Typical knots: size, link component count, and writhe

Margaret Doig
Mathematics Department, Creighton University

We model the typical behavior of knots and links using grid diagrams. Links are ubiquitous in the sciences, and their normal or typical behavior is of significant importance in understanding situations such as the topological state of DNA or the statistical mechanics of ring polymers. We examine three invariants: the expected size of a random knot; the expected number of components of a random link; and the expected writhe of a random knot. We investigate the first two numerically and produce generating functions which codify the observed patterns, we perform an exploratory data analysis for the third. We continue this project in a future work, where we investigate genus and the effects of crossing change on it.

I. KNOTS AND LINKS AS OUTCOMES IN A PROBABILITY SPACE

It is common (and enlightening) to study knots and links with extremal properties or to search for classifications and definite statements: Can I draw this particular knot with fewer crossings? How many knots have a diagram with 7 crossings? Is a minimal diagram for an alternating knot alternating? Can this invariant detect the unknot?

This is not our approach; rather, we pursue a complementary path, to understand the normal behavior of knots and links. We wish to investigate the family of links as a probability space and ask about the properties a typical or random link. There are many concrete topological reasons to study random links: If I doodle some 7-crossing knots, will there be interesting examples, or will I end up with a bunch of unknots? If I study this invariant for pretzel knots, will it be representative of the behavior for knots in general? Seifert genus basically grows with crossing number, but how fast? If I implement my algorithm for this invariant, will it usually finish in a reasonable amount of time? Similar questions about the typical behavior of knots also arise from a number of applications: cellular processes are affected by the knotting and writhing of DNA, the statistical mechanics of ring polymers seem to depend on their knot types, and even a new cryptographic scheme depends on whether the knot used as a key is sufficiently hard to approximate.

II. MOTIVATIONS

A. Knots and links as models for DNA and other polymers

Knots appear everywhere in nature: in polymers, DNA, networks, even headphone cords. We mathematically define a knot to be a closed curve in 3-D space, or, more precisely, an embedding of the circle $S^1$ into the 3-sphere $S^3$; a link is a set of component knots which may be entangled with one another in space. See Figure 1.

Knots and links model DNA reasonably well - even though DNA consists of two strands, they are complementary and twist into a single curve with rare effects for topology other than that they resist twisting (which we study in Section VI). Additionally, while eukaryotic DNA is technically not a closed loop as our definition of a knot requires, it is divided into large domains whose ends are fixed in the cellular structure away from the location of cellular processes and cannot be practically recruited to alter local topology; therefore, we model it with closed curves. Similarly, long linear polymers often locally look entangled in a way which meets our intuition of a knot, and we may reasonably model them by imagining the ends to be fixed in space some distance away from the action.

We have selected the grid diagram as our basic model for knots and links: we place arcs of variable length within a grid subject to some reasonable constraints, and we add crossings to resolve singularities. Under Even-Zohar’s classification of knot models, grid diagrams fall into the category of models with variable-length segments. Surprisingly, they do not appear to be limited by their right-angled or superficially 2-D nature. There are good biological and physical reasons to consider this model to be sufficiently representative of many situations in which random knots appear in nature. Some of our questions below have previously been addressed for other models, mostly from the category of constant-length segments, which seem to display distinct behavior. See Section II.5 for details.

B. Knots and polymer dynamics

The topology of polymers appears to influence their statistical mechanics. For example, the complexity of knotting of a long polymer chain can affect its response to weak mechanical stretching, its friction when passing through a constrained hole or pore, and its mobility when forced to pass through a gel or other resistant environment, as in electrophoresis. Knotting of a polymer ring can likewise affect the standard
size of the ring, the equilibrium relaxation time, its general motion, and the diffusion constant, among other topics.\textsuperscript{[13]} From the other direction, the formation of knotting and its complexity can be affected by degree of polymerization, quality of solvent, temperature, and confinement of the polymer.\textsuperscript{[15]}

Our first invariant, studied in Section \textsuperscript{[IV]} is motivated by the idea of building a random knot one segment at a time until it closes when the ends collide. We reproduce this by constraining ourselves within an \( n \times n \) grid and placing consecutive horizontal and vertical arcs of varying sizes (with some minor restrictions) until the knot closes, and we count the number of arcs used. We find that it is uniformly distributed among all possible values (all even numbers from 4 to \( 2n \)).

A number of theoretical and experimental works have previously examined scaling behavior of knots such as the relative dimensions of the knot compared to the number of arcs which make it up\textsuperscript{[16–25]} and this effort grows out of the investigation of random walks, both self-avoiding and \( n \times n \). The previous theoretical results have primarily used models from the category of constant-length segments, which seem to behave differently from grid diagrams.

Our second invariant, studied in Section \textsuperscript{[V]} is motivated by the idea of encountering a random link of a given size and counting how many components make up the link. We model this by selecting an \( n \times n \) grid diagram (see Section \textsuperscript{[II A]} and checking the number of components. As one may suspect, it is normally distributed, with mean and moments we derive from a generating function. Additionally, links of any given number of components become vanishingly rare as the link size grows. A similar question has previously been studied for random links in the braid and bridge presentation model,\textsuperscript{[26–28]} which fall into the same category as grid diagrams, but this question has surprisingly not been analyzed for most models, including grid diagrams.

C. DNA topology and writhe

Our third and final invariant in Section \textsuperscript{[VI]} writhe, is motivated by DNA topology. Writhe is the twisting or looping of a curve projected onto the plane. It reflects an effect most gardeners are familiar with: a carelessly coiled garden will resist being straightened out and will form kinks or supercoils. DNA molecules experience similar effects on a grander scale, and the resulting topology has a significant effect on cellular processes.

