Deterministic Model(s) for Topoisomerase II

Eric Lybrand
with Jason Cantarella and Harrison Chapman

February 13, 2017

Introduction

Fact: your body is composed of 37.2 trillion cells.

- Of those 37.2 trillion, roughly 30 billion go through mitosis each day.
- Given that DNA is really really tangled, how does a mother cell guarantee that each sister cell gets the right amount of DNA?

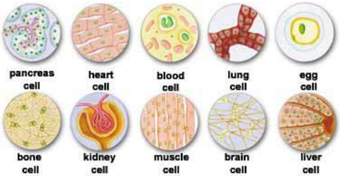


Figure: Cells in your body.

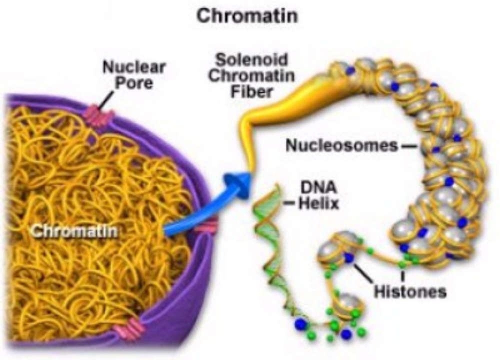


Figure: DNA in a nucleus.

Introduction

Rybenkov et. al. in 1997 found that Topoisomerase II, an enzyme active during mitosis, magically reduced knotting fraction of DNA in steady state by 80 times compared to thermodynamic equilibrium!

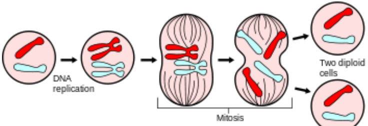


Figure: Mitosis!

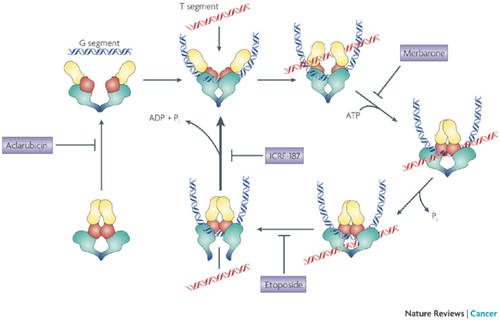


Figure: Topoisomerase II in action!

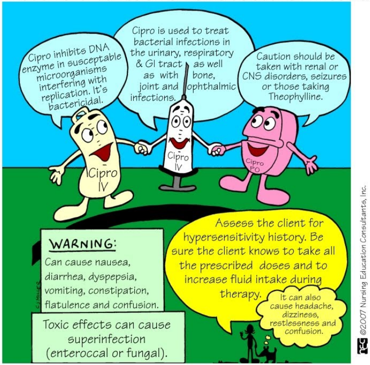
Introduction

Applications

- Cancer Prevention
- Antibiotics (Cipro)
- Knot Theory!

Ciprofloxacin (Cipro)

"For those hard to reach chronic bacterial infections"



Introduction

Goals for the talk:

- Brief overview of Knot Theory
- Propose model(s) for Topo II
- Look at performance of model(s).

1. Crash Course on Knot Theory
2. Topological Model for Topo II
3. Numerical Experiments

1. Crash Course on Knot Theory

2. Topological Model for Topo II

3. Numerical Experiments

1. What is a Knot?

Definition

A map between topological spaces $f : X \to Y$ is called an **embedding** if it is smooth, injective, and has an everywhere injective derivative.

1. What is a Knot?

Definition

A map between topological spaces $f : X \to Y$ is called an **embedding** if it is smooth, injective, and has an everywhere injective derivative.

Definition

A **knot** is an embedding $f : S^1 \to R^3$.

1. What is a Knot?

Definition

A map between topological spaces $f : X \to Y$ is called an **embedding** if it is smooth, injective, and has an everywhere injective derivative.

Definition

A **knot** is an embedding $f : S^1 \to R^3$.

Definition

A **link** is a collection of non-intersecting knots.

