Rose-Hulman Undergraduate Mathematics Journal

Volume 24
Issue 2
Article 3

# The Mean Sum of Squared Linking Numbers of Random Piecewise-Linear Embeddings of $K_n$

Yasmin Aguillon
*University of Notre Dame*, yaguillo@nd.edu

Xingyu Cheng
*University of North Carolina at Chapel Hill*, xcheng1@unc.edu

Spencer Eddins
*University of Kentucky*, spencer.p.eddins@gmail.com

Pedro Morales
*Purdue University*, moralep@purdue.edu

## The Mean Sum of Squared Linking Numbers of Random Piecewise-Linear Embeddings of $K_n$

### Cover Page Footnote

This project was a result from the 2020 Rose-Hulman REU on spatial graph embeddings. We would like to acknowledge the National Science Foundation and Grant No. DMS-1852132 for funding for this REU. In particular, we would like to thank our REU supervisor and mentor Dr. Kenji Kozai.

ℝose-Hulman Undergraduate Mathematics Journal
VOLUME 24, ISSUE 2, 2023

# The Mean Sum of Squared Linking Numbers of Random Piecewise-Linear Embeddings of $K_n$

By *Yasmin Aguillon*, *Xingyu Cheng*, *Spencer Eddins*, and *Pedro Morales*

**Abstract.** DNA and other polymer chains in confined spaces behave like closed loops. Arsuaga et al. [2] introduced the uniform random polygon model in order to better understand such loops in confined spaces using probabilistic and knot theoretical techniques, giving some classification on the mean squared linking number of such loops. Flapan and Kozai [6] extended these techniques to find the mean sum of squared linking numbers for random linear embeddings of complete graphs $K_n$ and found it to have order $\Theta(n(n!))$. We further these ideas by inspecting random piecewise-linear embeddings of complete graphs and give introductory-level summaries of the ideas throughout. In particular, we give a model of random piecewise-linear embeddings of complete graphs where the number of line segments between vertices is given by a random variable. We find further that in our model of the random piecewise-linear embeddings, the order of the expected sum of squared linking numbers is still $\Theta(n(n!))$.

## 1 Introduction

A motivation behind the study of knot theory stems from the discipline's many applications to molecular biology, especially in the study of DNA. Long polymer strands (such as DNA) are packed tightly within the nucleus of a cell such that if the nucleus of a cell is the size of a basketball, the DNA inside it is equivalent to 3 km of fishing line [9]. One can easily imagine how these large DNA molecules might inevitably become entangled when they are all compressed within the relatively small confined space of the nucleus. There are theorems by Diao, Pippenger, Sumners and Whittington which state that in different models as the length of a chain approaches infinity, the knotting probability goes to 1 [10][8][7][5]. This supports the idea that the likelihood of DNA becoming entangled within the nucleus is extremely high.

This entanglement can lead to problems, as it makes essential processes like DNA replication and transcription more difficult for the cell to perform. However, nature solves this dilemma using enzymes, called topoisomerases, to cut through knots in the DNA and then to reconnect these DNA strands in a more orderly fashion. This process topologically alters the knots in the DNA [3][1, pg. 182].

*Mathematics Subject Classification.* 11A41
*Keywords.* $K_n$, knot theory, linking number, DNA, topology, geometry, graph theory, geometric topology.

An important question many biochemists and molecular biologists have sought to answer is, “How do enzymes act upon DNA?" To study this, biochemists and molecular biologists have been studying circular DNA in particular. There are two reasons for this. First, knots on circular DNA will not slip off the ends as they could within a regular, unclosed 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.

Knot theory and topology have become invaluable tools for scientists and researchers by providing scientists with a quantitative way to measure the properties of knots in DNA. A good example of such a tool from knot theory is the linking number, which is a link invariant that describes how two closed curves are “linked" or “tangled" in 3-space. A link in knot theory is defined as a set of closed curves or hoops. Biochemists have used this invariant to quantitatively study and analyze the entanglement of DNA.

To model long polymer chains in confined spaces, Arsuaga et al. introduced the Uniform Random Polygon (URP) model in [2]. This model represents polymer chains as random polygons, which are constructed from points generated uniformly and independently in a confined convex space. They prove that the mean squared linking number of two uniform random polygons in a cube (or in any other symmetric convex space) grows with order $O(n^2)$, where $n$ is the number of segments which make up the two uniform random polygons. In this paper, we extend the results of Arsuaga et al. to links of random chain lengths, and provide applications to the complete graph $K_6$.

Furthermore, Flapan and Kozai [6] proved that the mean sum of squared linking numbers of linear embeddings (i.e. every edge of the graph gets mapped to a line segment in $\mathbb{R}^3$) of the complete graph $K_n$ is of order $\Theta(n(n!))$, and we use similar techniques to show in Corollary 4.9 that this result can be expanded to piecewise-linear embeddings of $K_n$, generated under the URP model [2] as well.

## 2 Background

We begin our discussion with the definition of the fundamental object of our study: links. Links are essentially a set of closed loops or curves in 3-space.

**Definition 2.1** (Links)**.** A *link* of $m$ components is a subset of $\mathbb{R}^3$ that consists of $m$ disjoint, simple closed curves. A link of one component is called a *knot*. Further, if all components of a link L are oriented, then we say that L is an *oriented link*.

We will sometimes call the components of a link *cycles* to emphasize the closed nature of the curve (note the connection to graph theory). We will do this especially when discussing $K_6$ since all cycles of interest in $K_6$ will be ‘defined’ by three points. In the case when these cycles are defined by three points, we will refer to them as 3-cycles.

As with other topological spaces, it is natural to ask in what sense are links equivalent. That is, we wish to ask for bijective maps between links such that the main properties of links are preserved. This leads us to the notion of ambient isotopy.

**Definition 2.2** (Ambient isotopy)**.** Let $f : \mathbb{R}^3 \times [0,1] \to \mathbb{R}^3$ be a homotopy. We say that $f$ is an *ambient isotopy* if $f(x,0) = id_{\mathrm{M}}(x)$ and each $f_t(x) = f(x,t)$ is a homeomorphism. Two links L and L′ are said to be *ambient isotopic* if there is a homeomorphism $h$ such that $id_{\mathrm{M}}$ and $h$ are ambient isotopic and $h(\mathrm{L}) = \mathrm{L}'$.

In words, this means that the links can continuously be deformed and moved around the space, so long as no two strands pass through each other in 3-D space. Ambient isotopies define an equivalence relation on links, and this notion allows us to make precise exactly what it means for two knots or two links to be the same.

