Gaussian Random Embeddings of Multigraphs

Jason Cantarella,* Tetsuo Deguchi,† Clayton Shonkwiler,‡ and Erica Uehara†

(Dated: January 31, 2020)

This paper generalizes the Gaussian random walk and Gaussian random polygon models for linear and ring polymers to polymer topologies specified by an arbitrary multigraph $\mathbf{G}$. Probability distributions of monomer positions and edge displacements are given explicitly and the spectrum of the graph Laplacian of $\mathbf{G}$ is shown to predict the geometry of the configurations. This provides a new perspective on the James–Guth–Flory theory of phantom elastic networks. The model is based on linear algebra motivated by ideas from homology and cohomology theory. It provides a robust theoretical foundation for more detailed models of topological polymers.

1. INTRODUCTION

There has been increasing interest in polymers with topologies more complicated than the standard linear polymer in recent years. Branched, multicyclic, "tadpole" or "lasso," and bottlebrush polymers have all been studied. Very recently, the Tezuka lab [9, 10] has started to synthesize polymers with even more complicated topologies such as a $K_{3,3}$ graph.

Polymers are traditionally modeled by random walks, and for previous topologies, the random walk model was relatively simple. The walk is a sum of steps which are either independent (along a branch) or part of a collection of steps conditioned on the hypothesis that they sum to zero (along an isolated loop). This conditioning introduces a small dependence between steps, but the hypothesis is a single linear constraint which can be handled by elementary methods.

For polymers with multiple loops, the steps are conditioned on a hypothesis which is much more complicated– the sum of steps around *any* loop in the polymer must vanish. Further, the same edge is likely part of many loops at the same time. Understanding the dependency structure of the edges is considerably more complicated in this case and seemed somewhat daunting.

We solve the loop constraint problem by recasting it in vector calculus terms: the steps in the random walk can be interpreted as the (discrete) gradient of the vertex positions, so finding steps which satisfy the constraints is analogous to determining which vector fields on a complicated domain admit a scalar potential. From a mathematician's point of view we are translating a problem in homology theory (finding a basis for the space of loops) to a problem in de Rham cohomology (finding potentials for certain forms). Though our approach implicitly draws on these ideas, the

*Mathematics Department, University of Georgia, Athens GA
†Ochanomizu University, Tokyo, Japan
‡Department of Mathematics, Colorado State University, Fort Collins CO

solution we present will be entirely elementary: the paper is self-contained and requires no prior knowledge of homology or cohomology.

In particular, this provides a natural framework for understanding the probability distribution of edges in polymers of arbitrary complexity given by a multigraph $\mathbf{G}$. As an application, we give a simple algorithm for sampling Gaussian random embeddings of arbitrary (connected) multigraphs, and show that the spectral graph theory of $\mathbf{G}$ provides valuable insight into the geometry of these random embeddings. This is the right theoretical foundation for handling complicated polymer topologies, and hence should be the place to start building models which add more physically realistic constraints (self-avoidance, external fields, bending energies, and so forth).

A related problem was faced in the classical theory of elasticity [5–7], where rubbers and polymer gels were taken to have complicated topological structures from the start. There, the vertices x_i were assumed to have the canonical distribution based on the potential function $\sum_{i,j \text{ joined by edge}} \|x_i - x_j\|^2$. As a quadratic form, this potential is $\langle x, Lx\rangle$, where L is the graph Laplacian. We find that taking the simplest prior on edge distributions in our theory exactly recovers the James–Guth–Flory theory of phantom elastic networks, meaning that our model both explains and extends the classical picture. Furthermore, we have checked that our theory is also consistent with the standard approach via the Langevin equation: our distribution agrees with the equilibrium distribution obtained after thermal relaxation.

2. DEFINITIONS AND PRELIMINARIES

We start by fixing notation. We assume that we have a connected multigraph $\mathbf{G}$ with $\mathbf{e}$ edges $e_1, \ldots, e_{\mathbf{e}}$ and $\mathbf{v}$ vertices $v_1, \ldots, v_{\mathbf{v}}$. We assume that the graph is directed,[1] so that each edge has a head vertex $\mathrm{head}(e_i)$ and a tail vertex $\mathrm{tail}(e_i)$. When e_i is a loop edge, $\mathrm{head}(e_i) = \mathrm{tail}(e_i)$, and since we allow multiple edges between vertices, it is no problem if $\mathrm{head}(e_i) = \mathrm{head}(e_j)$ and $\mathrm{tail}(e_i) = \mathrm{tail}(e_j)$.