DNA topology is typically described by three values, the linking number (the total number of times the two strands link with each other), the twist (the total number of complete helical turns of a pair), and the writhe (which describes the greater topological structure, the number of times the helix crosses over itself, with sign). The linking number is invariant under local deformations of the DNA, while the twist and writhe are related to one another (under appropriate normalization, they sum to the linking number) and may be altered by cellular processes. The thermodynamically preferred relaxed B-DNA form has one full twist approximately every 10 base pairs, and torsional stress will tend to convert some of the over- or under-twisting into supercoiling of the molecule, or writhe. In nature, DNA is rarely found in its relaxed form and is on average underwound by about 6%\textsuperscript{[30]}

Genetic processes often alter the supercoiling or writhe of a molecule. Transcription involves the movement of RNA polymerase down the strand to unwind, transcribe, and rewind the twin strands. If unconstrained, the polymerase would follow the twist and proceed in a screw-like fashion down the helix. In the twin-domain supercoiling model of Liu and Wang,\textsuperscript{[31]} though, the polymerase is not unconstrained, and transcription results in positive coiling ahead and negative coiling behind. While these supercoils do not alter the global topology from a mathematical point of view, they will not locally resolve one other in nature. Replication also results in coiling: when the DNA forks for replication, positive supercoils form ahead of the fork, and a precatenane, or positive wending of the two partially replicated strands, forms behind the fork. If not removed after replication, the precatenane will convert to a catenane, a linkage between the resulting sister chromosomes.\textsuperscript{[32]}

This topological state of DNA significantly affects is availability for genetic processes. Negatively supercoiled molecules, as generally found in cells, favor reactions that require unwinding the helix such as transcription and replication, and positive supercoiling favors processes which require rewinding, such as the rescue of a stalled replication fork. The presence of a catenane itself is debilitating during cytokinesis. Topoisomerases generally police the topological state of the DNA, where they sever a strand, change a crossing, and reattach the strand, thereby removing a positive supercoiling after replication, eliminating a linkage between chromosomes, or otherwise maintaining the topological stability of the molecule.\textsuperscript{[33–35]} Other enzymes appear to act by a related set of moves, where they replace one tangle by another.\textsuperscript{[36]}

We model writhe by selecting a random knot in an \( n \times n \) grid diagram and counting its crossings, with sign. We compare the observed writhe to both the size of the diagram \( n \) and the actual length of the knot.

III. MATHEMATICAL PREREQUISITES

A. Grid diagrams as a model for random links

Numerous models are available for studying the typical or average knot. Following the naming convention of Even-Zohar\textsuperscript{[3]} the first set, which can be called the set of 1-D models, includes random walks in grids and random polygonal walks, both of which involve successively adding small steps of standardized size and which are inclined towards local knotting and easily generate knots and links which are connected sums. The next set, or the 2-D models, includes models which are in some sense less linear, like random planar diagrams or planar curves, or the knot tables themselves; these are still inclined towards non-prime knots (except for the knot tables, where they have been deliberately removed) and satellite knots (ex, double figure 8 configurations are common). The last set of the so-called 3-D models includes random jumps (whose difference from polygonal walks is that the length of each suc-
cessive edge varies), the Petaluma model, and the grid diagram. In each of these, the successive steps tend to be comparable to the knot diameter, and it is conjectured that they are overwhelming prime and hyperbolic and various invariants obey distributions which seem intuitively reasonable.

It is not clear which category DNA knots and other polymers fall into. If one imagines a polymer forming by the slow addition of successive base pairs or other small units to a longer chain, then an ideal model would be from the 1-D category, such as Brownian motion or lattice walks. On the other hand, 3-D category models show some promise for modeling knot formation by topoisomerase and other cellular actors which spontaneously change crossings or tangles or otherwise alter a preexisting knot or link.

We have selected the grid diagram as our model. A grid diagram of size \( n \) as in Figure 2 consists of an \( n \times n \) grid drawn on a torus with black dots and \( n \) white dots filled in so that each row and each column has a single black dot and a single white dot, in different cells. A grid diagram corresponds to a link: Insert an arc in each column from the black dot to the white dot; in each row, insert an arc from the white dot to the black dot. When two arcs intersect, replace the intersection by a crossing with the vertical arc on top. There are three Cromwell moves corresponding to the Reidemeister moves which change the appearance of the diagram but do not change the type of the underlying knot or link: commutations (where two arcs in successive rows or columns are exchanged); translations (where the arc in the top row is moved to the arc in the bottom or the left to the right); and stabilizations/destabilizations (there are four types). Figure 2 shows a trefoil and its corresponding grid diagram. See Cromwell for a thorough introduction to grid diagrams.

B. Combinatorial details

We have several choices to encode a grid diagram combinatorially. For example, we may record the \( (x, y) \) coordinates of each dot; for simplicity, we could record the \( y \) coordinates only, reading from the top row to the bottom, which would give two permutations of the numbers \( 1, 2, \ldots, n \) (equivalently, two \( n \)-cycles from \( S_n \)). This serves well for describing any known grid diagram, but it is challenging to generate a random diagram this way: two such permutations would generate a set of black and white dots, a pair for each row and a pair for each column, but we could have collisions between the black and white dots. In Sections V and VI we will use this method to generate random links: we will select two permutations of \( 1, 2, \ldots, n \) and eliminate any pair that induces collisions.

In the case of knots, we could select a starting point at some black dot on the knot and then trace through the knot, recording in parallel the order in which the knot visits the columns and the rows; in this case, we again generate two permutations \((\rho_1 \rho_2 \cdots \rho_n)\) and \((\kappa_1 \kappa_2 \cdots \kappa_n)\). Then the black dots would be located at the points with coordinates \((\rho_i, \kappa_i)\) as the white dots would be located at the \((\rho_{i+1}, \kappa_i)\). Observe that any knot corresponds to exactly \( n \) different permutations, depending on which black dot was selected as the starting point. We may easily randomly generate knots by selecting two permutations. We use this encoding or a slight variation for the knots in Sections V and VI.

C. Implementation

All calculations were performed within the author’s toolkit and are available to the public for further experimentation. Random permutations were generated using the default_random_engine in c++14, seeded with system clock time. Data analysis (including \( R^2 \)) was performed in JMP; knot and grid diagrams were generated in xfig, line histograms in Microsoft Excel, and other graphs in JMP.