We can picture knots and links by representing them in $\mathbb{R}^2$ as diagrams. These diagrams are formed by projecting the links to 2 dimensions. However, sometimes these projections may be pathological. For example, three strands may cross at the same point rather than having three individual crossings, one for each pair of strands. To avoid these pathologies, we have the notion of a *regular projection*. To define a regular projection, we need to first characterize the various pathologies which we wish to remove.

**Definition 2.3.** A *polygonal link* is a set of connected line segments such that the topological space formed by the set of line segments is a link. We call a line segment of a polygonal link L in $\mathbb{R}^3$ an *edge* of L and an end point of the line a *vertex*.

**Definition 2.4.** Let L be a link and $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ be an orthogonal projection. We call a point $c \in \pi(\mathrm{L})$ a *multiple point* if $\pi^{-1}(c) \cap \mathrm{L}$ contains more than one point. The cardinality of $\pi^{-1}(c) \cap \mathrm{L}$ is called the *order* of $c$ and $c$ is called an $n$*-multiple point* if the order of $c$ is $n$. A two-multiple point is called a *double point*.

Being a multiple point means that in a knot or link diagram, we would have multiple strands crossing at the point in the knot projection. If the order of these multiple points are greater than two, then that means we have three or more strands crossing in the same place. However, our calculations rely on at most two strands ever crossing a single point, since we care about the notions of over- and under-crossings in a projection. Over- and under-crossings do not make sense for multiple strands crossing at the same point, so higher order multiple points are pathological. To rectify this, there is a notion of a regular projection.

**Definition 2.5.** Let $\pi : \mathbb{R}^3 \to \mathbb{R}^2$ be an orthogonal projection. We say $\pi$ is a *regular projection* for a polygonal link L if the following conditions hold:

1. The set of multiple points of the image $\pi(\mathrm{L})$ consists of finitely many double points.
2. No point in the preimage $\pi^{-1}(c) \cap \mathrm{L}$ of any double-point $c \in \pi(\mathrm{L})$ is a vertex of L.

Essentially, this definition means that we only have finitely many intersections in the projection, and that those intersections are only between two strands. It turns out that we can always get a regular projection of a link since links can be continuously deformed to one that has a regular projection. Now that we have the necessary definitions for links, we can provide some examples of links along with their projections below.

**Example 2.6** (The trivial link)**.** We have a picture of a projection of the trivial link in Figure 1. This is the simplest link of two components; it is also called the unlink.

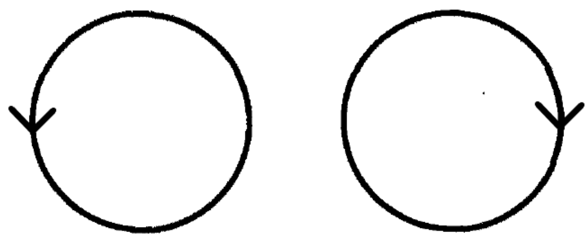

Figure 1: Unlink or Trivial Link

**Example 2.7** (The Hopf link)**.** A projection of the Hopf link is pictured in Figure 2. It is two hoops intertwined with each other once, and is the second simplest link of two components after the unlink.

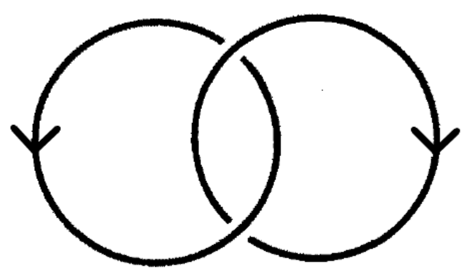

Figure 2: Hopf Link

Next, we define the linking number of two cycles $C_1$ and $C_2$. The linking number is a convenient link invariant which will allow us to quantify the complexity of links. The linking number is called a link invariant because any two (ambient) isotopic links share the same linking number.

**Definition 2.8.** Suppose $L = \{C_1, C_2\}$ is an oriented link with two components. We then take a regular projection of L, and we look at the crossing points between these two cycles.

That is, we look at the crossings in the diagram of the regular projection where the two strands come from different components of the link (i.e. one from $C_1$ and the other from $C_2$), and we do not consider crossings that come from the same component intersecting itself. For each of these crossings, there will be an overstrand and understrand. We look at each crossing from the perspective of the overstrand, and we define the sign of each crossing as

1. If the understrand is moving left relative to the perspective of the overstrand, we assign this crossing the value $+1$.
2. If the understrand is moving right relative to the perspective of the overstrand, we assign this crossing the value $-1$.

The values of the crossings are illustrated in Figure 3. The crossing pictured on the left of that figure is assigned $+1$, while the crossing pictured on the right of that figure is assigned $-1$. We say that a crossing is positive if it is assigned the value $+1$ and that a crossing is negative if it is assigned the value $-1$.

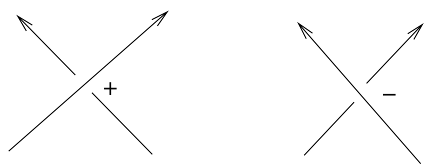


Figure 3: Positive and negative crossings

The *linking number* is defined as the sum of all the positive and negative crossings divided by 2; it is denoted as $\text{lk}(C_1, C_2)$, and it is given in equation form by

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

It is worth noting that the linking number will always be an integer, since

$$(\#\text{ of positive crossings}) - (\#\text{ of negative crossings})$$

will always be even.

Now for general two-component links (which may or may not be oriented), we note that for all four possible orientations that we can give such links, we will get the same linking number up to sign. This is because if we flip the orientation of any single component in the link, we will change the sign of every single crossing in the link, which means that the linking number of the new link will be the negative of the original link. This means that the linking number is invariant up to sign under changes in orientation.

## 3 Uniform random polygon model

The uniform random polygon (URP) model introduced by Arsuaga et al. in [2] generates two random polygons in a confined convex space. Because these two polygons are two closed curves, they form a link in the confined space. Therefore we can study their nature as links. A precise description of the model is given in the following definition.

**Definition 3.1** (The uniform random polygon model [2])**.** A set of points $R_1 = \{v_1, \dots, v_n\}$ are picked uniformly and independently in a bounded convex space $K \subset \mathbb{R}^3$. The points are connected by line segments so that $v_1$ is connected to $v_2$, $v_2$ to $v_3$, ..., and $v_{n-1}$ is connected to $v_n$; $v_n$ is further connected to $v_1$ so that the points $v_1$, ..., $v_n$ form a polygon in $\mathbb{R}^3$. Another set of points $R_2 = \{v'_1, \dots v'_n\}$ are also picked randomly and uniformly and connected similarly to the points of $R_1$ such that the points of $R_2$ forms a second polygon. We will use $R_1$ and $R_2$ to denote both the set of points defining the polygon as well as (slight abuse of notation) the polygon itself.