Examples:

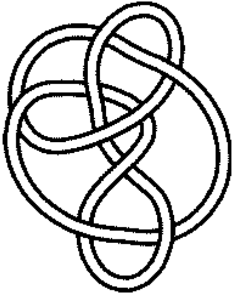


Figure: Knot 9n44 in Rolfsen table

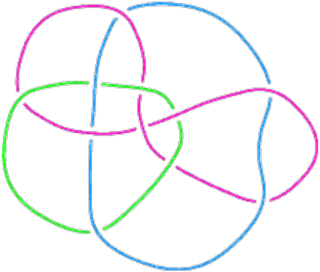


Figure: Link 11n431 in Thistlethwaite table

2. Equivalence of Knots

2. Equivalence of Knots

Given two knot diagrams, how do we determine whether or not they represent the same knot?

2. Equivalence of Knots

Given two knot diagrams, how do we determine whether or not they represent the same knot?

Definition

We say two knots K, K' are equivalent if they are related via an ambient isotopy.

A simple way of thinking of this is as a movie reel, where every still is a topological space that is an embedding of the original space.

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Examples of invariants:

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Examples of invariants:

- Fary-Milnor
  - Given a knot K embedded in $\mathbb{R}^3$, if $\int_K \kappa(p)\mathrm{dp} \leq 4\pi$, then K is the unknot.

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Examples of invariants:

- Fary-Milnor
  - Given a knot K embedded in $\mathbb{R}^3$, if $\int_K \kappa(p)\mathrm{d}p \leq 4\pi$, then K is the unknot.
- Jones Polynomial
  - Assigns to each oriented knot (resp. link) diagram a univariate polynomial over the integers.

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Examples of invariants:

- Fary-Milnor
  - Given a knot K embedded in $\mathbb{R}^3$, if $\int_K \kappa(p)\text{dp} \leq 4\pi$, then K is the unknot.
- Jones Polynomial
  - Assigns to each oriented knot (resp. link) diagram a univariate polynomial over the integers.
- HOMFLY Polynomial
  - Generalization of Jones Polynomial in that it is a bi-variate polynomial.

Checking knot equivalence is hard to do from the definition, so instead we use knot invariants.

Examples of invariants:

- Fary-Milnor
    - Given a knot K embedded in $\mathbb{R}^3$, if $\int_K \kappa(p)\mathrm{dp} \leq 4\pi$, then K is the unknot.
- Jones Polynomial
    - Assigns to each oriented knot (resp. link) diagram a univariate polynomial over the integers.
- HOMFLY Polynomial
    - Generalization of Jones Polynomial in that it is a bi-variate polynomial.

CAVEAT: The latter two are not complete invariants! There exist non-isotopic knots/links which have the same HOMFLY.

How do we represent knots on computers?

Cantarella has a C-library called plCurve which provides structures for representing polygonal walks in 3-space as well as planar diagrams of knots. In particular, he has exhaustive lists of all planar diagrams with 11 crossings or less up to labelled graph isomorphisms.

```
pd Bw4JBAQEAwMDAwICAQ4AAgIEBvQAAAA 1
nv 7
0 7 1 8        Edge labels going CC around
0 11 13 10
1 7 2 6
2 10 3 9
3 13 4 12
4 11 5 12
5 8 6 9
ne 14
1,0 -> 0,0     (row, col) in adj. matrix
0,2 -> 2,0
2,2 -> 3,0
3,2 -> 4,0
4,2 -> 5,0
5,2 -> 6,0
6,2 -> 2,3
2,1 -> 0,1
0,3 -> 6,1
6,3 -> 3,3
3,1 -> 1,3
1,1 -> 5,1
5,3 -> 4,3
4,1 -> 1,2
nc 1
14 : 0 1 2 3 4 5 6 7 8 9 10 11 12 13
nf 9
4 : + 0 + 8 - 5 - 11  Face orientation
4 : - 0 - 10 - 2 + 7  w.r.t. edges
4 : + 3 - 12 + 5 + 9
3 : + 1 - 6 - 8
3 : + 2 - 9 + 6
3 : - 3 + 10 - 13
3 : - 4 + 13 + 11
2 : - 1 - 7
2 : + 4 + 12
```

