

# Radius of gyration, contraction factors, and subdivisions of topological polymers

Jason Cantarella,<sup>\*</sup> Tetsuo Deguchi,<sup>†</sup> Clayton Shonkwiler,<sup>‡</sup> and Erica Uehara<sup>†</sup>

(Dated: April 13, 2020)

We consider the topologically constrained random walk model for topological polymers. In this model, the polymer forms an arbitrary graph whose edges are selected from an appropriate multivariate Gaussian which takes into account the constraints imposed by the graph type. We recover the result that the expected radius of gyration can be given exactly in terms of the Kirchhoff index of the graph. We then consider the expected radius of gyration of a topological polymer whose edges are subdivided into  $n$  pieces. We prove that the contraction factor of a subdivided polymer approaches a limit as the number of subdivisions increases, and compute the limit exactly in terms of the degree-Kirchhoff index of the original graph. This limit corresponds to the thermodynamic limit in statistical mechanics and is fundamental in the physics of topological polymers. Furthermore, these asymptotic contraction factors are shown to fit well with molecular dynamics simulations.

## 1. INTRODUCTION

We consider a classical model of polymers consisting of a collection of monomers connected by displacement vectors representing effective bond vectors between adjacent monomers. Here each bond vector represents a Kuhn length (or a multiple of the Kuhn length) along the polymer, as in [16]. For a linear polymer, we may think of the displacement vectors as independently sampled from multivariate Gaussian distributions, yielding a Gaussian random walk. The effective bond vectors (or displacement vectors) in a ring polymer must obey the additional condition that they must sum to zero, meaning that they are not independently sampled. Recently, polymers with more complicated topologies have been synthesized [18, 20], leading to an interest in modeling *topological* polymers where the underlying structure is not a path or a cycle but an arbitrary connected multigraph  $\mathbf{G}$ . This introduces a more complicated dependence structure between displacement vectors.

To describe the model in this case, it's helpful to introduce some notation:

**Definition 1.** Let  $\mathbf{G}$  be an arbitrary connected multigraph (loop edges and multiple edges are allowed) with an orientation on each edge. A *vertex vector* for  $\mathbf{G}$  is an  $x \in (\mathbb{R}^d)^\nu$  where  $x_i \in \mathbb{R}^d$  is the position of vertex  $v_i$  and  $x^k \in \mathbb{R}^\nu$  is the vector of  $k$ -th coordinates of all vertex positions. An *edge vector*  $w$  for  $\mathbf{G}$  is a  $w \in (\mathbb{R}^d)^e$  where  $w_j \in \mathbb{R}^d$  is the displacement along edge  $e_j$  and

---

<sup>\*</sup>Mathematics Department, University of Georgia, Athens, GA, USA

<sup>†</sup>Department of Physics, Ochanomizu University, Tokyo 112-8610, Japan

<sup>‡</sup>Department of Mathematics, Colorado State University, Fort Collins, CO, USA

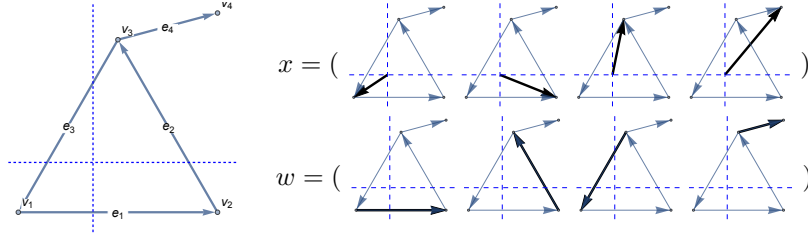


FIG. 1: A particular graph embedding in  $\mathbb{R}^2$ , along with the components of its vertex vector  $x \in (\mathbb{R}^2)^4$  and edge vector  $w \in (\mathbb{R}^2)^4$ .

$w^k \in \mathbb{R}^e$  is the vector of all  $k$ -th coordinates of the edge displacements. These are illustrated in Figure 1.

The vertex and edge vectors are related by the  $v \times e$  incidence matrix  $B$  of the graph  $\mathbf{G}$ , where

$$B_{ij} = \begin{cases} +1, & \text{if } v_i \text{ is (only) the head of } e_j \\ -1, & \text{if } v_i \text{ is (only) the tail of } e_j \\ 0, & \text{if } v_i \text{ is both the head and tail of } e_j \\ 0, & \text{if } v_i \text{ is neither the head nor tail of } e_j \end{cases}$$

and for each coordinate  $k$ ,  $w^k = B^T x^k$ . We note that multiple  $e_j$  may share the same head and tail vertices; these are simply repeated columns in  $B$ . While any  $x^k \in \mathbb{R}^v$  may be a vertex vector, only  $w^k \in \text{im } B^T$  can be an edge vector. These  $w^k$  have the special property that the sum of  $w_i^k$  around any loop in  $\mathbf{G}$  vanishes. We call  $w$  an *embeddable* edge vector if every  $w^k \in \text{im } B^T$ . We can then define

**Definition 2.** A Gaussian topological polymer or topologically constrained random walk (tcrw) in  $\mathbb{R}^d$  with underlying (multi)graph  $\mathbf{G}$  has all  $w^k$  sampled independently from a standard normal distribution on the embeddable edge vectors  $\text{im } B^T \subset \mathbb{R}^e$ .

We proved in previous work [1] that

**Theorem 3.** Let  $A^+$  denote the Moore–Penrose pseudoinverse of a matrix  $A$ , and  $L(\mathbf{G}) = BB^T$  denote the graph Laplacian of a connected multigraph  $\mathbf{G}$ . If  $x$  is a vertex vector for the tcrw with underlying graph  $\mathbf{G}$ , then the  $x^k \in \mathbb{R}^v$  are sampled independently from multivariate normals with mean 0 and covariance matrix  $L(\mathbf{G})^+$ .

In this paper, we will study the distribution of squared distances  $\|x_i - x_j\|^2$  between vertices in tcrows. We will first note that the expectation  $\mathcal{E}(\|x_i - x_j\|^2; \mathbf{G})$  is equal to the resistance distance [11, 12, 14] between  $v_i$  and  $v_j$  in  $\mathbf{G}$  and show that the expectation  $\mathcal{E}(R_g^2; \mathbf{G})$  of the squared radius of gyration of a tcrw is given in terms of the Kirchhoff index (or quasi-Wiener index) of  $\mathbf{G}$ .

We then recall the *contraction factor* or *g-factor* of a polymer, which is defined to be the ratio of the (expected) radius of gyration of the polymer to the expected radius of gyration of a linear polymer of the same length embedded in a space of the same dimension. Our main theorem concerns the asymptotic behavior of the contraction factor as we subdivide edges of the graph. This corresponds to the thermodynamic limit in statistical mechanics as the system size goes to infinity. In particular, we show that

**Theorem 4.** *For any connected multigraph  $\mathbf{G}$  (including loop and multiple edges), if  $\mathbf{G}_n$  is the graph created by dividing each edge of  $\mathbf{G}$  into  $n$  pieces, then the contraction factor obeys*

$$\lim_{n \rightarrow \infty} g(\mathbf{G}_n) = \frac{3}{e(\mathbf{G})^2} \left( \text{tr } \mathcal{L}^+(\mathbf{G}) + \frac{1}{3} \text{Loops}(\mathbf{G}) - \frac{1}{6} \right).$$

Here  $\text{Loops}(\mathbf{G}) = e(\mathbf{G}) - v(\mathbf{G}) + 1$  is the cycle rank of  $\mathbf{G}$ ,  $\mathcal{L}(\mathbf{G})$  is the normalized graph Laplacian of  $\mathbf{G}$ , and  $\mathcal{L}^+(\mathbf{G})$  is the Moore–Penrose pseudoinverse of  $\mathcal{L}(\mathbf{G})$ .