Note that since $R_1$ and $R_2$ form two closed curves, this means that $L = \{R_1, R_2\}$ forms a link. We will consider the projection diagram of $L = \{R_1, R_2\}$ as a regular projection (since every link projection is isotopic to a regular projection with probability one).

Consider the projection of two disjoint edges, $\ell_1$ and $\ell_2$. 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 is defined in [2] 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 $v_{i+1}$. Arsuaga et. al defined a random variable $\varepsilon$ in the following ways: $\varepsilon = 0$ if the projection of $\ell_1$ and $\ell_2$ 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. We have defined positive and negative crossings as in Figure 3. This means that $\varepsilon$ represents the crossing number of two random edges from the two uniform random polygons.

For simplicity, in this paper we use a uniform distribution for the independent random points that form the uniform random polygons. We also assume that we are choosing points from the unit cube $C^3 = [0,1]^3$. It is worth saying that even if the distribution is changed, every argument in this section still holds, only the hypothesis of independence and identical distribution needs to hold.

Arsuaga et. al proved that if $R_1$ and $R_2$ are two random polygons generated by $n$ points each, the mean squared linking number between these two polygons is given by

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

where $q > 0$, and $\varepsilon_{ij}$ is the crossing sign between edges $\ell_i$ of $R_1$ and $\ell'_j$ of $R_2$ after labeling the edges of $R_1$ as $\ell_1, \dots, \ell_n$ and respectively the edges of $R_2$ as $\ell'_1, \dots, \ell'_n$. This is the main result we wish to generalize for our paper.

To prove the above theorem, we need the following lemma, which gives us sufficient conditions to cancel out terms in the expected value. Although the proof is given in [2], we give an expanded proof with additional details as this is helpful for the rest of our discussion.

**Lemma 3.2.** *(Modified version of Lemma 1 in [2]) Consider 4 edges: $\ell_1, \ell_2, \ell_1'$, and $\ell_2'$ with randomly and independently chosen end points. Some of these edges may be identical or share a common end point. Define $\varepsilon_1$ as the number $\varepsilon$ between $\ell_1$ and $\ell_1'$, and let $\varepsilon_2$ be the number $\varepsilon$ between edges $\ell_2$ and $\ell_2'$, where $\varepsilon$ is defined above as the crossing between two disjoint edges.*

1. *If the end points of $\ell_1$, $\ell_2$, $\ell_1'$ and $\ell_2'$ are distinct, then $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0$ (This is the case when there are 8 random points involved);*

2. *If $\ell_1 = \ell_2$, and the end points of $\ell_1'$ and $\ell_2'$ are distinct (this is the case when there are 6 independent points involved with 3 distinct edges), then $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0$.*

3. *If $\ell_1$ and $\ell_2$ are adjacent and $\ell_1'$ and $\ell_2'$ are distinct (7 independent random points) then $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0$.*

4. *In the case that $\ell_1 = \ell_2$ and $\ell_1'$ and $\ell_2$ are distinct (so there are only 5 independent random points involved), let $u = \mathrm{E}(\varepsilon_1\varepsilon_2)$. In the case where $\ell_1$ and $\ell_2$ share a common point and where $\ell_1'$ and $\ell_2'$ also share a common point (so there are 4 edges with 6 independent points involved in this case), let $\mathrm{E}(\varepsilon_1\varepsilon_2) = v$. Lastly, let $p = \mathrm{P}(\varepsilon = 1)$ for $\varepsilon$ being the crossing sign of any two edges without restriction. Define $q := p + 2(u + v)$, then we have that $q > 0$.*

*Proof.*

1. Since the vertices in $\varepsilon_1$ and $\varepsilon_2$ are chosen without dependence on each other, we know $\varepsilon_1$ and $\varepsilon_2$ must be independent variables satisfying $\mathrm{E}(\varepsilon_1\varepsilon_2) = \mathrm{E}(\varepsilon_1)\mathrm{E}(\varepsilon_2)$. For any crossing $\varepsilon$ we have that $\mathrm{E}(\varepsilon) = 0$, since the probability $\mathrm{P}(\varepsilon = 1) = \mathrm{P}(\varepsilon = -1)$. Then $\mathrm{E}(\varepsilon_1\varepsilon_2) = \mathrm{E}(\varepsilon_1)\mathrm{E}(\varepsilon_2) = 0$.

2. For each configuration where both $\ell_1'$ and $\ell_2'$ 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 via permuting the points (See Figure 4 below). For 4 of these 8 orientations, $\varepsilon_1\varepsilon_2 = 1$, and for the other 4, $\varepsilon_1\varepsilon_2 = -1$. These permutations are equally likely, so $\mathrm{E}(\varepsilon_1\varepsilon_2) = 0 \cdot \mathrm{P}(\varepsilon_1\varepsilon_2 = 0) + \mathrm{P}(\varepsilon_1\varepsilon_2 \neq 0) \cdot (1(\frac{4}{8}) - 1(\frac{4}{8})) = 0$.

3. For the scenario when $\ell_1, \ell_2$ are adjacent and $\ell_1', \ell_2'$ are distinct, we notice that it is equally probable that edge $\ell_1'$ has reverse order by symmetry. Notice that this changes the sign of $\varepsilon_1$ and thus the sign of $\varepsilon_1\varepsilon_2$ too. Then $\varepsilon_1\varepsilon_2$ is equally likely to be positive or negative. That is, $\mathrm{E}(\varepsilon_1\varepsilon_2)$ is also 0.

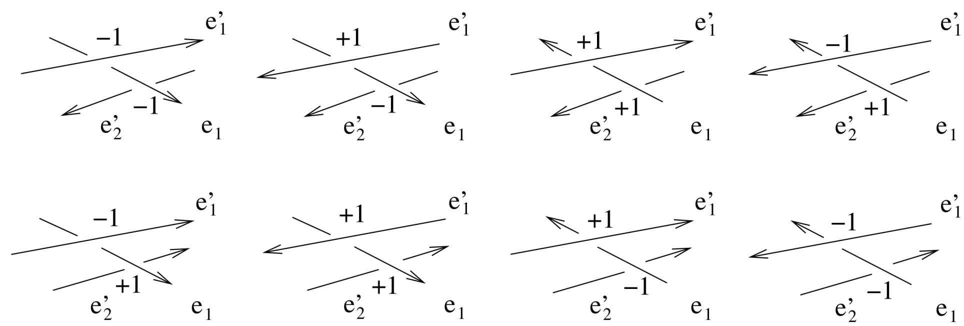


Figure 4: For each configuration described in Lemma 3.2 part 2 where $\varepsilon_1\varepsilon_2 \neq 0$, there are 8 symmetric ways of assigning the orientations ([2]).