IV. THE SIZE OF A TYPICAL KNOT

We warm up by studying the size of a knot randomly drawn within an \( n \times n \) grid. As a polymer might form and then close up in 3-D space, we envisage a knot building up arc by arc inside a grid diagram, but we allow the knot to close off at any time that the end touches the beginning. In particular, we do not require that the knot fill out the entire grid diagram. We will see that the knot size is uniformly distributed, which supports the use of grid diagrams to model polymeric behavior: if a polymer’s building blocks are of width 1 and variable length, and we wish to study a polymer constrained within an \( n \times n \) area, then we should model the polymer by looking at an \( n \times n \) grid diagram. As expected, the polymer is equally likely to close up after any given number of building blocks are added, within the constraint imposed by the area.

We model the question numerically, and then we produce a probability generating function to summarize behavior and verify the reasonableness of our model. One may argue that the combinatorics of this example are obvious; nevertheless, we produce the simulation to argue for the applicability of the grid diagram model to questions of polymer topology.

A. Numerical simulation

Recall the combinatorial encoding of a knot from Section III of an \( n \times n \) knot is represented by two \( n \)-permutations \( \rho = (\rho_1 \rho_2 \cdots \rho_n) \) and \( \kappa = (\kappa_1 \kappa_2 \cdots \kappa_n) \) which tell us the order in which we visit the rows and the columns. To generate
random knots living inside an \( n \times n \) grid, we assume without loss of generality that \( \rho_1 = 1 \) and then select two random \( n \)-permutations \( \rho \) and \( \kappa \). We read off \( k_1 \) from the front of \( \kappa \), then \( \rho_2 \) from \( \rho \), and so on; note that the \( k_i \)'s will not repeat, and the \( \rho_j \)'s will not repeat with the possible exception that some \( \rho_j \) may be \( 1 \), in which case we will declare that the knot closes up at size \( s \). (We omit the artificial case where \( \rho_2 = 1 \) and the knot is empty.) This is not exactly the combinatorial encoding for a grid diagram presented in Section III B, but it could be converted easily by removing the remaining \( n - s \) unused row/column numbers.

Randomly generating 10,000 knots in an \( n \times n \) grid for each of \( n = 10, 20, \cdots, 1000 \), we see that the size \( s \) of these knots is uniformly distributed from 2 to \( n \) with an average size of

\[
x(n) = \frac{n}{2} + 1.
\]

Below are the mean and standard error for the first ten cases, and a histogram appears in Figure (b). The data are comparable for larger values of \( n \).

| \( n \)  | 10  | 20  | 30  | 40  | 50  | 60  | 70  | 80  | 90  | 100 |
|---|---|---|---|---|---|---|---|---|---|---|
| \( \bar{s} \) | 6.0 | 11.1 | 16.0 | 21.2 | 26.1 | 31.0 | 36.1 | 41.3 | 46.1 | 50.6 |
| \( SE \) | 0.03 | 0.06 | 0.08 | 0.11 | 0.14 | 0.17 | 0.20 | 0.23 | 0.26 | 0.29 |

B. Mathematical analysis

Conveniently, we may directly calculate the expected size of a knot randomly drawn in an \( n \times n \) grid, and it agrees with our experimental estimate and expands our observations:

\[ p_s = \frac{1}{n-1}, \quad 1 < s \leq n. \]

The expected value of the knot size is \( EV(s) = \frac{n}{2} + 1 \), and the \( k^{th} \) moment is

\[ m_k(s) = \frac{1}{n-1} \sum_{t=2}^{n} t^k. \]

**Proof.** We see immediately that the probability of a 1 appearing in the \( s^{th} \) position in a random permutation of 1, 2, \cdots, \( n \) is exactly \( \frac{1}{n} \); omitting the cases where 1 appears in the first term, we have \( p_s = \frac{1}{n-1} \).

Observe that \( \sum p_s = 1 \) as it should be, and the expected value of the knot size is therefore

\[ EV(s) = \sum_{s=1}^{n} s p_s = \frac{1}{n-1} \sum_{s=2}^{n} s = \frac{n}{2} + 1. \]

We observe also that we could establish an (ordinary) probability generating function for \( p_s \); that is, we can build a function \( P(t) = \sum_{s=0}^{\infty} p_s t^s \) which neatly encapsulates all the \( p_s \) within a single series. We can recover the \( p_s \) either from the series expansion of \( p_s \) or by noting \( \frac{d^n p(t)}{dt^n}(0) = s! p_s \). In this case:

\[ P(t) = \frac{1}{n-1} \sum_{s=2}^{n} t^s = \frac{t^2 + t^3 + \cdots + t^n}{n-1} = \frac{t^2(1-t^{n-1})}{(n-1)(1-t)}. \]

Recall that an exponential generating function \( g(x) \) for a sequence \( a_k \) is a series expansion \( g(x) = \sum_{k=0}^{\infty} a_k x^k \); it is an elementary fact that, if \( P(t) \) is a probability generating function for a variable, then \( P(e^t) \) is an exponential generating function.
for its moments $m_k$,
\[
P(e) = e^{2t} + e^{3t} + \cdots + e^{nt} = \frac{1}{n-1} \left( \sum_{k=0}^{n} \binom{n}{k}^2 + \sum_{k=0}^{n} \binom{n}{k}^3 + \cdots + \sum_{k=0}^{n} n^k \binom{n}{k} \right)
\]
so
\[
m_k(s) = \frac{2^k + 3^k + \cdots + n^k}{n-1}.
\]
This formula agrees with our direct calculation of $m_1(s)$, a.k.a, $EV(s)$.

\section{The Number of Components of a Typical Link}

We next consider random links. These model the behavior of self-assembling polymers, which are not guaranteed to form a single chain when left to their own devices. We first consider the number of components $\#_e$ of such a link in an $n \times n$ grid diagram.

We use the second combinatorial encoding of Section III B of generating two $n$-permutations to give the $y$-coordinates for the black and white dots and removing any impossible grid diagrams (i.e., pairs of permutations with collisions). We count the number of valid link diagrams $c_n$ and the number with $k$ components $c_{n,k}$. There are naively $n!n!$ ways to place the black and white dots in a grid without accounting for collisions, and we will also make use of a renormalization, $\bar{c}_n = \frac{c_n}{n!n!}$ and $\bar{c}_{n,k} = \frac{c_{n,k}}{n!n!}$.