Along the way, we will prove some independently interesting results about resistances measured between points along edges in a resistor network and about the degree-Kirchhoff index of a graph. We define the normalized Laplacian below.

We will then use this theorem to analyze the relative sizes of the topological polymers in solution synthesized by Tezuka and compare the scaling coefficients with a molecular dynamics calculation of expected radii of gyration. We will see that there is an excellent linear fit between our coefficients and the results of the simulation.

The asymptotic contraction factor should be one of the most fundamental physical or dynamical quantities for topological polymers in solution. In fact, one can use this quantity to estimate various physical quantities of topological polymers in solution such as the viscosity coefficient. The asymptotic contraction factor should also be relevant to experiments on the viscosity of polymer solutions [16].

## 2. THE MULTIGRAPH LAPLACIAN AND RESISTOR NETWORKS

We commented above that the graph  $L = BB^T$  was known as the graph Laplacian of  $\mathbf{G}$ . We now expand on this point, summarizing some of the widely developed and rich theory of graph Laplacians. First, we observe that  $L$  is well-defined for graphs  $\mathbf{G}$  with multiple edges and loop edges because  $B$  is defined for such graphs. For multigraphs, the *degree*  $\text{deg } v_i$  of a vertex  $i$  counts the number of edges leaving or arriving at  $v_i$  (loop edges count twice), the degree matrix  $D$  is the diagonal matrix of vertex degrees and the adjacent matrix  $A$  is defined by  $-A_{ij} = \#$  one edge paths from  $v_i$  to  $v_j$ . Every loop edge  $v_i \rightarrow v_i$  provides two different paths from  $v_i$  to  $v_i$ , and so contributes 2 to the diagonal matrix.

An easy computation shows  $L = D - A$ , or

$$L_{ij} = \begin{cases} \deg v_i - 2(\# \text{ edges } v_i \rightarrow v_i \text{ in } \mathbf{G}) & \text{if } i = j, \\ -\# \text{ edges } (v_j \rightarrow v_i \text{ or } v_i \rightarrow v_j) \in \mathbf{G} & \text{if } i \neq j. \end{cases}$$

For functions  $p: \mathbf{v}(\mathbf{G}) \rightarrow \mathbb{R}$  defined on the vertices of a multigraph  $\mathbf{G}$  viewed as vectors in  $\mathbb{R}^{\mathbf{v}}$  we can interpret the Laplacian as a linear operator. Explicitly

$$Lp(v_i) = (\deg v_i) p(v_i) - \sum_{v_j \text{ adjacent to } v_i} p(v_j) \quad (1)$$

where  $v_j$  is *adjacent* to  $v_i$  in  $\mathbf{G}$  if there is an edge  $v_i \rightarrow v_j$  or  $v_j \rightarrow v_i$  in  $\mathbf{G}$ . Note that for each loop edge  $v_i \rightarrow v_i$  we increment  $\deg v_i$  by 2 *and* count  $v_i$  as adjacent to itself in two ways. These contributions cancel in (1).

For functions  $p: \mathbf{v}(\mathbf{G}) \rightarrow \mathbb{R}$  defined on the vertices of a graph  $\mathbf{G}$  viewed as vectors in  $\mathbb{R}^{\mathbf{v}}$  we can interpret  $B^T$  as the *gradient* operator. If  $e_k = v_i \rightarrow v_j$ ,

$$\text{grad } p(e_k) = p(v_j) - p(v_i) = (B^T p)_k.$$

For functions  $u: \mathbf{e}(\mathbf{G}) \rightarrow \mathbb{R}$  defined on the edges of  $\mathbf{G}$  and viewed as vectors in  $\mathbb{R}^{\mathbf{e}}$ , we can interpret  $B$  as the negative of the *divergence* operator

$$\text{div } u(v_i) = \sum_{e_j = v_i \rightarrow \cdot} u(e_j) - \sum_{e_j = \cdot \rightarrow v_i} u(e_j) = (-Bu)_i.$$

This means that the graph Laplacian  $L = -\text{grad div}$ . This agrees with the sign convention in Riemannian geometry [4], but is opposite from the sign convention in mathematical physics [8].

If the graph  $\mathbf{G}$  is a conductive network, then a function  $p: \mathbf{v} \rightarrow \mathbb{R}$  is called a *potential*. Along every edge of  $\mathbf{G}$ , the difference in potential  $\text{grad } p = U$  is the corresponding *voltage*. Ohm's law states that  $U = IR$  where  $I$  is the current and  $R$  the resistance, so if each edge has unit resistance,  $I = U$ . Kirchhoff's first law says that the total current flowing into and out of each vertex are equal, or that  $\text{div } I = 0$ . Combining these, we see that  $Lp = 0$  in an isolated system.

If  $\mathbf{G}$  is connected, the kernel of  $L$  is known to be one-dimensional and spanned by the constant functions. This means that in an isolated system,  $p$  is a constant and no current flows. However, if we introduce unit current at  $v_i$  and remove unit current at  $v_j$ , then the potential is the minimum-norm solution  $p^{ij}$  to the Poisson problem

$$(Lp^{ij})(v_k) = e_i - e_j = \begin{cases} +1, & \text{if } k = i, \\ -1, & \text{if } k = j. \end{cases} \quad (2)$$

Note that the right hand side is in  $(\ker L)^\perp = \text{im } L^T = \text{im } L$ , so  $p^{ij}$  is an exact solution and not a least-squares solution. The resistance  $r_{ij}$  between  $v_i$  and  $v_j$  is then

$$r_{ij} := p^{ij}(v_i) - p^{ij}(v_j). \quad (3)$$

In graph theory [12], this value is called the *resistance distance* from  $v_i$  to  $v_j$  in  $\mathbf{G}$ .

To analyze the resistance distance, we will need to recall the theory of the Moore–Penrose pseudoinverse of a matrix  $A$ . First, if  $A$  has singular value decomposition  $A = U\Sigma V^T$  with singular values  $\sigma_i$  then  $A^+ := V\Sigma^+U^T$ , where  $\Sigma^+$  is a diagonal matrix with entries

$$\sigma_i' = \begin{cases} \frac{1}{\sigma_i}, & \text{if } \sigma_i \neq 0, \\ 0, & \text{if } \sigma_i = 0. \end{cases}$$

The pseudoinverse has many useful properties, but the one that we will use the most often is the fact that  $x = A^+b$  is the unique minimum-norm least-squares solution to the linear system  $Ax = b$ . Explicitly, this means that  $x \in \ker A^\perp$  and  $Ax$  is the orthogonal projection of  $b$  to  $\text{im } A$ . It is also true that  $A^+A$  is orthogonal projection to  $\ker A^\perp$  and  $AA^+$  is orthogonal projection to  $\text{im } A$ .