4. Let us consider two random triangles. The first triangle will consist of sides $\ell_1, \ell_2, \ell_3$ while the second triangle will consist of sides $\ell'_1, \ell'_2, \ell'_3$. Let $\varepsilon_{ij}$ be the number $\varepsilon$ between edges $\ell_i$ and $\ell'_j$. 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 $\mathrm{Var}(\varepsilon) = \mathrm{E}(\varepsilon^2) - [\mathrm{E}(\varepsilon)]^2$. However, given that $\mathrm{P}(\varepsilon = 1) = \mathrm{P}(\varepsilon = -1) = p$, it follows that $\mathrm{E}(\varepsilon) = 0$. Therefore, we get that $\mathrm{Var}(\varepsilon) = \mathrm{E}(\varepsilon^2)$.
Now, by considering the variance of the summation $\sum_{i,j=1}^{3} \varepsilon_{ij}$, we have

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

This equation can be simplified into the 3 following sums:

$$\begin{aligned}\mathrm{Var}\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right) =& \mathrm{E}\left(\left(\sum_{i,j=1}^{3} \varepsilon_{ij}\right)^2\right)\\ &= \sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}^2) + 2\sum_{i,j=1}^{3} \left[\mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j-1)}) + \mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j+1)})\right]\\ &\quad + 2\sum_{i,j=1}^{3} \left[\mathrm{E}(\varepsilon_{ij}\varepsilon_{i+1,j+1}) + \mathrm{E}(\varepsilon_{ij}\varepsilon_{i-1,j+1})\right],\end{aligned}$$

where the indices in this sum are all taken modulo 3. Note that since

$$\sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j-1)}) = \sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}\varepsilon_{(i-1)j}),$$

we can omit one of these terms (for us we omitted $\sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}\varepsilon_{(i+1)j})$) to get the term

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

and similarly for the $2\sum_{i.j=1}^{3} \mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j+1)})$ term.

These 3 sums can be simplified even further. Each term in the first summation, $\sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}^2)$, 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 $\varepsilon$ in this scenario can only equal $\pm 1$ or 0, we know $\varepsilon_{i,j}^2$ will always equal 1 or 0. That is, $\varepsilon^2$ equals 1 if the two edges intersect and 0 otherwise. Since we previously defined the probability of two edges intersecting within the projection diagram as $2p$, for each term in the sum $\mathrm{E}(\varepsilon_{ij})^2 = 1 \cdot 2p = 2p$. So, the first sum

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

Each term of the second sum, $2\sum_{i,j=1}^{3} \mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j-1)}) + \mathrm{E}(\varepsilon_{ij}\varepsilon_{i(j+1)})$, yields $u$, where $u = \mathrm{E}(\varepsilon_1\varepsilon_2)$ in the case that the edges $\ell_1 = \ell_2$ and the edges $\ell_1'$ and $\ell_2'$ 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.

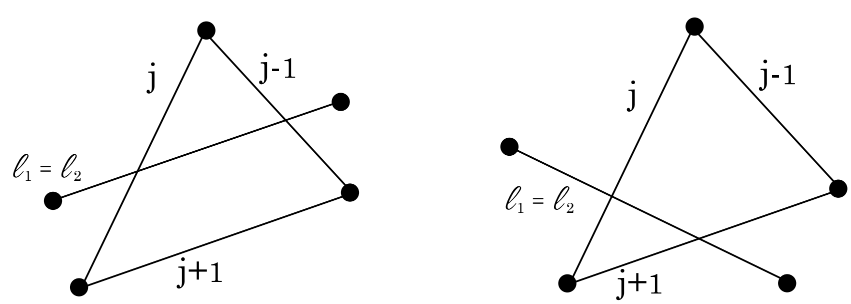


Figure 5: Let the disjoint edge be $\ell_1 = \ell_2$. In the left triangle, edge $j$ and $j-1$ are the adjacent edges that share a point i.e. $\ell_1'$ and $\ell_2'$. For the right triangle edge $j$ and $j+1$ are the adjacent edges that share a point i.e. $\ell_1'$ and $\ell_2'$.

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 $\varepsilon$'s is switched. This is analogous to expanding the expression $(a+b)^2$ in that we get $ab + ba = 2ab$. We multiply the sum by 2 to account for both of these terms. The sum will yield a total of 18 terms. Thus,

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

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 $\ell_1$ and $\ell_2$ share a common point and $\ell_1'$ and $\ell_2'$ 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.

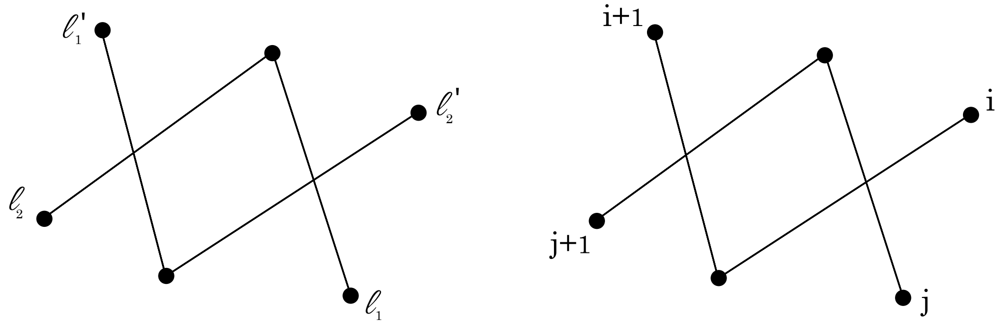


Figure 6: Illustration of the $v$ case.

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 $\varepsilon$ is switched and also results in a total of 18 different terms. So,

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

Thus,

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

Since $\mathrm{Var}\left(\sum_{i,j=1}^{3}\varepsilon_{ij}\right)>0$, this implies that $q=p+2(u+v)>0$. Thus, the claim of Arsuaga et al. holds.

□

As observed in the proof of the above lemma, when the difference of indices $i - i'$ mod $n$ and $j - j'$ mod $n$ are greater than 1, then $E(\varepsilon_{ij}\varepsilon_{i',j'}) = 0$. Using this idea, Arsuaga et al. proved the following theorem about mean squared linking numbers.

**Theorem 3.3.** *(Theorem 1 of [2]) 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 3.2.*

The main argument for this theorem relies on Lemma 3.2. In that lemma, we reduced all the possible crossing configurations into four cases, and we noted that when passing to the expected value, only the fourth case of the above lemma becomes nonzero. Simplifying the summation will then yield the expression $\frac{1}{2}n^2q$.