We discuss embeddings of our multigraphs using the following terminology:

Definition 1. Let $\mathbf{G}$ be a connected, directed multigraph. A *vertex vector* for $\mathbf{G}$ is an $x \in (\mathbb{R}^d)^{\mathbf{v}}$ where $x_i \in \mathbb{R}^d$ is the position of vertex v_i and $x^k = (x^k_1, \ldots, x^k_{\mathbf{v}}) \in \mathbb{R}^{\mathbf{v}}$ is the vector of kth coordinates of all vertex positions.

An *edge vector* w for $\mathbf{G}$ is a $w \in (\mathbb{R}^d)^{\mathbf{e}}$ where $w_j \in \mathbb{R}^d$ is the displacement along edge e_j and $w^k = (w^k_1, \ldots, w^k_{\mathbf{e}}) \in \mathbb{R}^{\mathbf{e}}$ is the vector of all kth coordinates of the edge displacements. See Fig. 1.

In addition to thinking of the vector x^k of kth coordinates of vertices as a vector in $\mathbb{R}^{\mathbf{v}}$, it will

[1] The direction picked for each edge is arbitrary and won't affect the theory. The directions just need to be consistent throughout any particular set of calculations.

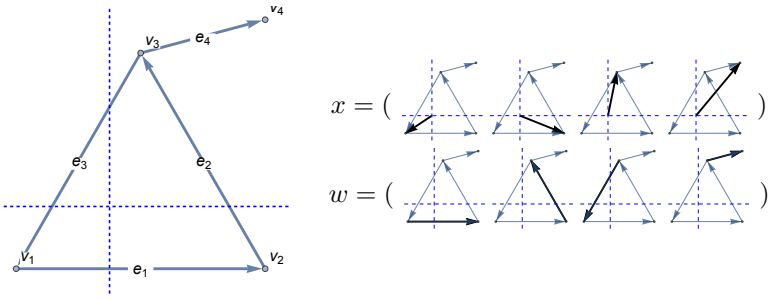


FIG. 1: A particular graph embedding in $\mathbb{R}^2$, along with the components of its vertex vector $x \in (\mathbb{R}^2)^4$ and edge vector $w \in (\mathbb{R}^2)^4$.

also be useful to think of it as a (scalar) function $x^k \colon \{v_1, \ldots, v_{\mathbf{v}}\} \to \mathbb{R}$ on the vertices of the graph.

By contrast, the vector w^k of kth coordinates of edge displacements can be interpreted either as a vector in $\mathbb{R}^{\mathbf{e}}$ or as a map $w^k \colon \{e_1, \ldots, e_{\mathbf{e}}\} \to \mathbb{R}$, which we think of as a vector field on the multigraph. Since $\mathbf{G}$ is directed, the sign of $w^k(e_j)$ uniquely determines a direction of flow along edge e_j.

We now recall the relationship between scalar functions and vector fields on multigraphs. In analogy to vector calculus, it's standard to define two linear maps between the spaces of functions and vector fields. The gradient field $\operatorname{grad} f$ of a function $f \colon \{v_1, \ldots, v_{\mathbf{v}}\} \to \mathbb{R}$ is the vector field defined by

$$(\operatorname{grad} f)(e_j) = f(\operatorname{head}(e_j)) - f(\operatorname{tail}(e_j)), \tag{1}$$

and the divergence $\operatorname{div} u$ of a vector field $u \colon \{e_1, \ldots, e_{\mathbf{e}}\} \to \mathbb{R}$ is the function

$$(\operatorname{div} u)(v_i) = \sum_{j=1}^{\mathbf{e}} \begin{cases} -u(e_j) & \text{if } v_i \text{ is (only) the head of } e_j, \\ +u(e_j) & \text{if } v_i \text{ is (only) the tail of } e_j, \\ 0 & \text{if } v_i \text{ is both head and tail of } e_j, \\ 0 & \text{else.} \end{cases}$$

As a matrix, $\operatorname{div} = -B$, where B is the $\mathbf{v} \times \mathbf{e}$ incidence matrix with

$$B_{ij} = \begin{cases} +1 & \text{if } v_i \text{ is (only) the head of } e_j, \\ -1 & \text{if } v_i \text{ is (only) the tail of } e_j, \\ 0 & \text{if } v_i \text{ is both head and tail of } e_j, \\ 0 & \text{else.} \end{cases} \tag{2}$$

Loop edges contribute zero columns, and when there are multiple edges connecting two vertices B will have repeated columns.