We note that this property of the pseudoinverse implies that minimum-norm least-squares solutions to linear systems are linear in their right-hand sides. If  $x$  and  $y$  solve  $Ax = b$  and  $Ay = c$  in this sense, then  $x = A^+b$  and  $y = A^+c$ . But if  $w$  is the minimum-norm least squares solution to  $Aw = b + c$ , then  $w \in (\ker A)^\perp$  so applying  $A^+$  to both sides we get

$$w = A^+Aw = A^+(b + c) = A^+b + A^+c = x + y. \quad (4)$$

In differential equations, this property is usually called “superposition of solutions”.

We now define the potential function  $p^i$  to be the unique minimum-norm least-squares solution to the Poisson problem  $Lp^i = e_i$ . Since  $e_i$  is not in  $\text{im } L$ , the solution  $p^i$  exactly solves the slightly different Poisson problem

$$Lp^i(w) = \begin{cases} 1 - \frac{1}{v}, & \text{if } w = v_i \\ -\frac{1}{v}, & \text{otherwise.} \end{cases} \quad (5)$$

We can now prove

**Proposition 5.** *If  $L$  is the graph Laplacian of a connected multigraph  $\mathbf{G}$ , the resistance distance  $r_{ij}$  between vertices  $v_i$  and  $v_j$ , the potential functions  $p^{ij}$ ,  $p^i$ , and  $p^j$  and the matrix  $L^+$  are related in the following ways:*

$$p^{ij} = L^+(e_i - e_j), \quad p^i = L^+e_i, \quad p^j = L^+e_j, \quad p^{ij} = p^i - p^j,$$

while

$$\begin{aligned} r_{ij} &= p^{ij}(v_i) - p^{ij}(v_j) = p^i(v_i) - p^j(v_i) - p^i(v_j) + p^j(v_j) \\ &= (L^+)_{ii} - (L^+)_{ij} - (L^+)_{ji} + (L^+)_{jj}. \end{aligned} \quad (6)$$

Further,  $L^+Lp^{ij} = p^{ij}$  while  $L^+Lp^i = p^i$ .

*Proof.* Since  $p^{ij}$  and  $p^i$  are defined to be minimum-norm least squares solutions to Poisson problems, they are given by  $p^{ij} = L^+(e_i - e_j)$  and  $p^i = L^+e_i$ . The linearity property (4) then implies that  $p^{ij} = p^i - p^j$ . Then (6) follows from (3) since  $p^i = L^+e_i$  is the  $i$ th row of  $L^+$ , so its values at  $v_i$  and  $v_j$  are the matrix elements  $(L^+)_{ii}$  and  $(L^+)_{ij}$ . Finally, since  $p^{ij}$  and  $p^i$  are *minimum-norm* solutions to Poisson problems, they are both in  $(\ker L)^\perp$  and hence fixed by  $L^+L$  (the orthogonal projection onto  $(\ker L)^\perp$ ).  $\square$

The sum of all the resistance distances between vertices in a graph is known as the Kirchhoff index (or quasi-Wiener index) of the graph and denoted

$$\text{Kf}(\mathbf{G}) = \sum_{i < j} r_{ij}.$$

The Kirchhoff index is a measure of the connectivity of the graph— graphs which are very well connected with short paths between vertices have small Kirchhoff indices while graphs with longer paths between vertices have larger ones. There is a well-developed theory of the Kirchhoff index in mathematical chemistry.

### 3. EXPECTED SQUARED CHORDLENGTHS AND RESISTANCES

From Theorem 3, it follows immediately that the marginal distribution of the displacement vector  $x_i - x_j$  between vertices  $v_i$  and  $v_j$  is Gaussian with mean zero. The variance of the Gaussian can be computed by taking

$$\mathcal{E} \left( (x_i^k - x_j^k)^2; \mathbf{G} \right) = \mathcal{E} \left( (x_i^k)^2 + (x_j^k)^2 - 2(x_i^k x_j^k); \mathbf{G} \right) = L_{ii}^+ + L_{jj}^+ - L_{ij}^+ - L_{ji}^+$$

since  $L^+$  is the covariance matrix of the  $x_i^k$ . This proves

**Proposition 6.** *Let  $v_i$  and  $v_j$  be vertices in a multigraph  $\mathbf{G}$  and  $r_{ij}$  be the resistance distance between them. The expected squared distance between  $v_i$  and  $v_j$  in a tcrrw with multigraph  $\mathbf{G}$  in  $\mathbb{R}^d$  is given by  $\mathcal{E} \left( \|x_i - x_j\|^2; \mathbf{G} \right) = d(L_{ii}^+ + L_{jj}^+ - L_{ij}^+ - L_{ji}^+) = d r_{ij}$ .*

We now give two examples. It is well known that in a Gaussian random polygon (ring polymer) in  $\mathbb{R}^3$  with  $n$  vertices, the expected squared chordlength between vertices  $v_1$  and  $v_j$  separated by  $j$  edges is given by  $3j(n-j)/n$  [26]. We may recompute this result rather simply using resistances.

There are two paths from  $v_1$  to  $v_j$ , one composed of  $j$  resistors in series and the other composed of  $n-j$  resistors in series. The two paths are in parallel, so the total resistance is

$$r_{1j} = \frac{1}{\frac{1}{j} + \frac{1}{n-j}} = \frac{j(n-j)}{n}. \quad (7)$$

Extending this idea, suppose we have a “multitheta” graph consisting of  $m$  arcs of  $n$  edges joining two vertices. The total resistance of each arc is  $n$ , and  $m$  such arcs are in parallel, so the resistance between junctions is

$$\frac{1}{\frac{1}{n} + \dots + \frac{1}{n}} = \frac{1}{\frac{m}{n}} = \frac{n}{m}.$$

This recovers an asymptotic result of Deguchi and Uehara [21]. It also shows the interesting phenomenon that the junctions are expected to be closer together when more arcs join them, even if the length of the arcs remains constant.

#### 4. THE RADIUS OF GYRATION

A standard measure of the effective size of a polymer in solution is the radius of gyration.

**Definition 7.** The radius of gyration of a tcrw with vertex vector  $x$  in  $\mathbb{R}^d$  is given by

$$R_g^2(x) = \frac{1}{2\mathbf{v}^2} \sum_{i,j \in 1}^{\mathbf{v}} \|x_i - x_j\|^2.$$

The expected radius of gyration of a tcrw with underlying graph  $\mathbf{G}$  is given by the expectation of  $R_g^2(x)$  when each coordinate vector  $x^k$  is chosen according to a multivariate Gaussian with mean zero and covariance matrix  $L^+(\mathbf{G})$ . In polymer science, this is the mean-square radius of gyration  $\langle s^2 \rangle$  of the molecule (cf. [17, 1.17]); we denote it by  $\mathcal{E}(R_g^2; \mathbf{G})$ .

We can use our result connecting resistance distances with expectations of chordlengths to compute the expected radius of gyration rather simply, recovering a classical formula for the radius of gyration of a molecule of arbitrary topology [9, eq. 18a]:

**Theorem 8.** For any tcrw in  $\mathbb{R}^d$  with multigraph  $\mathbf{G}$ , we have

$$\mathcal{E}(R_g^2; \mathbf{G}) = \frac{d}{\mathbf{v}^2} \text{Kf}(\mathbf{G}) = \frac{d}{\mathbf{v}} \text{tr } L^+ = \frac{d}{\mathbf{v}} \sum_{i=1}^{\mathbf{v}-1} \frac{1}{\lambda_i}$$

where  $\lambda_i$  are the nonzero eigenvalues of the graph Laplacian  $L(\mathbf{G})$ .