\subsection{Numerical simulation}

To explore the number of components $k$, we generate all possible grid diagrams up to $n = 7$, where computation becomes prohibitive. We count the number of knots and 2- and 3-component links ($c_{n,1}$, $c_{n,2}$, and $c_{n,3}$, respectively), and we calculate the average number of components $\bar{k}$:

\begin{table}[h]
\begin{tabular}{|c|c|c|c|c|c|}
\hline
$n$ & $c_{n,1}$ & $c_{n,2}$ & $c_{n,3}$ & $c_n$ & $\bar{k}$ \\
\hline
2 & 2 & & & 2 & 1 \\
3 & 12 & & & 12 & 1 \\
4 & 144 & 72 & & 216 & 1.3333 \\
5 & 2,880 & 2,400 & & 5,280 & 1.4545 \\
6 & 86,400 & 93,600 & 10,800 & 190,800 & 1.6038 \\
7 & 3,628,800 & 4,656,960 & 1,058,400 & 9,344,160 & 1.7249 \\
\hline
\end{tabular}
\end{table}

To explore component size for larger $n$, we generate 10,000 grid diagrams for each size $n = 10, 20, \ldots, 1000$ and count the number of components of each. Each grid diagram came from two random permutations (generated by the default_random_engine in c++14 seeded with clock time) of the numbers 1 through $n$, after eliminating ones which generated collisions between the black and white dots. The number of components follows a normal distribution, as seen in Figure 4 with mean and standard deviation:
\[
\bar{k}(n) = 1.1334\ln n + 1.3358,
\]
\[
SD(n) = 0.3356\ln n + 0.7129,
\]
where $R^2 = 0.9792$ for $\bar{k}$ and 0.9259 for $SD$. Loosely speaking, then, for large $n$, most random grid diagrams have between $0.5\ln n$ and $2\ln n$ components; 95% of the obey the inequality:
\[
0.4622\ln n - 0.0900 \leq k \leq 1.8046\ln n + 2.7616.
\]

Additionally, we observe that, as the grid size increases,
knots become vanishingly rare; in fact, for any fixed $k$, the $k$-component links become vanishingly rare. If $\Pr(k \mid n)$ is the probability that an $n \times n$ grid diagram represents a $k$-component link, then

$$\Pr(k \mid n) \to 0 \text{ as } n \to \infty$$

B. Mathematical analysis

We may also calculate directly how many $n \times n$ grid diagrams have a given number of components.

**Theorem 2.** For any $n$, the total number of $n \times n$ grid diagrams for knots is:

$$c_{n,1} = n!(n-1)!$$

The total number of link diagrams is:

$$c_n = n! n! \sum_{i=0}^{n} (-1)^i \frac{1}{i!}$$

and, in fact, $c_n$ displays asymptotic behavior similar to $n! n!$

$$\lim_{n \to \infty} \frac{c_n}{n! n!} = \frac{1}{e}$$

Equivalently, we could say $\tau_{n,1} = \frac{1}{n}$ and $\tau_n = \sum_{i=0}^{n} (-1)^i \frac{1}{i!}$.

**Proof.** To see $c_{n,1}$, recall the combinatorial description of a knot used in Section [IV] we select one permutation $\rho$ and one permutation $\kappa$; there are $n! n!$ pairs of such permutations. Each grid diagram corresponds to such a pair, or rather, to a permutation of these pairs, depending on which of the $n$ different black dots we select as our starting point; that is, there are $n!(n-1)!$ knot diagrams.

For the other results, we will use a generating function for $\tau_n$, that is, some function $g(x) = \sum_{n=0}^{\infty} \tau_n x^n$. Then we may recover the value by setting $g(x) = \frac{1}{n!} \frac{\partial^n}{\partial x^n}(0)$. We derive generating functions for the appropriate invariants in Theorem 2 of Section [VII]. Therefore,

$$\tau_n = \frac{1}{n!} \frac{\partial^n}{\partial x^n}\left[ \frac{1}{1-x} e^{-x} \right]_{x=0}$$

$$= \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} \frac{\partial^{n-i}}{\partial x^{n-i}}\left[ \frac{1}{1-x} \right]_{x=0} \frac{\partial^{i}}{\partial x^{i}}[e^{-x}]_{x=0}$$

$$= \frac{1}{n!} \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \left[ \frac{(n-i)!}{(1-x)^{n-i+1}} \right]_{x=0} \left[ (-1)^i e^{-x} \right]_{x=0}$$

$$= \sum_{i=0}^{n} (-1)^i \frac{1}{i!}$$

The asymptotic behavior follows immediately:

$$\lim_{n \to \infty} \tau_n = \sum_{i=0}^{\infty} (-1)^i \frac{1}{i!} = e^{-1}$$

It is more difficult to calculate the number of $k$-component links, although we may bound it and verify that they appear ever more rarely as the diagram enlarges.

**Theorem 3.** For any $n$ and $k$, the number of $n \times n$ grid diagrams representing $k$-component links obeys:

$$c_{n,k} \leq n!(n-1)! (\log n)^{k-1}$$

and $k$-component links become vanishingly rare as the grid size $n$ grows.

$$\Pr(k \mid n) \to 0 \text{ as } n \to \infty.$$ 

Indeed, any given link type $[L]$ is vanishingly rare as $n$ grows.

$$\Pr([L] \mid n) \to 0 \text{ as } n \to \infty.$$ 

This is a variation on the traditional conjecture that a given knot type grows vanishingly rare among knots as $n \to \infty$.

**Proof.** Assume $n > 1$. We will actually verify the bound

$$\tau_{n,k} \leq \frac{(\log n)^{k-1}}{n}.$$ 

This bound is sharp for $k = 1$ since $\tau_{n,1} = \frac{1}{n}$ by Theorem 2.