On the other hand, grad is the $\mathbf{e} \times \mathbf{v}$ matrix B^T. As usual, we will say that f is a *potential function* for u if $\operatorname{grad} f = u$. We will say u is a *conservative* vector field if there exists some f so that $\operatorname{grad} f = u$. If u is conservative, there is a one-dimensional family of possible potential functions $f + C$ where C is a constant function on $\mathbf{G}$.

As a subspace of the space of all vector fields, we can characterize the conservative vector fields using a version of the Helmholtz decomposition:

Theorem 2. *The vector space $\mathbb{R}^{\mathbf{e}}$ of vector fields on $\mathbf{G}$ is spanned by a $(\mathbf{v}-1)$-dimensional subspace of* conservative *vector fields $u = \operatorname{grad} f$ and an orthogonal $(\mathbf{e}-\mathbf{v}+1)$-dimensional space of* divergence-free *vector fields u with $\operatorname{div} u = 0$.*

Proof. The divergence-free fields are by definition $\ker B$. Their orthogonal complement $(\ker B)^\perp = \operatorname{im} B^T = \operatorname{im} \operatorname{grad}$. Since $\mathbf{G}$ is connected, it is not hard to show directly that $\ker \operatorname{grad}$ is one-dimensional (only the constant functions have no gradient), so $\operatorname{im} \operatorname{grad}$ has dimension $\mathbf{v}-1$. The dimension of $\ker \operatorname{div}$ follows. □

Consider the problem of finding a potential function f given a vector field $u \in \operatorname{im} B^T$. Since $\ker \operatorname{grad}$ is not empty, this problem is underdetermined: adding something in $\ker \operatorname{grad}$ to any particular solution f still yields a function with $\operatorname{grad} f = B^T f = u$. We can define a unique canonical potential function f for u by taking the potential function of minimum norm (among all possible potential functions for u).

This minimum norm potential function can be computed conveniently using the Moore–Penrose pseudoinverse. We now recall the definition and some standard properties. The pseudoinverse A^+ of a matrix A is "as close as possible" to the inverse of a matrix which is not full rank. It may be computed by taking the singular value decomposition $A = U\Sigma V^T$ and then defining $A^+ = V\Sigma^+ U^T$, where Σ^+ is the diagonal matrix whose nonzero entries are the reciprocals of the corresponding *nonzero* entries in S.

Equivalently, the pseudoinverse is defined to be the unique matrix satisfying the four *Moore–Penrose conditions*:

$$AA^+A = A, \quad A^+AA^+ = A^+, \quad AA^+ = (AA^+)^T, \quad A^+A = (A^+A)^T. \tag{3}$$

It will also be helpful to recall that AA^+ is the orthogonal projector onto $\operatorname{im} A = (\ker A^T)^\perp$ and A^+A is the orthogonal projector onto $\operatorname{im} A^T = (\ker A)^\perp$ and that $(A^T)^+ = (A^+)^T$, so there is no ambiguity in writing A^{T+} for the combination.

It is a standard fact that

Theorem 3. *The smallest x minimizing $\|Ax - b\|^2$ is given by $x_0 = A^+ b$. Further, $x_0 \in (\ker A)^\perp$, and $Ax_0 = b \iff b \in \operatorname{im} A$.*

Using the theorem, we see that if $u \in \operatorname{im}\operatorname{grad} = \operatorname{im} B^T$, then there is a unique potential function $f = B^{T+}u$ in $(\ker B^T)^\perp$. As we noted above, $\ker B^T$ is the one-dimensional[2] space of constant functions, so $f \in (\ker B^T)^\perp$ must have $\sum f(v_i) = 0$. It is helpful to note that $B^{T+}u$ is defined for any vector field u (whether or not u is in $\operatorname{im} B^T$), though $B^T B^{T+} w = u$ if and only if $u \in \operatorname{im} B^T$.

The discussion above solves the embedding problem while adroitly sidestepping any mention of the loop constraints themselves. To recover the loops, observe that every loop in the graph has a divergence-free field that flows around it and every vector field perpendicular to that field obeys the corresponding loop constraint. The fields flowing around the loops span $\ker \operatorname{div}$ (for more details, see [8]), whose dimension $\mathbf{e} - \mathbf{v} + 1$ is exactly the cycle rank of the graph.