*Proof.* Substituting the result of Proposition 6 into Definition 7, we get

$$R_g^2 = \frac{d}{2\mathbf{v}^2} \sum_{i,j} (L^+)_{ii} + (L^+)_{jj} - 2(L^+)_{ij} = \frac{d}{2\mathbf{v}^2} \left( 2\mathbf{v} \text{tr } L^+ - 2 \sum_{i,j} (L^+)_{ij} \right).$$

To analyze the last term on the right, we note that  $\ker L = \ker L^+$ , so in particular the constant vector of 1's is in  $\ker L^+$ . This means that the row sums of  $L^+$  are zero, and hence that the “grand sum”  $\sum_{i,j} (L^+)_{ij}$  is zero as well. The statement on eigenvalues is simply that the (nonzero) eigenvalues of  $L^+$  are the reciprocals of those of  $L$ .  $\square$

We note that the connection between Kirchhoff index and the trace of  $L^+$  is well-known [11], though it is not always stated clearly whether the result is intended to apply to multigraphs  $\mathbf{G}$ .

We now give two examples for Theorem 8. First, consider the path graph with  $\mathbf{v}$  vertices  $v_1, \dots, v_{\mathbf{v}}$ . The resistance between vertex  $v_i$  and  $v_j$  is simply  $|i - j|$ . Summing, we get

$$\mathcal{E}(R_g^2; \text{path graph}) = \frac{d}{6} \frac{\mathbf{v}(\mathbf{v} + 1)}{\mathbf{v} - 1} \quad (8)$$

which agrees with the standard asymptotic expression of  $\frac{d\mathbf{v}}{6}$ .

The Gaussian ring polymer is based on the cycle graph with  $\mathbf{v}$  vertices. The graph Laplacian for this graph can be written

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

which is a circulant matrix (every row is a cyclic shift of the previous one). The eigenvalues of a circulant matrix are easy to compute; in this case they are

$$\lambda_j = 2 - 2 \cos \left( \frac{2\pi j}{\mathbf{v}} \right) = 4 \sin^2 \left( \frac{\pi j}{\mathbf{v}} \right)$$

for  $j \in 0, \dots, \mathbf{v} - 1$ . Only  $\lambda_0 = 0$ , so we have to sum  $\sum_{j=1}^{\mathbf{v}-1} \csc^2 \left( \frac{\pi j}{\mathbf{v}} \right) = \frac{1}{3}(\mathbf{v}^2 - 1)$  [15, 4.4.6.5, p. 644]. Zhang et al. [25] compute this along with formulae for the Kirchhoff index of other circulant graphs. This means that the expected radius of gyration is (exactly)

$$\mathcal{E}(R_g^2; \text{cycle graph}) = \frac{d}{12} \frac{\mathbf{v}^2 - 1}{\mathbf{v}},$$

which agrees with both the standard asymptotic approximation  $\frac{d\mathbf{v}}{12}$  [3, 13] and the result of summing (7) over all pairs of vertices in the cycle and dividing by  $\mathbf{v}^2$ .

For more general graphs, there are a number of useful ways to approach the numerical computation of the Kirchhoff index. Finding the eigenvalues of  $L$  is certainly the most straightforward. The fastest one in practice seems to be to calculate in terms of the last two coefficients of the characteristic polynomial of the graph Laplacian  $L$ , which is a sparse matrix.



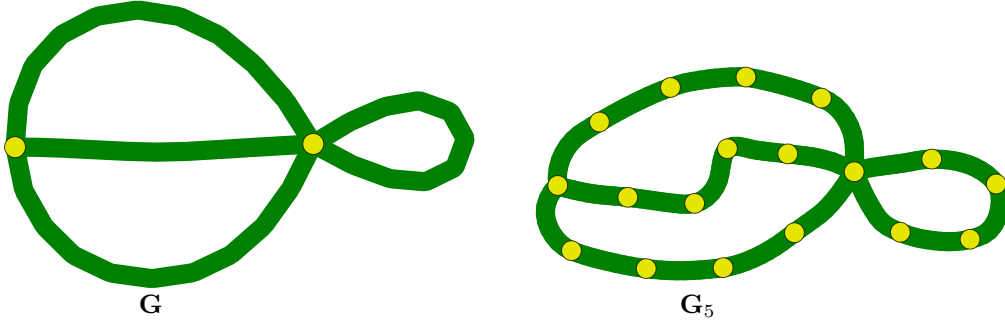


FIG. 2: The  $n$ -part edge subdivision divides each edge of  $\mathbf{G}$  into  $n$  pieces. On the left, we see a multigraph and on the right its 5-part edge subdivision.

**Proposition 9.** [14] If  $p(t) = \det(L - tI) = t^{\nu} + c_1 t^{\nu-1} + \dots + c_{\nu-1} t + c_{\nu}$  is the characteristic polynomial of the graph Laplacian  $L(\mathbf{G})$ , then  $\mathcal{E}(R_g^2; \mathbf{G}) = -\frac{1}{\nu} \frac{c_{\nu-2}}{c_{\nu-1}}$ .

We note that Proposition 9 has the perhaps surprising corollary:

**Corollary 10.** The expected radius of gyration  $\mathcal{E}(R_g^2)$  of any Gaussian TCRW is a rational number.

## 5. SUBDIVISION GRAPHS AND TOPOLOGICAL POLYMERS

The topological polymers so far synthesized in the laboratory have long chains of monomers joining a relatively small number of “junction” molecules. Thus, while the number of edges in their graph may be large, the relative complexity of the graphs is rather modest. Further, since each edge represents a persistence length of the chain of monomers, it is not always clear exactly how many edges should be in each path between junctions.

Therefore, we now study the behavior of the expected radius of gyration for a particular kind of subdivision graph as the number of subdivisions increases. Two cautions are in order: the usual definition of a subdivision graph places one vertex in the middle of each edge, so iterated subdivisions divide the original edges into a power of two subdivisions [23]. This is different from our model, where any number of subdivisions are allowed. Second, while previous authors subdivided simple graphs, we explicitly allow loop edges and multiple edges.

**Definition 11.** The  $n$ -part edge subdivision  $\mathbf{G}_n$  of a multigraph  $\mathbf{G}$  is the graph obtained by dividing each edge of  $\mathbf{G}$  into  $n$  pieces.

Figure 2 shows an example of the 5-fold edge subdivision of a multigraph.

We have already analyzed the expected radius of gyration of two  $n$ -part edge subdivisions of a graph— if  $\mathbf{G}$  has two vertices and one edge joining them, the path graph with  $n$  edges is  $\mathbf{G}_n$  and

$$\mathcal{E}(R_g^2; \mathbf{G}_n = \text{path graph}) = \frac{d}{6} \frac{(n+1)(n+2)}{n}. \quad (9)$$

If  $\mathbf{G}$  has one vertex and one loop edge, the cycle graph with  $n$  edges is  $\mathbf{G}_n$ . We saw that

$$\mathcal{E}(R_g^2; \mathbf{G}_n = \text{cycle graph}) = \frac{d}{12} \frac{n^2 - 1}{n}. \quad (10)$$

Notice that both of these are asymptotically linear in  $n$ . We will show that this is a general feature of edge-subdivision graphs, and show how to compute the leading coefficient.