### 3.1 Generalizations of the results from the uniform random polygon model

We extend the uniform random polygon model of Arsuaga et al. [2] in the following manner. Let $R_1 = \{v'_1, \ldots, v'_N\}$ and $R_2 = \{v'_1, \ldots, v'_M\}$, where N and M are two random variables and each $v_i$ and $v'_j$ are in $\Omega$, where $\Omega \subset \mathbb{R}^3$ is a bounded convex set. We prove that the mean squared linking number of such uniform random polygons is of the form $\frac{1}{2}E(M)E(N)q$ when N and M are independent. In the case of $N \equiv M$ (that is, $N = M$ in all instances), we have that the mean squared linking number is of form $\frac{1}{2}E(N^2)q$.

Before we begin the proofs of the above formulas, we will require a couple of standard results from probability theory.

**Theorem 3.4** (Law of total expectation)**.** *Let X, Y be random variables. Then*

$$E(X) = E(E(X|Y))$$

*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 above result can be rewritten as

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

**Theorem 3.5.** *Let X be a random variable with probability mass function $f_X$, and $g$ be a function of the random variable X. Then we have that*

$$E(g(X)) = \sum_x g(x)\,f_X(x).$$

This result is sometimes referred to as the law of the unconscious statistician and will be helpful to prove results about mean squared linking numbers for a novel modification of the uniform random polygon model.

In addition to some standard results from probability theory, we will also need the following theorem on uniform random polygons of different chain lengths. This result was proved in [6] as an application of results proven in [2].

**Theorem 3.6** ([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 [2].*

The above theorems are sufficient for us to describe the average linking behavior of our random uniform random polygons. And we will state our result below.

**Theorem 3.7.** *Let* M *and* N *be two random variables and let* $R_1$ *and* $R_2$ *be two uniform random polygons with vertices picked in* $\Omega$ *with* M *and* N *vertices respectively, where* $\Omega$ *is some convex region in* $\mathbb{R}^3$. *Further, let* $q = p+2(u+v)$ *be as in Lemma 3.3.*

1. *In the case that* $M \equiv N$*, the mean squared linking number is*

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

2. *In the case that* M *and* N *are independent, the mean squared linking number is*

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

*Proof.* 1. This is a result of a straightforward calculation involving conditional expectation. Note that by Theorem 3.5 we have that

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

The last equality comes from the fact that $E(g(X)) = \sum_x g(x)P(X = x)$ as per Theorem 3.5.

2. This result is achieved by a continuous application of conditional expectation. First note that we can take the expectation of N to get

$$\begin{aligned}\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{\mathrm{N}}\sum_{j}^{\mathrm{M}}\varepsilon_{ij}\right)^{2}\right) &= \mathrm{E}\left(\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{\mathrm{M}}\sum_{j}^{\mathrm{N}}\varepsilon_{ij}\right)^{2}\middle|\mathrm{N}\right)\right)\\ &= \mathrm{E}\left(\mathrm{E}\left(\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{\mathrm{M}}\sum_{j}^{\mathrm{N}}\varepsilon_{ij}\right)^{2}\middle|\mathrm{N}\right)\middle|\mathrm{M}\right)\right).\end{aligned}$$

We now use Theorem 3.4 on M to get the expression to equal

$$\sum_{m\geq 3}\mathrm{E}\left(\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{\mathrm{M}}\sum_{j}^{\mathrm{N}}\varepsilon_{ij}\right)^{2}\middle|\mathrm{N}\right)\middle|\mathrm{M}=m\right)\mathrm{P}(\mathrm{M}=m).$$

Applying the substitution to $\mathrm{M}=m$ to sum inside the expected value, we get that the sum becomes

$$\sum_{m\geq 3}\mathrm{E}\left(\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{m}\sum_{j}^{\mathrm{N}}\varepsilon_{ij}\right)^{2}\middle|\mathrm{N}\right)\right)\mathrm{P}(\mathrm{M}=m).$$

From here, we can apply the inner expected value to get the sum to equal

$$\sum_{m\geq 3}\left[\sum_{n\geq 3}\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{m}\sum_{j}^{\mathrm{N}}\varepsilon_{ij}\right)^{2}\middle|\mathrm{N}=n\right)\mathrm{P}(\mathrm{N}=n)\right]\mathrm{P}(\mathrm{M}=m).$$

Applying the substitution $\mathrm{N}=n$ yields

$$\sum_{m\geq 3}\left[\sum_{n\geq 3}\mathrm{E}\left(\left(\frac{1}{2}\sum_{i}^{m}\sum_{j}^{n}\varepsilon_{ij}\right)^{2}\right)\mathrm{P}(\mathrm{N}=n)\right]\mathrm{P}(\mathrm{M}=m).$$

After applying Theorem 3.6 to the above sum we get that the sum becomes

$$\sum_{n}\sum_{m}\frac{1}{2}mnq\mathrm{P}(\mathrm{N}=n)\mathrm{P}(\mathrm{M}=m).$$

Factoring out constants, we get the following simplifications

$$\begin{aligned}\frac{1}{2}q\sum_{n}\left[n\mathrm{P}(\mathrm{N}=n)\sum_{m}m\mathrm{P}(\mathrm{M}=m)\right] &= \frac{1}{2}q\sum_{n}n\mathrm{P}(\mathrm{N}=n)\mathrm{E}(\mathrm{M})\\ &= \frac{1}{2}\mathrm{E}(\mathrm{M})\,\mathrm{E}(\mathrm{N})\,q.\end{aligned}$$

□

The motivation for the above theorem and models comes from a desire to model curved embeddings of $K_6$. While the uniform random polygon model of [2] only deals with the case of straight line embeddings, this case allows us to join each vertex of $K_6$ by a random polygonal curve in space. Since these polygonal curves can approximate arbitrary tame (e.g. differentiable) curves in space, we can intuitively take this to be at least a crude approximation of curved embeddings of $K_6$.

## 4 Spatial embeddings of $K_n$

We now turn our attention to spatial embeddings of complete graphs in space. We begin our discussion with some preliminary notions from graph theory.

**Definition 4.1.** 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 such that no two edges intersect and no two vertices map to the same point.

**Definition 4.2.** 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 this graph as $K_n$.

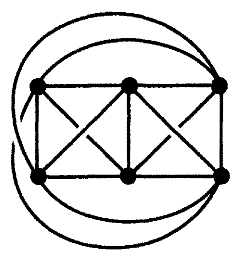

Figure 7: The complete graph $K_6$

For the simplest case consider an arbitrary spatial embedding of $K_6$. An important fact about $K_6$ 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 six vertices in $K_6$, the other three 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 represent nontrivial links, i.e. they are not (ambient) isotopic to the unlink.

Conway and Gordon [4] studied this case and a similar one for $K_7$. In doing so, they proved that every spatial embedding of $K_6$ has a nontrivial link and every spatial embedding of $K_7$ has a nontrivial knot. Our interest is in the former result.