For larger $k$, we use the generating function $h(x)$ from Theorem 5.

$$\tau_{n,k} = \frac{1}{n!} \frac{\partial^n}{\partial x^n}\left[ \frac{1}{k!} (-\ln(1-x) - x)^k \right]_{x=0}$$

$$= \frac{1}{n!} \frac{\partial^{n-1}}{\partial x^{n-1}}\left[ \frac{1}{(k-1)!} (-\ln(1-x) - x)^{k-1} \left( \frac{1}{1-x} - 1 \right) \right]_{x=0}$$

$$= \frac{1}{n!} \sum_{i=0}^{n-1} \binom{n-1}{i} \frac{\partial^{i}}{\partial x^{i}}\left[ (-\ln(1-x) - x)^{k-1} \right]_{x=0} \frac{\partial^{n-i-1}}{\partial x^{n-i-1}}\left[ \frac{1}{1-x} \right]_{x=0}$$

$$= \frac{1}{n!} \sum_{i=0}^{n-2} \frac{(n-1)!}{(n-i-1)!} \frac{1}{i!} \tau_{i,k-1}(n-i-1)! = \frac{1}{n} \sum_{i=1}^{n-2} \tau_{i,k-1}$$

which, inducting on $k$, gives

$$\tau_{n,k} \leq \frac{1}{n} \sum_{i=1}^{n-2} \frac{(\log i)^{k-2}}{i} \leq \frac{1}{n} (\log n)^{k-2} \sum_{i=1}^{n-1} \frac{1}{i} \leq \frac{1}{n} (\log n)^{k-1}$$

where the last inequality comes from the classical harmonic bound $H_{n-1} \leq \log n$.

Finally, observe that

$$\lim_{n \to \infty} \frac{(\log n)^{k-1}}{n} = 0$$

by a repeated application of l’Hôpital’s rule, so $\tau_{n,k} \to 0$ as
well. Since \( \tau_n \to \frac{1}{c} \) by Theorem 2:

\[
\lim_{n \to \infty} \frac{c_{n,k}}{c_n} = \lim_{n \to \infty} \frac{\tau_{n,k}}{\tau_n} = 0.
\]

In other words, \( \Pr(k \mid n) \to 0 \). The result for a fixed link type follows trivially: if all links of \( k \) components become vanishingly rare, then the particular ones which represent a given link type become even rarer.

At last, we verify our observation of mean and variance.

**Theorem 4.** In an \( n \times n \) grid, the expected value of the number of components \( k \) is

\[
EV(k) = \frac{1}{c_n} \sum_{i=1}^{n-1} (-1)^i \frac{H_{n-i} - 1}{i!}
\]

Here the \( H_i = \sum_{i=1}^{n} \frac{1}{i} \) are the Harmonic numbers.

**Proof.** To study expected value, note that (after fixing \( n \)) the probability of a randomly selected link having exactly \( k \) components is

\[
p_k = \frac{c_{n,k}}{c_n} = \frac{\tau_{n,k}}{\tau_n}
\]

which has probability generating function

\[
P(y) = \sum_{i=0}^{\infty} p_i y^i = \sum_{i=0}^{\infty} \frac{\tau_{n,i}}{\tau_n} y^i = \frac{1}{\tau_n} f(y),
\]

where \( f(y) \) is the generating function for \( \tau_{n,k} \) (with \( n \) fixed) from Theorem 5. The expected value will be

\[
EV = P'(1) = \frac{1}{\tau_n} \frac{\partial f}{\partial y}(1)
\]

and the variance will be

\[
Var = P''(1) - P'(1)
\]

Before we proceed, note that, if \( i < n \),

\[
\frac{\partial^2}{\partial y^2} \left[ \binom{y+n-i-1}{y-1} \right]_{y=1} = \frac{1}{(n-i)!} \frac{n-i-1}{\prod_{j=0}^{i-1} (y+j)} \sum_{k=0}^{n-i-1} \frac{1}{y+k}
\]

In particular, for \( 0 \leq i \leq n \) (if we assume the trivial case \( H_0 = 0 \)),

\[
\frac{\partial}{\partial y} \left[ \binom{y+n-i-1}{y-1} \right]_{y=1} = H_{n-i}
\]

and

\[
\frac{\partial^2}{\partial y^2} \left[ \binom{y+n-i-1}{y-1} \right]_{y=1} = \frac{\partial}{\partial y} \left[ \binom{y+n-i-1}{y-1} \sum_{k=0}^{n-i-1} \frac{1}{y+k} \right]_{y=1}
\]

\[
= \binom{y+n-i-1}{y-1} \left( \sum_{k=0}^{n-i-1} \frac{1}{y+k} \right)_{y=1}
\]

Therefore,

\[
EV(k) = \frac{1}{\tau_n} \sum_{i=0}^{n} \frac{\partial}{\partial y} \left[ (-1)^i \binom{y+n-i-1}{y-1} \frac{y^i}{i!} \right]_{y=1}
\]

and the variance obeys

\[
Var(k) + EV(k) = \frac{1}{\tau_n} \sum_{i=0}^{n} \frac{\partial^2}{\partial y^2} \left[ (-1)^i \binom{y+n-i-1}{y-1} \frac{y^i}{i!} \right]_{y=1}
\]

so

\[
Var(k) = \frac{1}{\tau_n} \sum_{i=0}^{n} (-1)^i \frac{H_{n-i}^2 - H_{n-i-2} + 2iH_{n-i} + i(i-1)}{i!}
\]

Theorem 3 is reminiscent of Ma, who studied the random braid model, where a link \( w_{n,k} \) is formed by closing up a random walk of \( k \) steps on the braid group \( B_n \). Ma found that, if \( n \) is fixed and \( k \) allowed to pass to infinity, then the expected value of the number of components approaches \( H_n \). Just as in our model, the EV grows with \( n \), although he does not derive separate results for the prevalence of a given number of components. He produces similar results for the bridge model, where a link \( w_{2,n,k} \) comes from closing up a random walk of \( k \) steps in the mapping class group \( \mathcal{M}_{0,2n} \) via an \( n \)-bridge presentation.

**VI. WRITHE \textit{wr}**

We finally approach writhe, which is algebraic count of the crossings (i.e., the number of positive crossings minus the number of negative crossings - see below).
The writhe \( w_r \) of a knot projected into the plane is dependent upon its projection, or its diagram; however, if the knot is framed or is presented as a pair of adjacent strands (as in the case of DNA), then a change to writhe is reflected in a change of framing or of the twist of the pair of strands.