To do so, we need to recall a definition:

**Definition 12.** The *normalized graph Laplacian*  $\mathcal{L}(\mathbf{G})$  is given by

$$\mathcal{L}_{ij} = \begin{cases} 1 - \frac{2 \times \# \text{ loop edges}}{\deg(v_i)}, & \text{if } i = j, \\ -\frac{k}{\sqrt{\deg(v_i) \deg(v_j)}}, & \text{if } v_i, v_j \text{ joined by } k \text{ edges,} \\ 0, & \text{otherwise.} \end{cases}$$

We can now state our next main result.

**Theorem 13.** *For any connected multigraph  $\mathbf{G}$  (including loop and multiple edges), if  $\mathbf{G}_n$  is the  $n$ -part edge subdivision of  $\mathbf{G}$ , then the expected radius of gyration of a tcrw with underlying graph  $\mathbf{G}_n$  embedded in  $\mathbb{R}^d$  obeys*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{E}(R_g^2; \mathbf{G}_n) = \frac{d}{2\mathbf{e}(\mathbf{G})} \left( \text{Tr } \mathcal{L}^+(\mathbf{G}) + \frac{1}{3} \text{Loops}(\mathbf{G}) - \frac{1}{6} \right).$$

Here  $\text{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$  is the cycle rank of  $\mathbf{G}$ .

For the graph  $\mathbf{G}$  with two vertices and one edge,  $\mathcal{L}$  is the  $2 \times 2$  matrix  $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix}$ . A computation reveals that  $\mathcal{L}^+ = \frac{1}{4}\mathcal{L}$  and  $\text{tr } \mathcal{L}^+ = \frac{1}{2}$ . Thus Theorem 13 recovers the coefficient  $\frac{d}{6}$  from (9). Similarly, for the graph  $\mathbf{G}$  with one vertex and one loop edge,  $\mathcal{L}$  is the  $1 \times 1$  matrix whose single entry is zero, so  $\text{tr } \mathcal{L}^+ = 0$ . Thus Theorem 13 also recovers the coefficient  $\frac{d}{12}$  from (10).

We will now develop several tools leading up to the proof of Theorem 13. Our basic idea is to calculate resistance distances on the subdivided graph  $\mathbf{G}_n$  by constructing solutions of the Poisson problem (2) using superposition of solutions. In this way, we'll be able to relate resistances on  $\mathbf{G}_n$  to resistances on  $\mathbf{G}$ , and so compute the Kirchhoff index of  $\mathbf{G}_n$  by summing these formulae. This approach was inspired by the Green's kernel method of Carmona, Mitjana, and Monsó [2].

### 5.1. Poisson problems on subdivided graphs

For this section, we establish the notation that we will refer to vertices along a subdivided edge in  $\mathbf{G}_n$  as  $v_0, \dots, v_n$ . The end vertices  $v_0$  and  $v_n$  correspond to vertices in  $\mathbf{G}$  which we will refer to as  $a$  and  $b$ . It is possible that  $a = b$  if the original edge in  $\mathbf{G}$  was a loop edge.

We first note that functions on subdivided edges are linear when their Laplacian vanishes.

**Lemma 14.** *If a vertex  $v_i$  is joined to two other vertices  $v_{i-1}$  and  $v_{i+1}$  then  $Lp(v_i) = 0$  if and only if  $p$  is a linear function of  $i$  on  $v_{i-1}, v_i, v_{i+1}$ .*

*Proof.* Using (1), the Laplacian at  $v_i$  is given by

$$Lp(v_i) = -p(v_{i-1}) + 2p(v_i) - p(v_{i+1}).$$

Rearranging the right hand side, we see  $Lp(v_i) = 0$  if and only if  $p(v_i) - p(v_{i-1}) = p(v_{i+1}) - p(v_i)$ , which is exactly the condition for  $p$  to be a linear function on these points.  $\square$

We can now define and solve a particular Poisson problem on a subdivided edge

**Proposition 15.** *Suppose  $a \rightarrow b$  is an edge of a connected multigraph  $\mathbf{G}$  subdivided in  $\mathbf{G}_n$  by  $v_0, \dots, v_n$ . The Poisson problem*

$$Lp^{ajb}(w) = \begin{cases} -\frac{n-j}{n}, & \text{if } w = a \neq b \\ -\frac{j}{n}, & \text{if } w = b \neq a \\ -1, & \text{if } w = a = b \\ +1, & \text{if } w = v_j \\ 0, & \text{otherwise} \end{cases} \quad (11)$$

*has minimum norm solution*

$$p^{ajb}(w) = C + \begin{cases} \frac{n-j}{n}k, & \text{if } w = v_k \text{ and } k \leq j \\ -\frac{j}{n}k + j, & \text{if } w = v_k \text{ and } k \geq j \\ 0, & \text{otherwise, including } w = a \text{ and } w = b \end{cases} \quad (12)$$

*for some  $C$ . Since  $Lp^{ajb}$  is in  $\text{im } L$ ,  $L^+Lp^{ajb} = p^{ajb}$ .*

*Proof.* The function  $p^{ajb}$  is piecewise linear with corners at  $a, b$  and  $v_j$ . This proves that  $Lp^{ajb} = 0$  away from these three vertices by Lemma 14. The Laplacian may be checked directly at those points using (1). Since all solutions of Poisson problems on a connected graph  $\mathbf{G}$  differ by a constant there must be some  $C$  which yields the minimum norm solution.  $\square$

Recalling that  $p^{ij}$  is our notation for a solution to (2), we can now prove a general decomposition formula for solutions of (2) on a subdivided graph  $\mathbf{G}_n$ .

**Proposition 16.** *Suppose that  $a \rightarrow b$  and  $c \rightarrow d$  are any edges in  $\mathbf{G}$  (they may be loop edges, connect the same endpoints, or even be the same edge). Suppose that  $q$  is  $j$  edges from  $a$  along the subdivision of  $a \rightarrow b$  in  $\mathbf{G}_n$  and  $r$  is  $k$  edges from  $c$  along the subdivision of  $c \rightarrow d$  in  $\mathbf{G}_n$ .*

Then as functions on  $\mathbf{G}_n$ , we have the following relation between minimum norm least-squares solutions to the three Poisson problems (2), (5) and (11):

$$p^{qr} = p^{ajb} + \frac{n-j}{n}p^a + \frac{j}{n}p^b - \frac{n-k}{n}p^c - \frac{k}{n}p^d - p^{ckd}$$

where we use the convention  $p^{aa} = 0$  and note that  $p^{a0b} = p^{anb} = 0$  in case there are coincidences among  $a, b, c, d, q$  and  $r$ .

*Proof.* We first observe that

$$Lp^q = Lp^{ajb} + \frac{n-j}{n}Lp^a + \frac{j}{n}Lp^b.$$

Both sides are equal to  $-\frac{1}{v}$  away from  $q, a$ , and  $b$ . At  $q, a$ , and  $b$  we can use the definitions of the functions to check the equality directly. Equality still holds if  $a = b, q = a$  (that is,  $j = 0$ ) or  $q = b$  (that is,  $j = n$ ). Applying  $L^+$  to both sides of the equation, we have shown

$$p^q = p^{ajb} + \frac{n-j}{n}p^a + \frac{j}{n}p^b.$$

regardless of any possible coincidences between  $a, b$  and  $q$ .