3. GAUSSIAN RANDOM EMBEDDINGS

With the language above, we see immediately that a collection of scalar functions $x^1, \ldots, x^d$ on $\mathbf{G}$ defines an embedding of $\mathbf{G}$ into $\mathbb{R}^d$ by letting x_i^k – the kth coordinate of vertex v_i – be $x^k(v_i)$. On the other hand, a collection of vector fields $w^1, \ldots, w^d$ on $\mathbf{G}$ can be interpreted as a collection of coordinates of edge displacements, but these displacements define an embedding of $\mathbf{G}$ *only if* $w^k \in \operatorname{im}\operatorname{grad}$ for each $k = 1, \ldots, d$. In this case, the w's define a unique embedding $x^k = B^{T+} w^k$ with $\sum x_i^k = 0$. We will call such an embedding *centered* because its center of mass is at the origin.

In analogy to the requirement that the displacements w of a Gaussian random walk are sampled from a standard Gaussian on $(\mathbb{R}^d)^{\mathbf{e}}$, we make the least restrictive assumption about the distribution of $w^k = \operatorname{grad} x^k$ that we can:

Definition 4. A Gaussian random embedding of $\mathbf{G}$ into $\mathbb{R}^d$ with kth coordinates of the edge displacements w^k is defined by the assumption that w^k is sampled from a standard normal on $\mathbb{R}^{\mathbf{e}}$ conditioned on the hypothesis that $w^k \in \operatorname{im}\operatorname{grad} = \operatorname{im} B^T$.

We can immediately compute the covariance matrix of w^k:

Theorem 5. *If w is the edge vector of a Gaussian random embedding of $\mathbf{G}$ into $\mathbb{R}^d$, then the vector w^k of kth coordinates is distributed as $\mathcal{N}(0, B^T B^{T+}) = \mathcal{N}(0, B^+ B)$.*

Proof. To condition on the hypothesis that a multivariate normal is restricted to a linear subspace, we transform the normal by orthogonal projection to that subspace. Using the fact that $B^T B^{T+}$ is the (symmetric) orthogonal projector onto $\operatorname{im} B^T$ we can compute that the covariance matrix of

[2] Recall that $\mathbf{G}$ is assumed to be connected.

the projected variable is

$$B^T B^{T+} I (B^T B^{T+})^T = B^T B^{T+} = (B^+ B)^T = B^+ B.$$

This completes the proof. □

Given the edge displacements, it is certainly possible to determine the vertex vector of the corresponding centered embedding directly: for the path, the vertex positions are just the partial sums of the displacement vectors. Strictly speaking, this is based on a choice of (the unique) spanning tree for the path graph, but for more complicated multigraph topologies one needs to find a spanning tree before computing the partial sums. It is generally simpler to compute B^{T+} and use the equation $x^k = B^{T+} w^k$.

Corollary 6. *If x is the vertex vector of a Gaussian random embedding of $\mathbf{G}$ into $\mathbb{R}^d$, the vector x^k of kth coordinates can be constructed by taking $x^k = B^{T+} y^k$ where y^k is distributed as $\mathcal{N}(0, I_{\mathbf{e}})$ on $\mathbb{R}^{\mathbf{e}}$.*

Proof. We know that we can construct x^k by taking $x^k = B^{T+} w^k$, where w^k is distributed as $\mathcal{N}(0, B^T B^{T+})$ on $\mathbb{R}^{\mathbf{e}}$. In the proof of Theorem 5, we showed that $w^k = B^T B^{T+} y^k$, where y^k is distributed as $\mathcal{N}(0, I_{\mathbf{e}})$. Thus $x^k = B^{T+} B^T B^{T+} y^k = B^{T+} y^k$ using (3). □

We are now going to directly compute the covariance matrix of x^k. It will help to recall that $BB^T = \operatorname{div}\operatorname{grad}$ is also known as the (multi)*graph Laplacian* L. It's a standard fact that $L = D - A$ where D is the diagonal matrix with entries equal to the degrees of the vertices of the graph and A is the adjacency matrix with A_{ij} recording the number of edges connecting v_i and v_j. For example, if $i = j$ then A_{ii} is twice the number of loops based at v_i; if $i \neq j$ then A_{ij} is the total number of edges e_k with $\{v_i, v_j\} = \{\operatorname{head}(e_k), \operatorname{tail}(e_k)\}$.

Theorem 7. *If x is the vertex vector of a Gaussian random embedding of $\mathbf{G}$ into $\mathbb{R}^d$, then the vector x^k of kth coordinates is distributed as $\mathcal{N}(0, L^+)$ where L is the graph Laplacian of $\mathbf{G}$. Hence x^k may be constructed by*

$$x^k = (L^+)^{1/2} y^k \tag{4}$$

where $(L^+)^{1/2}$ is any symmetric square root of L^+ and y^k is distributed as $\mathcal{N}(0, I_{\mathbf{v}})$ on $\mathbb{R}^{\mathbf{v}}$.