### A. Numerical simulation

We check the writhe of a random \( n \times n \) knot diagram. Writhe should average 0 and be symmetric around the mean (every diagram has an equally likely mirror diagram where all crossings are changed; for grid diagrams, reflecting right-to-left is equivalent to changing all crossings).

We consider the second description from Section III B: a knot is two \( n \)-permutations which read off the order in which the columns and rows and encountered. See Figure 5 for knots up to \( n = 1000 \).

First, we ran 10,000 cases for each of \( n = 10, 20, \cdots, 1000 \) and compared writhe to grid size. The mean writhe \( \overline{w_r}(n) \) for an \( n \times n \) grid diagram may be fit to its average:

\[
\overline{w_r}(n) = -0.187818
\]

with standard error of only 0.126266. In fact, 93 of the 100 cases have a 95\% confidence interval for \( \overline{w_r} \) which contains 0. The standard deviation may be modeled by:

\[
SD(n) = 0.2358955n - 0.215688
\]

with \( p < 0.0001 \). We conjecture that the population obeys:

\[
SD(n) = 0.25n - 0.25.
\]

We also calculate the writhe-to-length ratio. We ran 1,000,000 cases randomly selected from \( 2 \leq n \leq 1000 \). The longest knot had length just over 700,000, and knots on the upper end of this range were not well represented; we restricted our set to those of length at most 650,000 and rounded length to the nearest 10,000. Once more, we fit \( \overline{w_r} \) to its average:

\[
\overline{w_r}(l) = 0.486494
\]

with standard error of only 0.174848 and model standard deviation as:

\[
SD(l) = 0.2854956 l^{1/2} + 1.5235323.
\]

with \( p < 0.0001 \).

### B. Mathematical analysis

The behavior of writhe suggests it is susceptible to analysis via generating functions, but it was resistant to our efforts. We leave this project for future work.

---

FIG. 5: Writhe vs grid size for randomly generated \( n \times n \) knot grid diagrams: (a) a box plot for \( n \leq 1000 \); (b) a histogram for \( n \leq 100 \); (c) standard deviation of writhe vs. grid size.
VII. GENERATING FUNCTIONS

We derive several results used in Section V.

Lemma 5. $\tau_n$ has generating function

$$g(x) = (1 - x)^{-1} e^{-x}$$

and $\tau_{n,k}$ has generating function

$$G(x, y) = (1 - x)^{-y} e^{-xy}.$$  

If we fix $k$ and vary $n$, then $G(x, y)$ simplifies to

$$h(x) = (-1)^k \left( \frac{\ln(1-x) + x}{k!} \right)^k.$$  

On the other hand, if we fix $n$ and vary $k$, it simplifies to

$$f(y) = \sum_{i=0}^{n} (-1)^i \left( \frac{y + n - i - 1}{y - 1} \right)^i i!.$$  

Proof. We begin by deriving formulae for $c_n$ and $c_{n,k}$. Consider an arbitrary link in an $n \times n$ grid. Let $e_i$ be the number of components that use $i$ rows and columns. We assume $e_1 = 0$ since this is a grid diagram, and $2e_2 + 3e_3 + \cdots = n$. To count the number of links which satisfy a given choice of component sizes $e_2, e_3, e_4, \cdots$, we must first divide the $n$ rows among the components: that is, we must divide them into subsets with $e_2$ of size 2, $e_3$ of size 3, and so on. There are $\frac{n!}{(2!)^{e_2}(3!)^{e_3}\cdots}$ ways to do this. Repeat for the rows. Next, since this was not a proper partition (we can distinguish between two subsets of the same size), we now divide by $e_2! e_3! \cdots$. Finally, for each subset

$$c_n = \sum_{2e_2 + 3e_3 + \cdots = n} \frac{n!n!(2!)^{e_2}(3!)^{e_3} \cdots}{(2!)^{e_2}(3!)^{e_3} \cdots} \cdot \frac{1}{e_2! e_3! \cdots}.$$  

To find the number of links with exactly $k$ components $c_{n,k}$, we need merely add the condition that $e_2 + e_3 + \cdots = k$:

$$c_{n,k} = \sum_{2e_2 + 3e_3 + \cdots = n} \frac{n!n!}{2^{e_2} 3^{e_3} \cdots} \cdot \frac{1}{e_2! e_3! \cdots}.$$  