Figure: pdstor

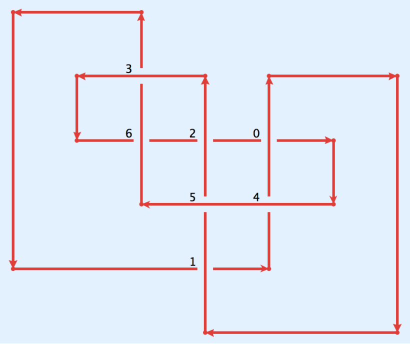


Figure: Planar Diagram 7n1

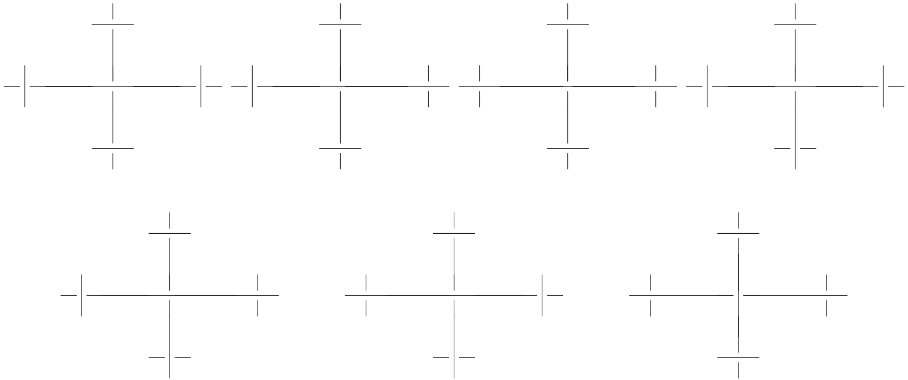
1. Crash Course on Knot Theory
2. Topological Model for Topo II
3. Numerical Experiments

Other models for Topo II have been proposed:

- Buck et. al. modeled by having strand passage occur when certain "hook" patterns are present on integer lattice walks (2004).
- Hua et. al. proposed a Monte Carlo method implemented on integer lattice walks (2007).

The above models deal with local geometry of a lattice walk. Could we do better with a topological model instead?

Idea: what if Topo II at any given crossing can view crossing information at crossings 1 edge away? We could then exhaustively enumerate the possibilities of configurations and create rules for Topo II to fire on.



Above are the 7 distinct possible crossing configurations, up to symmetry of the square. In other words, there are 2^7 rules under this model.

1 Crash Course on Knot Theory

2 Topological Model for Topo II

3 Numerical Experiments

We can think of our rules as discrete Markov chains!

Definition

A **discrete Markov chain** is a stochastic process over a finite state space in which the probability of an event depends only on the current state.

Discrete Markov chains have the nice feature that they can be represented as a matrix!

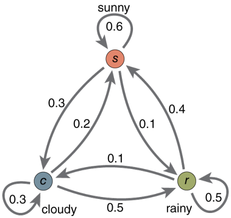


Figure: A simple Markov chain over a state space with three elements. We can represent this as $\begin{bmatrix} 0.3 & 0.5 & 0.2 \\ 0.1 & 0.5 & 0.4 \\ 0.3 & 0.1 & 0.6 \end{bmatrix}$, ordering rows and columns by alphabetical order.

Under suitable conditions on the matrix, the steady state of a discrete Markov chain is given by its leading (eigenvalue 1) eigenvector. How should we compare steady state distributions?

Definition

The **entropy** of a discrete distribution over N events is given by $\sum_{i=1}^{N} p_i \log_2\left(\frac{1}{p_i}\right)$, where p_i is the probability of event i occurring.

- Low entropy $\iff$ distribution is concentrated.
- High entropy $\iff$ distribution has a wide spread.

We're ready to define the procedure for our experiment:

- Fix n to be the maximal number of crossings to consider.
- For each rule, generate the Markov matrix over the space of HOMFLYs with ncross $\leq n$.
- Find steady state for each rule and compare knot-type entropy against the null hypothesis: Topo II always fires.