Proof. For the first part, we know that $x^k = B^{T+} y^k_{\mathbf{e}}$, where $y^k_{\mathbf{e}}$ is distributed as $\mathcal{N}(0, I_{\mathbf{e}})$ on $\mathbb{R}^{\mathbf{e}}$. Therefore x^k is a multivariate normal whose covariance matrix is

$$B^{T+} I_{\mathbf{e}} (B^{T+})^T = B^{T+} B^+ = (BB^T)^+ = L^+,$$

using the Moore-Penrose conditions and the fact that $(AA^T)^+ = A^{T+} A^+$ for any matrix A. The construction (4) is justified by the fact that $(L^+)^{1/2} y^k$ is also a multivariate normal with covariance matrix L^+. To construct a symmetric square root for L^+, we note that since L is a real symmetric matrix, it has a singular value decomposition in the form $U \Sigma U^T$. This means we can let $(L^+)^{1/2} = U (\Sigma^+)^{1/2} U^T$. □

We can sample random conformations of a multigraph $\mathbf{G}$ in two ways: either by generating edge vectors normally and converting to vertex positions as in Corollary 6, or by sampling Gaussians with covariance matrix L^+ as in Theorem 7. The latter is almost always preferable, since any multigraph with loops has at least as many edges as vertices and the covariance matrix L^+ only has to be computed once. This, then, gives a powerful computational tool for estimating arbitrary quantities of interest using Monte Carlo integration.

We now do an example. For ring polymers, the graph $\mathbf{G}$ is a cycle graph and the Laplacian is:

$$L = \begin{pmatrix} 2 & -1 & \cdots & \cdots & -1 \\ -1 & 2 & -1 & \cdots & 0 \\ 0 & -1 & 2 & -1 & 0 \\ \vdots & \vdots & \ddots & \ddots & -1 \\ -1 & \cdots & \cdots & -1 & 2 \end{pmatrix}$$

A singular value decomposition $L = U\Sigma U^T$ is well-known, as L is the discrete Fourier transform. We can write it as

$$\sigma_j = 4\sin^2\frac{\pi j}{\mathbf{v}}, \qquad U_{\ell j} = \alpha_j \begin{cases} \sin\frac{2\pi j\ell}{\mathbf{v}}, & 1 \le j < \frac{\mathbf{v}}{2}, \\ \cos\frac{2\pi j\ell}{\mathbf{v}}, & \frac{\mathbf{v}}{2} \le j \le \mathbf{v}. \end{cases} \tag{5}$$

where $\alpha_j = \sqrt{2/\mathbf{v}}$ unless $j = \mathbf{v}$ or $j = \mathbf{v}/2$ (if $\mathbf{v}$ is even), in which case $\alpha_j = \sqrt{1/\mathbf{v}}$. This allows us to construct a square root $(L^+)^{1/2}$ whose singular values are $\frac{1}{2}\csc\frac{\pi j}{\mathbf{v}}$ for $1 \le j < \mathbf{v}$ together with a single 0 corresponding to $\sigma_{\mathbf{v}} = 0$ in (5). We note this Fourier description for Gaussian ring polymers was considered by Bloomfield and Zimm [1] and Eichinger [3].

4. FOURIER-TYPE ANALYSIS OF RANDOM GRAPH EMBEDDINGS

In the proof of Theorem 7, we saw that the vertex vectors of a Gaussian random embedding could be generated by taking

$$x^k = U(\Sigma^+)^{1/2}U^T y^k, \quad y^k \text{ distributed as } \mathcal{N}(0, I_{\mathbf{v}}),$$

where U is the matrix of eigenvectors of L. Another way to look at this equation is to see that x^k is a weighted linear combination of the eigenvectors with random (normal) coefficients where the weights are given by the singular values on the diagonal of $(\Sigma^+)^{1/2}$:

$$x^k = \sum_{\substack{j\in\{1,\ldots,\mathbf{v}\} \\ \sigma_j \neq 0}} \frac{1}{\sqrt{\sigma_j}} u_j \cdot y_j$$

where each y_j is distributed as $\mathcal{N}(0,1)$. It is clear that the eigenfunctions of L with small eigenvalues σ_j are expected to play a much larger role in determining the vertex vector x^k than the eigenfunctions with larger eigenvalues.