Exactly the same argument shows

$$p^r = p^{ckd} + \frac{n-k}{n}p^c + \frac{k}{n}p^d$$

regardless of coincidences among  $c, d$  and  $r$ .

Since  $p^{qr} = p^q - p^r$  this proves the result. We note that this last equality holds regardless of any coincidences between  $\{a, b, q\}$  and  $\{c, d, r\}$ .  $\square$

## 5.2. Relating resistances between middle points and endpoints of subdivided edges

We now use the decomposition formula of Proposition 16 to compute the resistance between vertices in the middle of subdivided edges in terms of resistances between their endpoints.

**Proposition 17.** *Suppose that  $a \rightarrow b$  and  $c \rightarrow d$  are any edges in  $\mathbf{G}$  (they may be loop edges, connect the same endpoints, or even be the same edge). Suppose that  $q$  is  $j$  edges from  $a$  along the subdivision of  $a \rightarrow b$  in  $\mathbf{G}_n$  and  $r$  is  $k$  edges from  $c$  along the subdivision of  $c \rightarrow d$  in  $\mathbf{G}_n$ . Further, let us define “interpolation coefficients” associated to  $a, b, c$  and  $d$  by*

$$\mu_a = \frac{n-j}{n}, \quad \mu_b = \frac{j}{n}, \quad \mu_c = -\frac{n-k}{n}, \quad \mu_d = -\frac{k}{n}.$$

Further, to simplify notation, let us write  $\mathcal{S} = \{a, b, c, d\}$ . We may express the resistance  $r_{qr}$  in terms of the resistances among  $a, b, c$  and  $d$  as follows. If  $a \rightarrow b$  and  $c \rightarrow d$  are different edges,

$$r_{qr} = \frac{j(n-j) + k(n-k)}{n} - \frac{1}{2} \sum_{v,w \in \mathcal{S}} \mu_v \mu_w r_{vw} \quad (13)$$

while if  $a \rightarrow b = c \rightarrow d$ , we modify the right-hand side of (13) by adding

$$-2 \frac{\min(j, k)(n - \max(j, k))}{n}.$$

We note that this is a generalization of Theorem 4.1 in [2] and Example 6 in [5] in two ways: our result applies to multigraphs with multiple and loop edges and also treats the cases of subdivisions with any number of inserted vertices.

We can use Proposition 17 – in particular (13) – to evaluate various statistical physical quantities for topological polymers in solution such as the hydrodynamic radius.

*Proof.* We first observe that  $\sum_{v \in \mathcal{S}} \mu_v = 0$  (regardless of the values of  $j$  and  $k$ ). We will use this observation regularly below. Next, let us define a new function  $p^\mu := \sum_{v \in \mathcal{S}} \mu_v p^v$ . If  $v \in \mathcal{S}$  and  $w \notin \mathcal{S}$ , then  $Lp^v(w) = -\frac{1}{v}$ . Thus, if  $w \notin \mathcal{S}$ ,  $(Lp^\mu)(w) = -\frac{1}{v} \sum_{v \in \mathcal{S}} \mu_v = 0$ .

This means that Lemma 14 tells us that  $p^\mu$  is linear on  $a \rightarrow b$  and  $c \rightarrow d$ . Thus we can evaluate  $p^\mu$  at  $q$  by linearly interpolating values at  $a$  and  $b$  (and likewise for  $r, c$  and  $d$ ). In particular, keeping track of the signs in  $\mu_c$  and  $\mu_d$ ,

$$p^\mu(q) - p^\mu(r) = \sum_{w \in \mathcal{S}} \mu_w p^\mu(w). \quad (14)$$

Now (3) tells us that  $r_{qr} = p^{qr}(q) - p^{qr}(r)$ . Further, using the definition of  $p^\mu$ , Proposition 16 can be rephrased as  $p^{qr} = p^{ajb} - p^{ckd} + p^\mu$ . Using the definition of  $p^\mu$  and (14), we then have

$$r_{qr} = (p^{ajb} - p^{ckd})(q) - (p^{ajb} - p^{ckd})(r) + \sum_{v,w \in \mathcal{S}} \mu_v \mu_w p^v(w).$$

We now consider the right hand side. Using (3) we can expand

$$\sum_{v,w \in \mathcal{S}} \mu_v \mu_w r_{vw} = \sum_{v,w \in \mathcal{S}} \mu_v \mu_w (p^v(v) + p^w(w) - p^v(w) - p^w(v)) = -2 \sum_{v,w \in \mathcal{S}} \mu_v \mu_w p^v(w)$$

where we used  $\sum_{v \in \mathcal{S}} \mu_v = 0$  in the second equality. Now assume that  $a \rightarrow b$  and  $c \rightarrow d$  are different. It remains only to prove that

$$(p^{ajb} - p^{ckd})(q) - (p^{ajb} - p^{ckd})(r) = \frac{j(n-j) + k(n-k)}{n}. \quad (15)$$

Using (12) in Proposition 15, we can compute that

$$p^{ajb}(q) - p^{ajb}(r) = \frac{j(n-j)}{n}$$

since  $r$  is not on  $a \rightarrow b$  so  $p^{ajb}(r) = C$ . This and the corresponding statement for  $p^{ckd}(r) - p^{ckd}(q)$  establish (15). If  $a \rightarrow b = c \rightarrow d$  the situation is slightly more vexing, since  $r$  is on  $a \rightarrow b$ . In this case, we see that

$$p^{ajb}(r) = C + \begin{cases} \frac{k(n-j)}{n}, & \text{if } k \leq j \\ \frac{j(n-k)}{n}, & \text{if } k \geq j \end{cases} = C + \frac{\min(j, k)(n - \max(j, k))}{n}.$$

□

### 5.3. Computing the Kirchhoff index of $\mathbf{G}_n$

We are now interested in computing  $\text{Kf}(\mathbf{G}_n)$ . It will be easier to compute  $2\text{Kf}(\mathbf{G}_n) = \sum_{v,w \in \mathbf{G}_n} r_{vw}$ . For convenience, we define

**Definition 18.** We call a vertex  $v \in \mathbf{G}_n$  an *interior* vertex if it is one of the vertices  $v_1, \dots, v_{n-1}$  along the subdivision in  $\mathbf{G}_n$  of an edge  $a \rightarrow b$  in  $\mathbf{G}$  (that is, it is not  $v_0 = a$  or  $v_n = b$ ). The set of interior vertices will be called  $\mathcal{I}$ . The interior vertices along  $a \rightarrow b$  will be  $\mathcal{I}(a \rightarrow b)$ . We call vertices that are not interior *boundary* vertices, and note that they are also vertices of  $\mathbf{G}$ .