These formulas are not particularly easy to manipulate on their own, so we employ generating functions to make them manageable. Herbert Wilf wrote, “A generating function is a clothesline on which we hang a sequence of numbers for display.” Recall that an ordinary generating function for $\tau_n$ is some $g(x)$ with MacLaurin series $g(x) = \sum_{n=0}^{\infty} \tau_n x^n$, and so $\tau_n = \frac{1}{n!} \frac{\partial^n g}{\partial x^n}(0)$. See, e.g., Tucker (Ch. 6) or Wilf (for $\tau_n$).
above, consider the function
\[
g(x) = \left( 1 + \frac{x^2}{2} + \frac{\left( \frac{x^2}{2} \right)^2}{2!} + \cdots \right) \left( 1 + \frac{x^3}{3} + \frac{\left( \frac{x^3}{3} \right)^2}{2!} + \cdots \right) \ldots
\]
We may think of \( e_2 \) as telling us which term to select from the first parentheses, then \( e_3 \) which term to select from the next, and so on. The power series expansion has as its degree \( n \) term the sum of all formal products
\[
\left( \frac{x^2}{2} \right)^{e_2} \left( \frac{x^3}{3} \right)^{e_3} \cdots \frac{x^{2e_2+3e_3+\cdots}}{2e_2+3e_3} \cdots = \frac{x^n}{2e_2+3e_3} \cdots
\]
because \( 2e_2 + 3e_3 + \cdots = n \); this sum happens to be the same as \( n! e_n x^n = \tau_n x^n \). That is,
\[
g(x) = \sum_{n=0}^{\infty} \tau_n x^n.
\]
The expansion of \( g(x) \) above is a little awkward; we may rewrite it as
\[
g(x) = e^{x^2} e^{x^3} \cdots = e^{x^2 + x^3 + \cdots}.
\]
Additionally,
\[
\frac{\partial}{\partial x} \left[ \frac{x^2}{2} + \frac{x^3}{3} + \cdots \right] = x + x^2 + \cdots
\]
\[
= \frac{1}{1-x} - 1 = \frac{\partial}{\partial x} \left[ -\ln |1-x|-x \right].
\]
We restrict ourselves to the case of \( x < 1 \), and we perform a quick check at \( x = 0 \) to verify that \( \frac{x^2}{2} + \frac{x^3}{3} + \cdots = -\ln |1-x|-x \), so
\[
g(x) = e^{-\ln |1-x|-x} = (1-x)^{-1} e^{-x}.
\]
We next need a multivariable generating function for \( \tau_{n,k} \).
Let
\[
G(x,y) = \left( 1 + \frac{x^2 y}{2} + \frac{\left( \frac{x^2 y}{2} \right)^2}{2!} + \cdots \right) \left( 1 + \frac{x^3 y}{3} + \frac{\left( \frac{x^3 y}{3} \right)^2}{2!} + \cdots \right) \ldots
\]
which has as its degree \( (n,k) \) term the sum of formal products
\[
\left( \frac{1}{2} x^2 y \right)^{e_2} \left( \frac{1}{3} x^3 y \right)^{e_3} \cdots = \frac{x^{2e_2+3e_3} y^k}{2e_2+3e_3} \cdots
\]
for \( 2e_2 + 3e_3 + \cdots = n \) and \( e_2 + e_3 \cdots = k \). This sum is again exactly \( \tau_{n,k} x^n y^k \), so \( G(x,y) \) is a generating function for \( \tau_{n,k} \), or
\[
G(x,y) = \sum_{n,k=0}^{\infty} \tau_{n,k} x^n y^k.
\]
We may also write it:
\[
G(x,y) = e^{\left( \frac{x^2 y}{2} + \frac{x^3 y}{3} + \cdots \right)} = (g(x))^y = (1-x)^{-y} e^{-y},
\]
and of course \( g(x) = G(x,1) \).
We observe that we could fix \( n \) and find a generating function for \( \tau_{n,k} \) as \( k \) varies,
\[
f(y) = \sum_{k=0}^{\infty} \tau_{n,k} y^k = \frac{1}{n!} \frac{\partial^n G}{\partial x^n} (0,y).
\]
In other words,
\[
f(y) = \frac{1}{n!} \frac{\partial^n}{\partial x^n} \left[ (1-x)^{-y} e^{-y} \right]_{x=0}
\]
\[
= \frac{1}{n!} \sum_{i=0}^{n} \binom{n}{i} \left[ \frac{\partial^{n-i}}{\partial x^{n-i}} \left[ (1-x)^{-y} \right] \right]_{x=0}
\]
\[
= \frac{1}{n!} \sum_{i=0}^{n} \frac{n!}{i!(n-i)!} \left[ (1-y)^i e^{-y} \right]_{x=0} \left[ \frac{(-y)^i}{(y-1)!} \right]_{x=0}
\]
\[
= \sum_{i=0}^{n} \binom{n}{i} \frac{(-y)^i}{(y-1)!} \left[ \frac{(y+n-i-1)!}{i!(y-1)!} \right]_{x=0}.
\]
On the other hand, we could also fix \( k \) and find a generating function for \( \tau_{n,k} \) as \( n \) varies. This function should take the form:
\[
h(x) = \sum_{k=0}^{\infty} \frac{\partial^k G}{\partial y^k} \left( x,0 \right)
\]
\[
= \frac{1}{k!} \left[ \left( (1-x)^{-1} e^{-x} \right)^y \right]_{y=0}
\]
\[
= \frac{1}{k!} \left[ \left( (1-x)^{-1} e^{-x} \right)^y \left( \ln \left( (1-x)^{-1} e^{-x} \right) \right) \right]_{y=0}
\]
\[
= \frac{(-\ln(1-x)-x)^k}{k!} = (-1)^k \left( \ln(1-x)+x \right)^k.
\]
\[]

VIII. FUTURE WORK

The exploratory data analysis on write above begs to be verified mathematically.

We are also involved in a future exploration of genus as a model of knot complexity; in particular, we investigate the effects of a crossing change and tangle change on genus.

IX. ACKNOWLEDGEMENTS

Thanks to Creighton University for giving me an environment in which I am free to wander about the mathematical world and explore new things. Thanks also to my student
Billy Duckworth, whose ideas during his summer research inspired me to start digging into the world of typical knots. Finally, thanks to the many previous researchers referenced here whose work was inspiring and interesting as well as highly informative.