We can make this observation precise by recalling a few facts from linear algebra. An optimal rank-p approximation A_p to a matrix A with singular value decomposition $U\Sigma V^T$ is given by replacing Σ with another diagonal matrix Σ_p keeping a collection[3] S_p of the p largest singular values of Σ and setting the remaining singular values to zero. We can now define the rank-p approximation to a Gaussian random graph embedding

$$x_p^k = (L^+)_p^{1/2} y^k$$

and note that

$$x^k - x_p^k = ((L^+)^{1/2} - (L^+)_p^{1/2}) y^k$$

is itself a Gaussian random vector. If σ_j^+ are the singular values of L^+, and S_p^+ is a collection of the p largest σ_j^+, then the expected squared norm of this difference vector is

$$\mathcal{E}\left(\left\|x^k - x_p^k\right\|^2\right) = \operatorname{tr}\left(L^+ - L_p^+\right) = \sum_{\sigma_j^+ \notin S_p} \sigma_j^+. \tag{6}$$

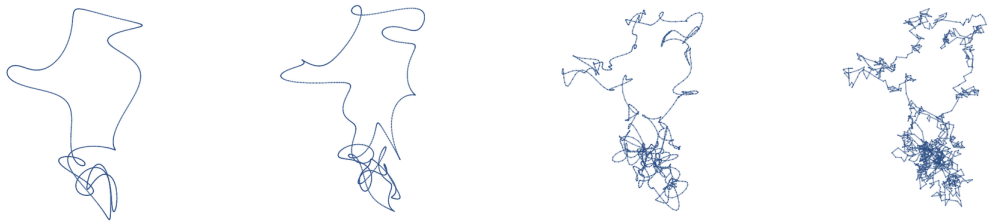
For the cycle graph, we know from (5) that $\sigma_j^+ = \frac{1}{4}\csc^2\frac{\pi j}{\mathbf{v}}$ for $j \in \{1,\ldots,\mathbf{v}-1\}$ and $\sigma_{\mathbf{v}}^+ = 0$. Summing the σ_j^+ tells us that $\mathcal{E}(\left\|x^k\right\|^2) = \frac{1}{12}(\mathbf{v}^2-1)$ and (for p even)

$$\mathcal{E}\left(\left\|x^k - x_p^k\right\|^2\right) = \sum_{j=p/2+1}^{\mathbf{v}-p/2-1} \sigma_j^+ \approx \int_{\frac{p+1}{2}}^{n-\frac{p+1}{2}} \frac{1}{4}\csc^2\frac{\pi t}{\mathbf{v}}\,\mathrm{d}t = \frac{\mathbf{v}}{2\pi}\cot\left(\frac{\pi}{2}\frac{p+1}{\mathbf{v}}\right). \tag{7}$$

Fig. 2 shows examples of low rank approximations to the ring polymer with $\mathbf{v} = 1000$ for $p \in \{20, 50, 250, 999\}$.

We see a similar phenomenon for a θ-curve where each edge has been subdivided 500 times. Here, we can compute the spectrum of the Laplacian numerically and calculate expected error estimates using (6), but we do not have explicit formulae for these estimates as we did for the cycle. The results are shown in Fig. 3.

[3] This collection is not always unique if the singular values of A are not all distinct; in this case, all matrices A_p constructed in this way are equally good rank-p approximations to A.



$x^k_{20} = (L^+)^{1/2}_{20} y^k$ $x^k_{50} = (L^+)^{1/2}_{50} y^k$ $x^k_{250} = (L^+)^{1/2}_{250} y^k$ $x^k = L^+ y^k$

FIG. 2: In these pictures, we construct a Gaussian ring polymer $x \in \mathbb{R}^3$ with 1000 edges by sampling $y^1, y^2, y^3 \in \mathbb{R}^{1000}$ from $\mathcal{N}(0, I_{1000})$ and computing $x^k_p = (L^+)^{1/2}_p y^k$ for various low-rank approximations $(L^+)^{1/2}_p$ of $(L^+)^{1/2}$. We can see that the low-rank approximations model the polymer rather well. We can measure the relative difference between shapes by computing $\|x - x_p\|^2 / \|x\|^2$. This can be compared to the estimate $\mathcal{E}(\|x - x_p\|^2)/\mathcal{E}(\|x\|^2)$ computed with (6). For these p, we get 0.056, 0.024, and 0.004 for the actual differences, which are rather close to the estimates 0.058, 0.024, and 0.005. We note that the cotangent estimate (7) of the expected error agrees with the actual expectation at this level of precision.