**Theorem 4.3** ([4])**.** *Every spatial embedding of* $K_6$ *has a nontrivial link.*

### 4.1 Applications of random polygon model methods to $K_6$ and $K_n$

With tools from the uniform random polygon model, and our extensions thereof, we will now be able to apply these models to spatial embeddings of $K_6$. Conway and Gordon [4] proved that any spatial embedding of $K_6$ must have at least one nontrivial link, particularly with an odd linking number, and further that the sum of linking numbers over all pairs of cycles in spatial embeddings of $K_6$ must be odd. However, Conway and Gordon's proof does not give further information about the explicit values of such linking numbers. The linking number provides information about the complexity of a link or a set of links, and therefore it is important to see, given a random embedding, how complex the embedding is in the sense of the linking number. The uniform random polygon model and our extensions provide one such tool to analyze this complexity.

We begin our discussion on applications of random polygon methods to $K_6$ with a definition of random piecewise-linear embeddings.

**Definition 4.4.** A *random piecewise-linear embedding* of a graph G is an embedding of G into space where the vertices of G are placed randomly into a confined space $\Omega \subset \mathbb{R}^3$ and the edges are represented by non-intersecting, piecewise-linear curves.

Both linear and piecewise-linear embeddings are in some sense very simple, and we can view cycles from such embeddings in a confined space as uniform random polygons. For linear embeddings, each closed loop of $K_6$ will be composed of three vertices, and hence these polygons will be triangles. There are $10 = \frac{1}{2}\binom{6}{3}$ unordered pairs of disjoint triangles from this set. This leads us to the following result from [6].

**Theorem 4.5.** *Let six points $\{v_1, \ldots, v_6\}$ be chosen independently and uniformly in the cube $C^3 = [0,1]^3$. We then connect each point $v_i$, $1 \le i \le 6$ to each point in $\{v_1, \ldots, v_6\} \setminus \{v_i\}$ with straight line segments. This represents a linear embedding of $K_6$, and the mean sum of squared linking numbers of such an embedding is given by*

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

*where $q > 0$ is the constant defined in Lemma 3.2, and we are summing over the ten pairs of triangles $C_1$ and $C_2$.*

Now that we have a good idea of the random linear embeddings of $K_6$, we can consider random piecewise-linear embeddings of $K_6$. This will provide an intuition for the general case of $K_n$.

**Theorem 4.6.** *Let $C_1$ and $C_2$ be two uniform random cycles in $K_6$. Define $X_{ij}$ as the random variable for the number of line segments in our linear embedding over the edge $\{v_i, v_j\}$, and further suppose for all $i, j$ that $E(X_{ij}) = U$ for some $U$, such as the case when* *the variables are identically distributed. Also, let the* $X_{ij}$*'s be pairwise independent. Then, the mean sum of squared linking numbers is given by*

$$E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{r}^{M(C_1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right)=45\,q\,U^2,$$

*where*

$$M(C_\iota):=\sum_{\substack{\{v_i,v_j\}\subset C_\iota\\ i\neq j}}X_{ij}.$$

*Proof.* We begin by iterating a sum over the $M(C_1)$ and $M(C_2)$ line segments in $C_1$ and $C_2$ respectively. Because these are just two cycles with a random number of line segments (i.e. two uniform random polygons), we can apply Theorem 3.7.

$$\begin{aligned}E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{r}^{M(C_1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right)&=\sum_{C_1,C_2}\left(E\left(\left(\frac{1}{2}\sum_{r}^{M(C1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right)\right)\\&=\sum_{C_1,C_2}\left(\frac{1}{2}\,q\,E(M(C_1))\,E(M(C_2))\right).\end{aligned}$$

Since each cycle has three edges, there will be three random variables for each selection of $C_1$ and $C_2$. Without loss of generality, consider the case when $C_1$ is the cycle containing $\{v_1, v_2, v_3\}$ and $C_2$ is the cycle containing $\{v_4, v_5, v_6\}$. Then $M(C_1) = X_{12} + X_{13} + X_{23}$ and $M(C_2) = X_{45} + X_{46} + X_{56}$. 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

$$E(M(C_1)) = E(X_{12} + X_{13} + X_{23}) = E(X_{12}) + E(X_{13}) + E(X_{23}).$$

A similar decomposition gives $E(M(C_2)) = E(X_{45})+E(X_{46})+E(X_{56})$. By hypothesis, we also have that $E(X_{ij}) = U$ for all $i, j$, so $E(M(C_1)) = 3U$ and similarly $E(M(C_2)) = 3U$. Hence, we can rewrite our summation as

$$\begin{aligned}E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{r}^{M(C_1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right)&=\sum_{C_1,C_2}\frac{1}{2}q(3U)(3U)\\&=\sum_{C_1,C_2}\left(\frac{9}{2}\,q\,U^2\right)\\&=10\left(\frac{9}{2}\,q\,U^2\right)\\&=45\,q\,U^2,\end{aligned}$$

where we are summing over the 10 different pairs of cycles $\{C_1, C_2\}$ in a random piecewise-linear embedding of $K_6$, hence the coefficient 10. □

It would be desirable to generalize the above theorem to $K_n$ and without the restriction of having the same expected number of line segments per edge. For $n > 6$, cycles may have more than 3 vertices, meaning the number of vertices in a cycle is no longer constant. Moreover, there is now more than one cycle containing the same set of 4 or more vertices due to the different orders they can be arranged in. Surprisingly, this only results in a small change in structure and a multiplicative factor we call $D(n)$.

Our following result characterizes the average linking behavior of random piecewise-linear embeddings of $K_n$. Here we build on the work of [6] and [2].

**Theorem 4.7.** *Define* $X_{ij}$ *as the independent random variable for the number of segments in the edge from* $v_i$ *to* $v_j$ *in* $K_n$. *Then the mean sum of squared linking numbers is given by*

$$E\left(\sum_{C_1,C_2}\left(\mathrm{lk}(C_1,C_2)^2\right)\right) = \frac{D(n)}{4} q \sum_{\substack{\{i,j\}\\ i\neq j}} E(X_{ij}) \sum_{\substack{k\neq l\\ k,l\notin\{i,j\}}} E(X_{kl}),$$

*where* $D(n)$ *is some positive integer depending on* $n$.