It is now clear that

$$2\text{Kf}(\mathbf{G}_n) = \sum_{q,r \in \mathbf{G}_n} r_{qr} = \sum_{q,r \in \mathcal{I}} r_{qr} + 2 \sum_{\substack{q \in \mathcal{I} \\ r \in \mathcal{B}}} r_{qr} + \sum_{q,r \in \mathcal{B}} r_{qr}. \quad (16)$$

We are ready to prove the heart of Theorem 13:

**Proposition 19.** *If  $\mathbf{G}$  is a connected multigraph and  $\mathbf{G}_n$  is the graph obtained by subdividing each edge of  $\mathbf{G}$  into  $n$  pieces, then*

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \text{Kf}(\mathbf{G}_n) = \frac{2\text{Loops}(\mathbf{G}) - 1}{12} \mathbf{e}(\mathbf{G}) + \frac{1}{4} \text{Kf}^*(\mathbf{G})$$

where  $\text{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$  is the cycle rank of  $\mathbf{G}$  and  $\text{Kf}^*(\mathbf{G})$  is the “degree-Kirchhoff index” introduced by [6]:

$$\text{Kf}^*(\mathbf{G}) = \sum_{v_i < v_j \in \mathbf{G}} (\deg_{\mathbf{G}} v_i \deg_{\mathbf{G}} v_j) r_{v_i v_j}.$$

We note that this result is compatible with the result of iteratively applying the subdivision operator given by Theorem 3.4 of [24], but that it doesn't follow from it—iterating the usual subdivision operator can only give  $2^k$ -fold edge subdivisions of  $\mathbf{G}$ , and we don't know that  $\lim_{n \rightarrow \infty} \frac{1}{n^3} \text{Kf}(\mathbf{G}_n) = \lim_{k \rightarrow \infty} \frac{1}{(2^k)^3} \text{Kf}(\mathbf{G}_{2^k})$  until we know that both limits exist. Further, [24] doesn't identify the number of loops as part of the result and depends on a result of [5] which is only proved for graphs without multiple edges.

*Proof.* We will repeatedly use the fact that if  $q$  and  $r$  are in  $\mathbf{G}$  and  $\mathbf{G}_n$ , then  $r_{qr}^{\mathbf{G}_n} = n r_{qr}^{\mathbf{G}}$  because (from the point of view of  $q$  and  $r$ ) we can regard  $\mathbf{G}_n$  as the electrical network obtained by replacing each edge of  $\mathbf{G}$  with a resistor of resistance  $n$ .

Using Proposition 17 we see that if  $a \rightarrow b$  and  $c \rightarrow d$  are different edges of  $\mathbf{G}$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{\substack{q \in \mathcal{I}(a \rightarrow b) \\ r \in \mathcal{I}(c \rightarrow d)}} r_{qr}^{\mathbf{G}_n} = \frac{1}{3} - \frac{1}{6}(r_{ab}^{\mathbf{G}} + r_{cd}^{\mathbf{G}}) + \frac{1}{4}(r_{ac}^{\mathbf{G}} + r_{ad}^{\mathbf{G}} + r_{bc}^{\mathbf{G}} + r_{bd}^{\mathbf{G}}). \quad (17)$$

We now need to sum (17) over pairs of (different) edges in  $\mathbf{G}$ . We start with

$$\sum_{\substack{a \rightarrow b \neq c \rightarrow d \\ \text{edges of } \mathbf{G}}} r_{ab}^{\mathbf{G}} + r_{cd}^{\mathbf{G}} = 2 \sum_{\substack{a \rightarrow b \neq c \rightarrow d \\ \text{edges of } \mathbf{G}}} r_{ab}^{\mathbf{G}} = 2(\mathbf{e}(\mathbf{G}) - 1) \sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} r_{ab}^{\mathbf{G}} = 2(\mathbf{e}(\mathbf{G}) - 1)(\mathbf{v}(\mathbf{G}) - 1).$$

The last step follows from Foster's theorem, originally proved by Foster [10] for graphs and generalized to weighted graphs by Tetali [19]. In turn, the weighted graph result implies the result for multigraphs: loop edges contribute nothing to resistance distance, since the resistance distance from any vertex to itself is clearly zero, and, since conductances add in parallel circuits, we can interpret  $k$  edges connecting two vertices as a single edge of weight (i.e., conductance)  $k$  for the purposes of computing resistance distances.

We next consider

$$\sum_{\substack{a \rightarrow b \neq c \rightarrow d \\ \text{edges of } \mathbf{G}}} r_{ac}^{\mathbf{G}} + r_{ad}^{\mathbf{G}} + r_{bc}^{\mathbf{G}} + r_{bd}^{\mathbf{G}} = \sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \left( \sum_{\substack{c \rightarrow d \text{ edge of} \\ \mathbf{G} - (a \rightarrow b)}} r_{ac}^{\mathbf{G}} + r_{ad}^{\mathbf{G}} + r_{bc}^{\mathbf{G}} + r_{bd}^{\mathbf{G}} \right).$$

Choose any vertex  $r \in \mathbf{G}$ . In the inner sum,  $r$  appears as  $c$  once for every edge with tail  $r$  in  $\mathbf{G} - (a \rightarrow b)$  and  $r$  appears as  $d$  once for every edge with head  $r$  in  $\mathbf{G} - (a \rightarrow b)$ . Note that loop edges pose no special difficulty— $r$  may appear as both  $c$  and  $d$  for a single  $c \rightarrow d$  in this case. Further, multiple edges also require no special treatment. Together,  $r_{ar}$  and  $r_{br}$  appear  $\deg_{\mathbf{G} - (a \rightarrow b)} r$  times in the inner sum, so

$$\sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \left( \sum_{\substack{c \rightarrow d \text{ edge of} \\ \mathbf{G} - (a \rightarrow b)}} r_{ac}^{\mathbf{G}} + r_{ad}^{\mathbf{G}} + r_{bc}^{\mathbf{G}} + r_{bd}^{\mathbf{G}} \right) = \sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \sum_{r \in \mathbf{G}} (\deg_{\mathbf{G} - (a \rightarrow b)} r) (r_{ar}^{\mathbf{G}} + r_{br}^{\mathbf{G}}).$$

Now  $\deg_{\mathbf{G}-(a \rightarrow b)} r = \deg_{\mathbf{G}} r$  for all  $r$  except  $a$  and  $b$ , where we have  $\deg_{\mathbf{G}-(a \rightarrow b)} a = \deg_{\mathbf{G}} a - 1$  and  $\deg_{\mathbf{G}-(a \rightarrow b)} b = \deg_{\mathbf{G}} b - 1$ . Thus

$$\sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \sum_{r \in \mathbf{G}} (\deg_{\mathbf{G}-(a \rightarrow b)} r) (r_{ar}^{\mathbf{G}} + r_{br}^{\mathbf{G}}) = \sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \left( -2r_{ab} + \sum_{r \in \mathbf{G}} (\deg_{\mathbf{G}} r) (r_{ar}^{\mathbf{G}} + r_{br}^{\mathbf{G}}) \right).$$

The first sum is equal to  $-2(\mathbf{v}(\mathbf{G}) - 1)$ , again by Foster's theorem. As above, if we fix any  $q \in \mathbf{G}$ , it appears as  $a$  in the sum once for every edge with tail  $q$  and as  $b$  in the sum above once for every edge with head  $q$ . Therefore, the term  $r_{qr}^{\mathbf{G}}$  appears  $\deg_{\mathbf{G}} q$  times in the sum. We have thus proved

$$\begin{aligned} \sum_{\substack{a \rightarrow b \\ \text{edge of } \mathbf{G}}} \left( -2r_{ab} + \sum_{r \in \mathbf{G}} (\deg_{\mathbf{G}} r) (r_{ar}^{\mathbf{G}} + r_{br}^{\mathbf{G}}) \right) &= -2(\mathbf{v}(\mathbf{G}) - 1) + \sum_{q, r \in \mathbf{G}} (\deg_{\mathbf{G}} q \deg_{\mathbf{G}} r) r_{qr} \\ &= -2(\mathbf{v}(\mathbf{G}) - 1) + 2 \text{Kf}^*(\mathbf{G}). \end{aligned}$$

