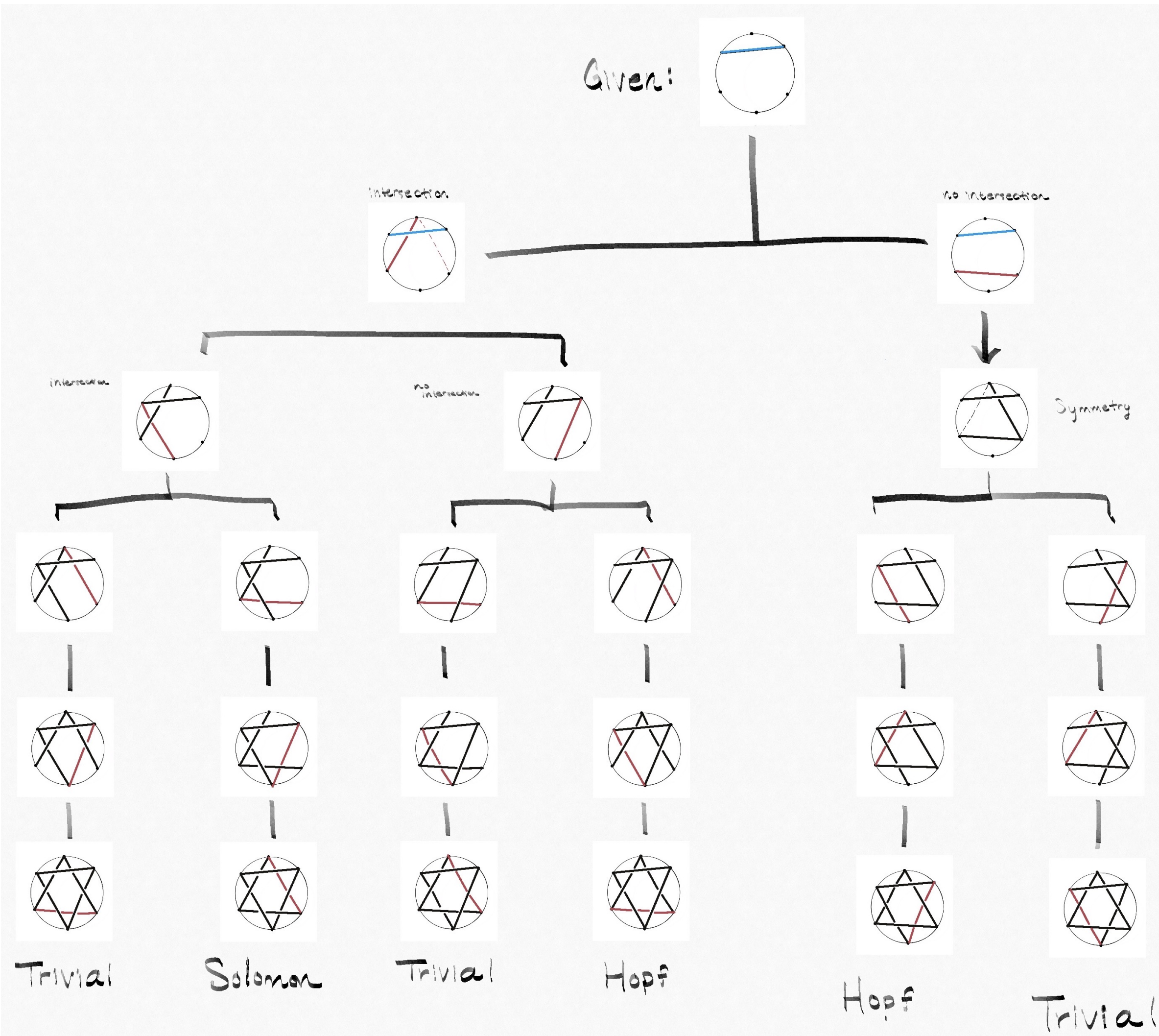


Figure 21: The ABABAB case.

Since two of the three edges in the B triangle intersect with the first edge we chose from triangle A, the left branch has a 2/3 chance of occurring while the right branch has a 1/3 probability of occurring.

The left branch then splits again into two more different branches, each with 1/2 probability of happening, depending on which edge of A we chose. The left most of these branches then branches once more into two different sub-branches. The left most branch forms an unlink. The branch to right of the unlink forms a Solomon 'knot'.

Now, the branch which follows to the right from the intersecting edge branch also splits into two different branches. The leftmost of these two branches becomes a Trivial link, while the other branch forms a Hopf link.

Now, the non-intersecting branch (which occurs with probability 1/3) splits into two distinct branches. The left branch forms a Hopf link, while the right branch forms a trivial link. Every time the tree splits into two branches, we have to multiply a 1/2 to the probability. Finally, multiplying all the probabilities, we find that there is a 1/6 chance of getting a Solomon link, 1/2 chance for a getting a trivial link, and a 1/3 chance of getting a Hopf link.

6 Code

All of our code simulations could be found at `https://github.com/Xyxyxx/RHREU2020`.

References

[1] Adams, C. C. The Knot Book. The American Mathematical Society, 1994.

[2] Alikoski, H. A. Über das sylvertsche Vierpunktproblem. Ann. Acad. Sci. Fenn. 51, 7 (1939), 1–10.

[3] Arsuaga, J., Blackstone, T., Diao, Y., Karadayi, E., and Saito, M. Linking of uniform random polygons in confined spaces. J. Phys. A 40, 9 (2007), 1925–1936.

[4] Beals, M., Gross, L., and Harrell, S. DNA and Knot Theory.

[5] Blaschke, W. Vorlesungen über Differential Geometrie II. Affine Differentialgeometrie (1923).

[6] Conway, J. H., and Gordon, C. M. Knots and links in spatial graphs. J. Graph Theory 7, 4 (1983), 445–453.

[7] joriki (https://math.stackexchange.com/users/6622/joriki). How is the number of points in the convex hull of five random points distributed? Mathematics Stack Exchange. URL:https://math.stackexchange.com/q/282148 (version: 2013-01-19).

[8] Numberphile. How DNA unties its own knots - Numberphile.

[9] Rowland, D. Classification of book representations of K_6. Journal of Knot Theory and Its Ramifications 26 (2017), 1750075.

[10] Solomon, H. Geometric probability. SIAM, 1978.

[11] Valtr, P. Probability that n random points are in convex position. Discrete & Computational Geometry 13, 3 (Jun 1995), 637–643.

[12] Warley, J. Random links via simplex embeddings, 2017.

[13] Y. Diao, N. J. P., and Sumners, D. W. L. On random knots. J. Knot Theory Ramifications 3, 3 (1994), 419–429.