*Proof.* Let $C_\iota$ be a closed loop in $K_n$, and define $M(C_\iota)$ to be the number of segments in $C_\iota$, given by $\sum_{\substack{v_i,v_j\in C_\iota\\ i\neq j}} X_{ij}$. Index the line segments in each $C_\iota$ and denote $\varepsilon_{rs}$ as the crossing between line segments $r$ in $C_1$ and $s$ in $C_2$. By Theorem 3.7, we have

$$\begin{aligned}
E\left(\sum_{C_1,C_2}\left(\frac{1}{2}\sum_{r}^{M(C_1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right) &= \sum_{C_1,C_2}\left(E\left(\left(\frac{1}{2}\sum_{r}^{M(C_1)}\sum_{s}^{M(C_2)}\varepsilon_{rs}\right)^2\right)\right)\\
&= \sum_{C_1,C_2}\left(\frac{1}{2}q\,E(M(C_1))\,E(M(C_2))\right)\\
&= \sum_{C_1,C_2}\left(\frac{1}{2}q\,E\left(\sum_{\substack{v_i,v_j\in C_1\\ i\neq j}} X_{ij}\right)E\left(\sum_{\substack{v_k,v_l\in C_2\\ k\neq l}} X_{kl}\right)\right)\\
&= \frac{1}{2}q\sum_{C_1,C_2}\left(\sum_{\substack{v_i,v_j\in C_1\\ i\neq j}} E(X_{ij})\right)\left(\sum_{\substack{v_k,v_l\in C_2\\ k\neq l}} E(X_{kl})\right).
\end{aligned}$$

Notice that this expands into a sum of products of pairs of expected values of random variables times some integer. That is, a sum of terms of the form $a_{ijkl}E(X_{ij})E(X_{kl})$ where $a_{ijkl}$ is the number of times the term $E(X_{ij})E(X_{kl})$ appears in the expanded sum. Because each pair of vertices comes from disjoint pairs of cycles, we have $v_k, v_l \notin \{v_i, v_j\}$, or, in other words, $i, j, k, l$ are distinct. Therefore $a_{ijkl} = 0$ if $i, j, k, l$ are not distinct, and if they are distinct, then $a_{ijkl}$ counts the number of unordered pairs of disjoint cycles

$\{C_1, C_2\}$ such that $\{i, j\}$ and $\{k, l\}$ are edges of $C_1$ and $C_2$ respectively. We now rewrite this as a sum over each nonzero $a_{ijkl}$ term in a few ways:

$$\begin{aligned}
\frac{1}{2}q\sum_{C_1,C_2}\left(\sum_{\substack{v_i,v_j\in C_1\\ i\neq j}}E(X_{ij})\right)\left(\sum_{\substack{v_k,v_l\in C_2\\ k\neq l}}E(X_{kl})\right) &= \frac{1}{2}q\sum_{\{\{i,j\},\{k,l\}\}}a_{ijkl}E(X_{ij})E(X_{kl})\\
&= \frac{1}{16}q\sum_{\substack{i,j,k,l\\ \text{distinct}}}a_{ijkl}E(X_{ij})E(X_{kl})\\
&= \frac{1}{4}q\sum_{\substack{\{i,j\}\\ i\neq j}}\left(E(X_{ij})\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}a_{ijkl}E(X_{kl})\right)
\end{aligned}$$

Note that the second-line summation iterates over the unordered pair of edges $\{\{i, j\}, \{k, l\}\}$ 8 times due to the double counting from swapping the edges, swapping the vertices in the first edge, and swapping the vertices in the second edge. Meanwhile the last-line only double counts by a factor of 2 by swapping the two edges. For the remainder of this paper we will use the last expression. Since $a_{ijkl}$ satisfies the hypothesis of Lemma 4.8, and since we are only iterating over $a_{ijkl}$ with distinct $i, j, k, l$, we have by 4.8 that each $a_{ijkl}$ term is equal to some positive integer dependent only on $n$ which we will call $D(n)$. That is $a_{ijkl} = D(n)$. Then

$$\begin{aligned}
\frac{1}{4}q\sum_{\substack{\{i,j\}\\ i\neq j}}\left(E(X_{ij})\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}a_{ijkl}E(X_{kl})\right) &= \frac{1}{4}q\sum_{\substack{\{i,j\}\\ i\neq j}}\left(E(X_{ij})\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}D(n)E(X_{kl})\right)\\
&= \frac{D(n)}{4}q\sum_{\substack{\{i,j\}\\ i\neq j}}\left(E(X_{ij})\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}E(X_{kl})\right).
\end{aligned}$$

□

We then derive the formula for $D(n)$.

**Lemma 4.8.** *Let $a_{ijkl}$ be the number of unordered pairs of cycles in $K_n$ such that the edge $\{v_i, v_j\}$ and $\{v_k, v_l\}$ appear in opposite cycles. If $i, j, k, l$ are distinct then $a_{ijkl} = D(n)$*

*where*

$$\mathrm{D}(n) = \sum_{a=3}^{n-3} \sum_{b=3}^{n-a} \frac{(n-4)!}{(n-a-b)!} = \sum_{i=6}^{n} \frac{(n-4)!(i-5)}{(n-i)!}.$$

*Proof.* Let $f(s,m)$ be the number of ways to make a cycle with a given edge and $s$ additional vertices chosen from a set of $m$ vertices. Then $f(s,m)$ is the number of ways to choose $s$ vertices from the $m$ vertices, given by $\binom{m}{s}$, times the number of ways the $s+2$ vertices can be arranged in a cycle that includes the given edge.

To find the number of arrangements, note that every cycle can be identified with an ordered list of the vertices along any path around that cycle. One has to be careful since both the starting vertex and direction admits a different ordered list despite coming from the same cycle. In our case, however, we know our cycle contains a fixed edge, say, $\{v_i, v_j\}$. Then we can write each cycle uniquely as an ordered list of the $s+2$ vertices starting with $v_i, v_j, \ldots$ so that there are exactly $s!$ arrangements. Thus, $f(s,m) = s!\binom{m}{s}$.

Since $a_{ijkl}$ equals the number of unordered pairs of cycles where the edges $\{v_i, v_j\}$ and $\{v_k, v_l\}$ show up in opposite cycles, we can then write $a_{ijkl}$ as the sum over all possible $a$ and $b$ of the number of ways $\{v_i, v_j\}$ shows up in a cycle with $a$ vertices and $\{v_k, v_l\}$ shows up in an disjoint cycle with $b$ vertices. This gives $f(a-2, n-4)f(b-2, n-2-a)$ such pairs of unordered cycles for each $a$ and $b$. Note that not all vertices have to be used, so $a+b \leq n$. As well, every cycle must have at least 3 vertices, so we may write that $3 \leq a \leq n-3$ and $3 \leq b \leq n-a$ are the ranges of our sum. Therefore we obtain the sum

$$\begin{aligned} a_{ijkl} &= \sum_{a=3}^{n-3} \sum_{b=3}^{n-a} f(a-2, n-4) f(b-2, n-2-a) \\ &= \sum_{a=3}^{n-3} \sum_{b=3}^{n-a} (a-2)!\,(b-2)! \binom{n-4}{a-2} \binom{n-2-a}{b-2} \\ &= \sum_{a=3}^{n-3} \sum_{b=3}^{n-a} \frac{(n-4)!}{(n-a-b)!}. \end{aligned}$$

Since $a_{ijkl}$ depends only on $n$ and not $i, j, k, l$, we have that $a_{ijkl} = a_{i'j'k'l'}$ whenever both $i, j, k, l$ are distinct and $i', j', k', l'$ are distinct. To reflect this, we define $\mathrm{D}(n) := a_{ijkl}$ for any four distinct vertices $v_i, v_j, v_k, v_l$.