5. THE JAMES–GUTH–FLORY PHANTOM NETWORK MODEL

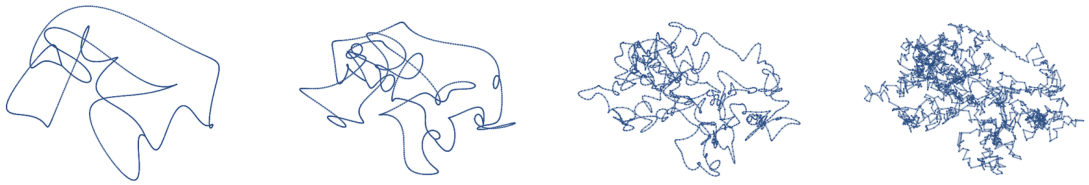
In our derivation, we started from the assumption that our conditions on the distribution of the displacements should be the least restrictive ones possible: that the displacements should be sampled from a standard normal on the linear subspace of displacements compatible with the graph topology. We then derived the matrix of covariances L^+ for the vertex coordinates via linear algebra. Now we compare this to the James–Guth–Flory approach. Consider the pdf of the one-dimensional normal distribution $\mathcal{N}(0, 1)$:

$$p(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}.$$

This is the canonical distribution of a system with potential energy $\frac{x^2}{2}$, for instance, a Hookean spring with rest length zero. By analogy, we may view the displacement vectors of a Gaussian random embedding as springs themselves, with energies proportional to the squares of their lengths. The total energy of embedding $x : \{v_1, \ldots, v_{\mathbf{v}}\} \to \mathbb{R}^d$ is then given by

$$\frac{1}{2} \sum_{j=1}^{\mathbf{e}} \|x(\mathrm{head}(e_j)) - x(\mathrm{tail}(e_j))\|^2 = \frac{1}{2} \sum_{k=1}^{d} \left\langle x^k, L x^k \right\rangle,$$

where L is the graph Laplacian of $\mathbf{G}$, and the equality follows from a straightforward computation expanding the middle summand. This suggests that the canonical distribution of the vertices should



$x^k_{30} = (L^+)^{1/2}_{30} y^k$ $x^k_{75} = (L^+)^{1/2}_{75} y^k$ $x^k_{375} = (L^+)^{1/2}_{375} y^k$ $x^k = (L^+)^{1/2} y^k$

FIG. 3: In these pictures, we construct a Gaussian random embedding $x \in \mathbb{R}^3$ of a 1499 vertex θ-curve where each edge has been subdivided into 500 pieces by sampling $y^1, y^2, y^3 \in \mathbb{R}^{1499}$ from $\mathcal{N}(0, I_{1499})$ and computing $x^k_p = (L^+)^{1/2}_p y^k$ for various low-rank approximations $(L^+)^{1/2}_p$ of $(L^+)^{1/2}$. Again, the low-rank approximations model the polymer rather well. We compute $\|x - x_p\|^2 / \|x\|^2$ and compare it to the estimate $\mathcal{E}(\|x - x_p\|^2)/\mathcal{E}(\|x\|^2)$ produced by (6). For these p, we get 0.0588, 0.0237, and 0.0047 for the actual differences, which are rather close to the estimates 0.0578, 0.0238, and 0.0046 obtained by numerically computing the singular values of L^+. We do not have a closed form for these difference estimates at present, but the example indicates that a similar phenomenon to the cycle is likely present– relatively few eigenfunctions describe most of the geometry of the embedding.

have pdf

$$p(x) \propto e^{-\frac{1}{2}\sum_{k=1}^{d} \langle x^k, Lx^k \rangle} = \prod_{k=1}^{d} e^{-\frac{1}{2}\langle x^k, Lx^k \rangle} \tag{8}$$

as given by (14) and (15) in Flory [5], (2) in Haliloglu et al. [6], or (4) in Estrada and Hatano [4]. The argument appears to originate with the foundational paper of Hubert James [7].

Usually, given a distribution with pdf proportional to $e^{-\frac{1}{2}\langle f, Qf \rangle}$, we can conclude that the random variable is normal with covariance matrix given by the inverse matrix Q^{-1}. However, this requires that the quadratic form Q be positive definite, so Q^{-1} exists. In our case, L is only positive *semi*definite: the constant embeddings in each coordinate form the one-dimensional subspace $\ker L$. This is to be expected, because our potential energy is invariant under translations, so we have to restrict the distribution to $(\ker L)^\perp$: the centered embeddings. If we agree to restrict the distribution to $(\ker L)^\perp$, the standard theorem is then that (8) is the pdf of a product of normals with covariance matrix[4] the *pseudo*inverse matrix L^+.