1. M. I. Doig, “Typical links: The effects of crossing change on genus,” In progress.
2. E. Farhi, D. Gosset, A. Hassidim, A. Lutomirski, and P. Shor, “Quantum money from knots,” in *Proceedings of the 3rd Innovations in Theoretical Computer Science Conference* (2012) pp. 276–289.
3. C. Even-Zohar, “Models of random knots,” Journal of Applied and Computational Topology 1, 263–296 (2017).
4. M. Caraglio, C. Micheletti, and E. Orlandini, “Stretching response of knotted and unknotted polymer chains,” Physical review letters 115, 188301 (2015).
5. C. Plesa, D. Verschuere, S. Pud, J. van der Torre, J. W. Ruitenberg, M. J. Witteveen, M. P. Jonsson, A. Y. Grosberg, Y. Rabin, and C. Dekker, “Direct observation of dna knots using a solid-state nanopore,” Nature nanotechnology 11, 1093 (2016).
6. J. Arsuaga, M. Vázquez, S. Trigueros, J. Roca, et al., “Knotting probability of dna molecules confined in restricted volumes: Dna knotting in phage capsids,” Proceedings of the National Academy of Sciences 99, 5373–5377 (2002).
7. J. Cebrián, M. J. Kadomatsu-Hermosa, A. Castán, V. Martínez, C. Parra, M. J. Fernández-Nestosa, C. Schaefer, M.-L. Martínez-Robles, P. Hernández, D. B. Krimer, et al., “Electrophoretic mobility of supercoiled, cated- nated and knotted dna molecules,” Nucleic acids research 43, e24–e24 (2015).
8. A. Stasiak, V. Katritch, J. Bednar, D. Michoud, and J. Dubochet, “Electrophoretic mobility of dna knots,” Nature 384, 122–122 (1996).
9. L. Olavarrieta, M. L. Martínez-Robles, J. Sogo, A. Stasiak, P. Hernandez, D. B. Krimer, and J. B. Schwartzman, “Supercoiling, knotting and replication fork reversal in partially replicated plasmids,” Nucleic acids research 30, 656–666 (2002).
10. S. Trigueros, J. Arsuaga, M. E. Vázquez, D. W. Summers, and J. Roca, “Novel display of knotted dna molecules by two-dimensional gel electrophoresis,” Nucleic acids research 29, e67–e67 (2001).
11. S. Trigueros and J. Roca, “Production of highly knotted dna by means of cosmid circularization inside phage capsids,” BMC biotechnology 7, 94 (2007).
12. C. Weber, A. Stasiak, P. De Los Rios, and G. Dietler, “Numerical simulation of gel electrophoresis of dna knots in weak and strong electric fields,” Biophysical journal 90, 3100–3105 (2006).
13. C. Weber, P. De Los Rios, G. Dietler, and A. Stasiak, “Simulations of electrophoretic collisions of dna knots with gel obstacles,” Journal of Physics: Condensed Matter 18, S161 (2006).
14. E. Orlandini, “Statics and dynamics of dna knotting,” Journal of Physics A: Mathematical and Theoretical 51, 053001 (2017).
15. E. Orlandini and S. G. Whittington, “Statistical topology of closed curves: Some applications in polymer physics,” Reviews of modern physics 79, 611 (2007).
16. J. des Cloizeaux, “Ring polymers in solution: topological effects,” Journal de Physique Lettres 42, 433–436 (1981).
17. E. Orlandini, M. C. Tesi, E. J. Van Rensburg, and S. G. Whittington, “Asymptotics of knotted lattice polygons,” Journal of Physics A: Mathematical and General 31, 5953 (1998).
18. J. Deutsch, “Equilibrium size of large ring molecules,” Physical Review E 59, R2539 (1999).
19. A. Y. Grosberg, “Critical exponents for random knots,” Physical review letters 85, 3858 (2000).
20. M. K. Shitamura and T. Deguchi, “Anomalous finite-size effects for the mean-squared gyration radius of gaussian random knots,” Journal of Physics A: Mathematical and General 35, L241 (2002).
21. A. Dobay, J. Dubochet, K. Millett, P.-E. Sottas, and A. Stasiak, “Scaling behavior of random knots,” Proceedings of the National Academy of Sciences 100, 5611–5615 (2003).
22. H. Matsuda, A. Yao, H. Tsukahara, T. Deguchi, K. Furuta, and T. Inami, “Average size of random polygons with fixed knot topology,” Physical Review E 68, 011102 (2003).
23. N. T. Moore, R. C. Lua, and A. Yu. Grosberg, “Under-knotted and over-knotted polymers 1: Unrestricted loops,” in *Physical and Numerical Models in Knot Theory: Including Applications to the Life Sciences* (World Scientific, 2005) pp. 363–384.
24. R. C. Lua, N. T. Moore, and A. Yu. Grosberg, “Under-knotted and over-knotted polymers 2: Compact self-avoiding loops,” in *Physical and Numerical Models in Knot Theory: Including Applications to the Life Sciences* (World Scientific, 2005) pp. 385–398.
25. E. Uehara and T. Deguchi, “Scaling behavior of knotted random polygons and self-avoiding polygons: Topological swelling with enhanced exponent,” The Journal of chemical physics 147, 214901 (2017).
26. P. J. Flory, *Principles of polymer chemistry* (Cornell University Press, 1953).
27. P.-G. De Gennes and P.-G. Gennes, *Scaling concepts in polymer physics* (Cornell university press, 1979).
28. M. Doi and S. F. Edwards, *The theory of polymer dynamics*, Vol. 73 (oxford university press, 1988).
29. J. Ma, “Components of random links,” Journal of Knot Theory and Its Ramifications 22, 1350043 (2013).
30. A. V. Vologodskii and N. R. Cozzarelli, “Conformational and thermodynamic properties of supercoiled DNA,” Annual review of biophysics and biomolecular structure 23, 609–643 (1994).
31. I. F. Liu and J. C. Wang, “Supercoiling of the DNA template during transcription,” Proceedings of the National Academy of Sciences 84, 7024–7027 (1987).
32. J. I. Champoux and M. D. Been, “Topoisomerases and the swivel problem,” in *Mechanistic studies of DNA replication and genetic recombination* (Elsevier, 1980) pp. 809–815.
33. J. E. Dewese, M. A. Osheroff, and N. Osheroff, “DNA topology and topoisomerases: teaching a “knotty” subject,” Biochemistry and Molecular Biology Education 37, 2–10 (2009).
34. Q. Espeli and K. J. Marians, “Untangling intracellular DNA topology,” Molecular microbiology 52, 925–931 (2004).
35. Y. Seol and K. C. Neuman, “The dynamic interplay between DNA topoisomerases and DNA topology,” Biophysical reviews 8, 101–111 (2016).
36. I. K. Darcy and M. Vázquez, “Determining the topology of stable protein–DNA complexes,” (2013).
37. A. Barbensi, D. Celoria, H. A. Harrington, A. Stasiak, and D. Buck, “Grid diagrams as tools to investigate knot spaces and topoisomerase-mediated simplification of dna topology,” Science Advances 6, eaay14493 (2020).
38. P. Cromwell, “Arc presentations of knots and links,” Banach Center Publications 42, 57–64 (1998).
39. M. I. Doig, “Doig Math Toolkit,” http://doigmath.maderak.com (2019).
40. A. Tucker, *Applied combinatorics* (John Wiley & Sons, Inc., 1994).
41. H. S. Wilf, *generatingfunctionology* (AK Peters/CRC Press, 2005).