Lastly, in the proof of Theorem 2.4 in [6], Flapan and Kozai show that

$$\sum_{k=3}^{n-3} \sum_{l=3}^{n-k} \frac{n!}{(n-k-l)!} = \sum_{i=6}^{n} \frac{n!(i-5)}{(n-i)!}$$

which implies

$$\sum_{k=3}^{n-3}\sum_{l=3}^{n-k}\frac{1}{(n-k-l)!}=\sum_{i=6}^{n}\frac{(i-5)}{(n-i)!}$$

and thus

$$D(n)=\sum_{a=3}^{n-3}\sum_{b=3}^{n-a}\frac{(n-4)!}{(n-a-b)!}=\sum_{i=6}^{n}\frac{(n-4)!(i-5)}{(n-i)!}.$$

□

**Corollary 4.9.** *Let $X_{ij}$ be the independent random variable for the number of segments in the edge from $v_i$ to $v_j$ in $K_n$, and let $E(X_{ij})=U$ for all $i, j$, such as the case when each $X_{ij}$ is independent and identically distributed. Then the mean sum of squared linking numbers is given by the expression*

$$E\left(\sum_{C_1,C_2}\left(\operatorname{lk}(C_1,C_2)^2\right)\right)=\frac{1}{16}U^2q\sum_{i=6}^{n}\frac{n!(i-5)}{(n-i)!}.$$

*Further, if $U$ is constant with respect to $n$, this has the order $\Theta(n(n!))$.*

*Proof.* Since $E(X_{ij})=U$ for all $i, j$, we may substitute and simplify for this case of Theorem 4.7.

$$\begin{aligned}E\left(\sum_{C_1,C_2}\left(\operatorname{lk}(C_1,C_2)^2\right)\right)&=\frac{D(n)}{4}q\sum_{\substack{\{i,j\}\\ i\neq j}}\left(E(X_{ij})\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}E(X_{kl})\right)\\&=\frac{D(n)}{4}U^2q\sum_{\substack{\{i,j\}\\ i\neq j}}\sum_{\substack{\{k,l\}\\ k\neq l\\ k,l\notin\{i,j\}}}1\\&=\frac{D(n)}{4}U^2q\sum_{\substack{\{i,j\}\\ i\neq j}}\binom{n-2}{2}\\&=\frac{D(n)}{4}U^2q\binom{n}{2}\binom{n-2}{2}.\end{aligned}$$

Substituting $D(n)$ with the first formula from Lemma 4.8 gives

$$E\left(\sum_{C_1,C_2}\left(\operatorname{lk}(C_1,C_2)^2\right)\right)=\frac{1}{4}U^2q\sum_{a=3}^{n-3}\sum_{b=3}^{n-a}\frac{(n-4)!}{(n-a-b)!}\binom{n}{2}\binom{n-2}{2}$$

$$= \frac{1}{4}\mathrm{U}^2 q \sum_{a=3}^{n-3}\sum_{b=3}^{n-a} \frac{n!(n-4)!}{4(n-a-b)!(n-4)!}$$
$$= \frac{1}{16}\mathrm{U}^2 q \sum_{a=3}^{n-3}\sum_{b=3}^{n-a} \frac{n!}{(n-a-b)!}.$$

By Flapan and Kozai [6], this is equal to

$$\frac{1}{16}\mathrm{U}^2 q \sum_{i=6}^{n} \frac{n!(i-5)}{(n-i)!}.$$

Flapan and Kozai [6] also showed that

$$\frac{1}{16} q \sum_{i=6}^{n} \frac{n!(i-5)}{(n-i)!}$$

has the order $\Theta(n(n!))$. By hypothesis, the factor of $\mathrm{U}^2$ is constant as $n$ increases, so the order is the same. That is, the order of the expected value of the sum of the squared linking numbers in a random piecewise-linear $\mathrm{K}_n$ is also $\Theta(n(n!))$.

□

It is worth comparing our generalization with the original result that inspired it. While Flapan and Kozai [6] showed that the expected value of the sum of squared linking numbers for the pairs of cycles found in URP generated linear embeddings of $\mathrm{K}_n$ had the order of $\Theta(n(n!))$, our result generalizes the embeddings to be piecewise-linear with a random variable per each edge dictating the number of line segments in the embedding for that edge. However, the stunning result is that this does not change the order of complexity.

## 5 Acknowledgments

This project was a result from the 2020 Rose-Hulman REU on spatial graph embeddings. We would like to acknowledge the National Science Foundation and Grant No. DMS-1852132 for funding for this REU. In particular, we would like to thank our REU supervisor and mentor Dr. Kenji Kozai.

## References


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

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

[3] M. Beals, L. Gross, and S. Harrell. DNA and Knot Theory. http://www.nimbios.org/~gross/LGrossTIEMwebsite/home/gross/public_html/bioed/webmodules/DNAknot.html, 1999. Online; accessed: 2022-08-07.

[4] John H. Conway and C McA. Gordon. Knots and links in spatial graphs. Journal of Graph Theory, 7(4):445–453, 1983.

[5] Yuanan Diao. The knotting of equilateral polygons in $\mathbb{R}^3$. Journal of Knot Theory and its Ramifications, 4(02):189–196, 1995.

[6] Erica Flapan and Kenji Kozai. Linking number and writhe in random linear embeddings of graphs. Journal of Mathematical Chemistry, 54(5):1117–1133, 2016.

[7] Nicholas Pippenger. Knots in random walks. Discrete Applied Mathematics, 25(3):273–278, 1989.

[8] D. W. Sumners and S. G. Whittington. Detecting knots in self-avoiding walks. Journal of Physics A: Mathematical and General, 23(8):1471, 1990.

[9] De Witt Sumners. Lifting the curtain: using topology to probe the hidden action of enzymes. Match, pages 51–76, 1996.

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



**Yasmin Aguillon**

University of Notre Dame

yaguillo@nd.edu

**Xingyu Cheng**

University of North Carolina at Chapel Hill xcheng1@unc.edu

**Spencer Eddins**

University of Kentucky

spencer.p.eddins@gmail.com

**Pedro Morales**

Purdue University

moralep@purdue.edu