In the following slides, we'll see results for $n = 6$. In total, we iterated over 51,384 knot diagrams. That's about 300,000 crossings y'all.

$$
\begin{pmatrix}
\frac{499483}{532485} & \frac{6359}{236660} & \frac{18593}{709980} & \frac{587}{70998} & \frac{4}{106497} & 0 & \frac{52}{177495} & \frac{47}{177495} & 0 & 0 & 0 & \frac{1}{212994} & \frac{1}{212994} & \frac{4}{106497} & 0 & \frac{1}{106497} \\
\frac{14033}{23355} & \frac{2906}{7785} & \frac{199}{23355} & \frac{146}{23355} & \frac{37}{15570} & 0 & \frac{109}{15570} & 0 & \frac{1}{1557} & \frac{1}{1557} & 0 & \frac{1}{4671} & 0 & \frac{1}{4671} & 0 & 0 \\
\frac{1588}{2595} & \frac{71}{7785} & \frac{934}{2595} & \frac{34}{4671} & 0 & \frac{151}{46710} & 0 & \frac{307}{46710} & 0 & \frac{1}{1557} & \frac{1}{1557} & 0 & \frac{1}{4671} & \frac{1}{4671} & 0 & \frac{1}{4671} \\
\frac{4541}{6465} & \frac{19}{862} & \frac{347}{12930} & \frac{1583}{6465} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \frac{1}{862} & \frac{1}{862} & 0 & 0 & \frac{2}{1293} \\
\frac{1}{7} & \frac{2}{3} & 0 & 0 & \frac{2}{21} & 0 & \frac{2}{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{21} & 0 & \frac{17}{21} & 0 & 0 & \frac{2}{21} & 0 & \frac{1}{21} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{151}{375} & \frac{323}{750} & 0 & \frac{1}{75} & \frac{1}{50} & 0 & \frac{2}{15} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{161}{375} & 0 & \frac{101}{250} & \frac{1}{75} & 0 & \frac{1}{50} & 0 & \frac{2}{15} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & \frac{1}{2} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{6} & \frac{1}{3} & 0 & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{6} & 0 & \frac{1}{3} & \frac{1}{2} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{1}{2} & \frac{1}{4} & \frac{1}{4} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
\frac{2}{3} & 0 & 0 & \frac{1}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{pmatrix}
$$

Figure: Markov matrix for the rule which always fires (rule 1) over HOMFLYs of 6 or fewer crossing diagrams.

0.908077053436438	0.989010931165697
0.0406080522634194	0.00510194843160136
0.0388126667852605	0.00524309784280032
0.0107224489084l101	0.000644022559901670
0.000158776907367551 5	0
0.000151501134395252 1	0
0.00065243049022519 2	0
0.00058011312905268 1	0
0.0000260809584222346	0
0.0000510088112066023	0
0.00002492785278436768	0
0.0000253960813109491	0
0.0000250117127649935	0
0.0000511100782441658	0
0	0
0.0000334214506974476	0

Figure: Steady state distributions for rule 1 (left) and rule 127 (right). Rule 127 is the rule which fires only on locally alternating crossings.

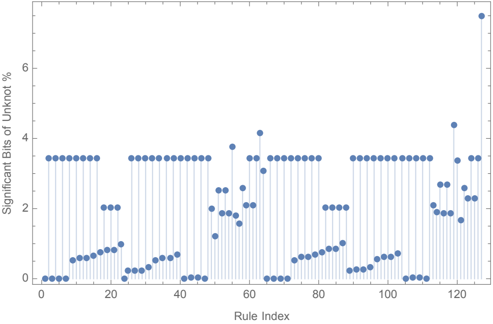


Figure: Change in significant bits of the unknot percentage from the null-hypothesis steady state. Rule 128 (never fire) omitted.