A third way to approach the problem would be to study the dynamics of relaxation using the Langevin equation. We have completed this calculation and found that the equilibrium distribution

[4] Here we follow the usual convention for multivariate normals $\mathcal{N}(0, \Sigma)$ that when the covariance matrix Σ is only positive semidefinite, the variable is understood to be restricted to the subspace where Σ is positive definite.

again gives the same result. Furthermore, the eigenmodes corresponding to larger singular values of L^+ decay slower and have longer relaxation times, suggesting that our approach may also be relevant to the study of non-equilibrium dynamics of topological polymers.

6. CONCLUSION

We have now given an explicit description of the distribution of vertex positions and displacements in a Gaussian model of topological polymers. Our description is computationally effective– one can use it to quickly and accurately sample ensembles of polymer shapes. Further, it provides insight into the shapes of the polymers by expressing them as random linear combinations of weighted eigenvectors of the graph Laplacian of the underlying multigraph $\mathbf{G}$.

In a companion paper [2], we use this setup to find exact expectations for vertex-vertex distances and radii of gyration for topological polymers, allowing us to compute their contraction factors precisely. This allows us to give an explicit formula for the asymptotic contraction factor of a topological polymer based on a subdivided graph as the number of subdivisions increases. We are also working on extending the theory to include self-avoidance constraints, and have computational "proof-of-concept" experiments generating configurations of hard spheres realizing various graph types. Finally, inhomogeneous and random networks should also be amenable to theoretical investigation using these methods.

Acknowledgements

The authors would like to acknowledge many friends and colleagues whose helpful discussions and generous explanations shaped this work. In particular we would like to acknowledge Yasuyuki Tezuka and Satoshi Honda for helpful discussions of topological polymer chemistry and thank Fan Chung for introducing us to spectral graph theory. This paper stemmed from a long series of discussions which started at conferences at Ochanomizu University and the Tokyo Institute of Technology. Cantarella and Shonkwiler are grateful to the organizers and the Japan Science and Technology Agency for making these possible. In addition, we are grateful for the support of the Simons Foundation (#524120 to Cantarella, #354225 to Shonkwiler), the Japan Science and Technology Agency (CREST Grant Number JPMJCR19T4) and the Japan Society for the Promotion of Science (KAKENHI Grant Number JP17H06463).

[1] Victor Bloomfield and Bruno H. Zimm. Viscosity, sedimentation, *et cetera*, of ring- and straight-chain polymers in dilute solution. *The Journal of Chemical Physics*, 44(1):315–323, 1966.

[2] Jason Cantarella, Tetsuo Deguchi, Clayton Shonkwiler, and Erica Uehara. Radius of gyration, contraction factors, and subdivisions of topological polymers. Preprint, 2020.

[3] Bruce E. Eichinger. Elasticity theory. I. Distribution functions for perfect phantom networks. *Macromolecules*, 5(4):496–505, 1972.

[4] Ernesto Estrada and Naomichi Hatano. Topological atomic displacements, Kirchhoff and Wiener indices of molecules. *Chemical Physics Letters*, 486(4-6):166–170, 2010.

[5] Paul J. Flory. Statistical thermodynamics of random networks. *Proceedings of the Royal Society of London. Series A. Mathematical, Physical and Engineering Sciences.*, 351(1666):351–380, 1976.

[6] Turkan Haliloglu, Ivet Bahar, and Burak Erman. Gaussian dynamics of folded proteins. *Physical Review Letters*, 79(16):3090–3093, 1997.

[7] Hubert M. James. Statistical properties of networks of flexible chains. *The Journal of Chemical Physics*, 15(9):651–668, 1947.

[8] Xiaoye Jiang, Lek-Heng Lim, Yuan Yao, and Yinyu Ye. Statistical ranking and combinatorial Hodge theory. *Mathematical Programming*, 127(1):203–244, 2011.

[9] Takuya Suzuki, Takuya Yamamoto, and Yasuyuki Tezuka. Constructing a macromolecular $K_{3,3}$ graph through electrostatic self-assembly and covalent fixation with a dendritic polymer precursor. *Journal of the American Chemical Society*, 136(28):10148–10155, 2014.

[10] Yasuyuki Tezuka. Topological polymer chemistry designing complex macromolecular graph constructions. *Accounts of Chemical Research*, 50(11):2661–2672, 2017.