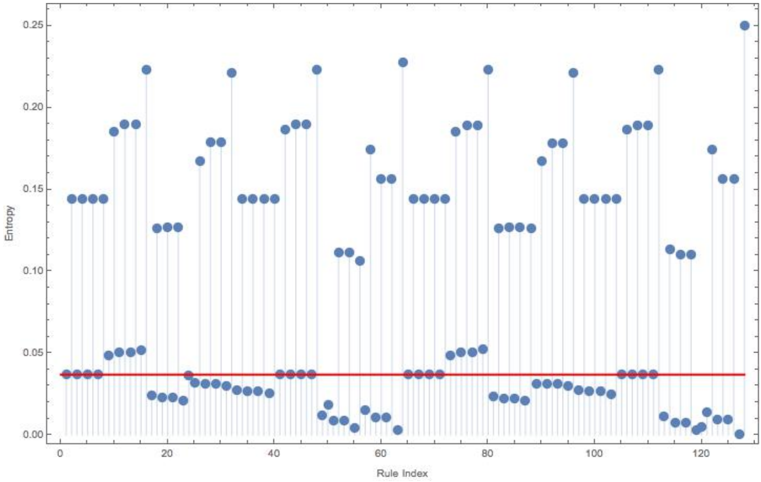


Figure: Entropy plot! The two rules with lowest knot type entropy are rule 119 and rule 127.

More fun plots!

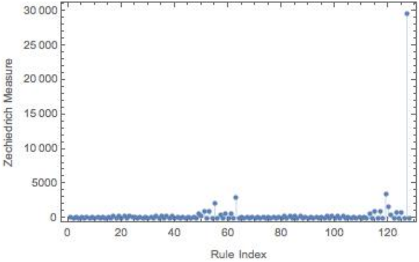


Figure: Zechiedrich measure plot. This is the percentage of unknots in steady state over the total of the remaining percentages.

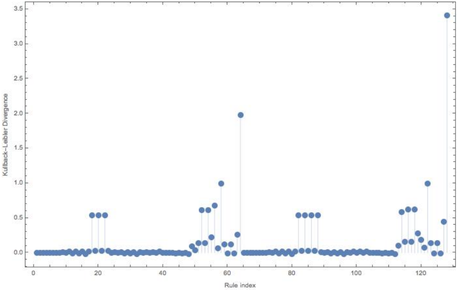


Figure: Kullback-Liebler divergence against rule 1. Intuitively measures how much info is lost when we use one distribution to approximate another. Rule 128 omitted.

Other neat things we discovered:

- We computed diagram entropy in addition to knot type (HOMFLY) entropy.
- Rule 127 is among the maximizers of diagram entropy!
- Rule 127 has the maximal covariance between knot and diagram entropy!
- Rule 119 comes in second for knot-type entropy, Zechiedrich measure, and significant bits of unknot in steady state.

Rule 127 seems to be a clear winner. But why is this mysterious rule 119 coming in second place?

Definition

Given an oriented knot K, define the **non-alternating measure** Nalt: $\{\text{Oriented knot diagrams}\} \to \mathbb{Z}$ via

$$\mathrm{Nalt}(K) = \sum_{\mathrm{edge} \in E_K} h(\mathrm{edge})$$

where $h(\mathrm{edge}) = 1$ if the edge is going over (resp. under) both head and tail crossing, and 0 otherwise.

Definition

Define nalt : $\{$ Oriented Crossings $\} \to \mathbb{Z}$ via

$$\mathrm{nalt}(x) = \sum_{\text{edges around } x} h(\mathrm{edge}).$$

Lemma

For an oriented knot diagram K with no 1-edge loops,

$$Nalt(K) = \frac{1}{2} \sum_{crossings\ x} nalt(x)$$

Proof.

Each edge is visited twice under the crossing nalt measure, as there is a distinct head and tail vertex for each edge. □

So if we're interested in maximizing Nalt, we should maximize each nalt in the sum! If you look at which rules maximize nalt under a toggle change, rule 119 is the maximizer with rule 127 in second.

There are a lot of questions that we still have. Here are just a few:

- Are our results just a low-crossing phenomena?
- How do our results on unknotting relate to unlinking?
- Who prevails in the end? Rule 119 or Rule 127? (or someone else?)
- If rule 127 remains the dominant unknotting rule, is there a mathematical reason why?

Thank you for your attention!

$\diamond \diamond \diamond$