Putting all of this together, we can conclude that

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{a \rightarrow b \neq c \rightarrow d} \sum_{\substack{q \in \mathcal{I}(a \rightarrow b) \\ r \in \mathcal{I}(c \rightarrow d)}} r_{qr}^{\mathbf{G}_n} = \frac{1 + 2\mathbf{e}(\mathbf{G})(\mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G})) - \mathbf{v}(\mathbf{G})}{6} + \frac{1}{2} \text{Kf}^*(\mathbf{G}).$$

Further, if  $a \rightarrow b$  is any single edge of  $\mathbf{G}$ , then

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{q, r \in \mathcal{I}(a \rightarrow b)} r_{qr}^{\mathbf{G}_n} = \frac{1}{6} + \frac{1}{6} r_{ab}^{\mathbf{G}},$$

so we have (using Foster's theorem again),

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{a \rightarrow b} \sum_{q, r \in \mathcal{I}(a \rightarrow b)} r_{qr}^{\mathbf{G}_n} = \frac{\mathbf{e}(\mathbf{G}) + \mathbf{v}(\mathbf{G}) - 1}{6}.$$

We are now ready to return to (16) and assemble the pieces. Recalling that the cycle rank  $\text{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$ , we get

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \sum_{q, r \in \mathcal{I}} r_{qr}^{\mathbf{G}_n} = \frac{2 \text{Loops}(\mathbf{G}) - 1}{6} \mathbf{e}(\mathbf{G}) + \frac{1}{2} \text{Kf}^*(\mathbf{G}). \quad (18)$$

Using Proposition 17 and summing directly, one finds that

$$\lim_{n \rightarrow \infty} \frac{1}{n^3} \left( 2 \sum_{\substack{q \in \mathcal{I} \\ r \in \mathcal{B}}} r_{qr} + \sum_{q, r \in \mathcal{B}} r_{qr} \right) = 0,$$

so in fact (18) proves the Proposition.  $\square$



## 6. PROOF OF THEOREM 13

We are now ready to prove Theorem 13.

*Proof.* Recall from Theorem 8 that

$$\mathcal{E}(R_g^2; \mathbf{G}_n) = \frac{d}{\mathbf{v}(\mathbf{G}_n)^2} \text{Kf}(\mathbf{G}_n).$$

Since  $\mathbf{v}(\mathbf{G}_n) = \mathbf{e}(\mathbf{G})(n-1) + \mathbf{v}(\mathbf{G})$ , we have

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathcal{E}(R_g^2; \mathbf{G}_n) = \lim_{n \rightarrow \infty} \frac{d}{n \mathbf{v}(\mathbf{G}_n)^2} \text{Kf}(\mathbf{G}_n) = \frac{d}{\mathbf{e}(\mathbf{G})^2} \lim_{n \rightarrow \infty} \frac{1}{n^3} \text{Kf}(\mathbf{G}_n).$$

Computing the last term with Proposition 19, we see that

$$\frac{d}{\mathbf{e}(\mathbf{G})^2} \lim_{n \rightarrow \infty} \frac{1}{n^3} \text{Kf}(\mathbf{G}_n) = \frac{d}{2\mathbf{e}(\mathbf{G})} \left( \frac{1}{3} \text{Loops}(\mathbf{G}) - \frac{1}{6} \right) + \frac{d}{4\mathbf{e}(\mathbf{G})^2} \text{Kf}^*(\mathbf{G}).$$

A result of Chen and Zhang [7], generalized to weighted graphs and hence to multigraphs by Chen [5, p. 1694], is that the degree-Kirchhoff index is related to the normalized graph Laplacian by

$$\text{Kf}^*(\mathbf{G}) = 2\mathbf{e}(\mathbf{G}) \text{tr } \mathcal{L}^+,$$

which completes the proof. □

## 7. CONTRACTION FACTORS

We can now study the asymptotic behavior of the expected radius of gyration in a subdivided graph. But it's still a little unclear why we should divide by  $n$ , and what these coefficients really mean from a polymer science point of view. We can give a nicer interpretation of our theorem by recalling the idea of the contraction factor (cf. [17, 1.48]).

**Definition 20.** The *contraction factor*  $g$  of a tcw with graph  $\mathbf{G}$  and  $\mathbf{v}$  vertices is the ratio

$$g(\mathbf{G}) = \frac{\mathcal{E}(R_g^2; \mathbf{G})}{\mathcal{E}(R_g^2; \text{path graph with } \mathbf{v} \text{ vertices})}.$$

Combining Theorem 13 with the formula (8) for the expected radius of gyration of the path graph yields the asymptotic contraction factor of an  $n$ -part edge subdivided graph  $\mathbf{G}_n$  as  $n \rightarrow \infty$ .

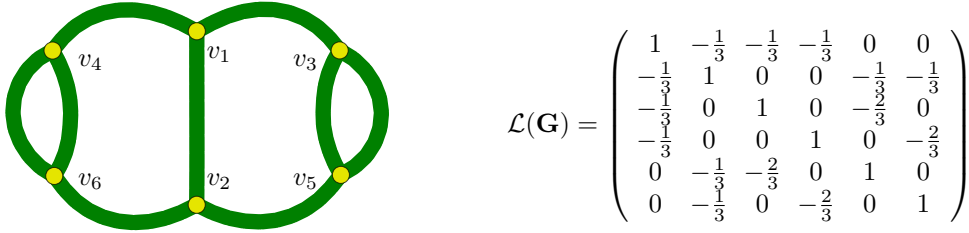


FIG. 3: A multigraph topological polymer synthesized in the Tezuka lab [18] together with its normalized graph Laplacian  $\mathcal{L}(\mathbf{G})$ .

**Theorem 21.** *For any connected multigraph  $\mathbf{G}$  (including loop and multiple edges), if  $\mathbf{G}_n$  is the  $n$ -part edge subdivision of  $\mathbf{G}$ , then the contraction factor  $g(\mathbf{G}_n)$  of a tcw with underlying graph  $\mathbf{G}_n$  embedded in  $\mathbb{R}^d$  obeys*

$$g(\mathbf{G}_\infty) := \lim_{n \rightarrow \infty} g(\mathbf{G}_n) = \frac{3}{e(\mathbf{G})^2} \left( \text{Tr } \mathcal{L}^+(\mathbf{G}) + \frac{1}{3} \text{Loops}(\mathbf{G}) - \frac{1}{6} \right).$$

We may now use Theorem 21 to estimate the relative sizes in solution of large topological polymers with different underlying graphs. Figure 3 shows an example of a topological polymer synthesized in the Tezuka lab. A direct computation shows that the eigenvalues of the  $6 \times 6$  matrix  $\mathcal{L}(\mathbf{G})$  are  $2, \frac{5}{3}, 1, 1, \frac{1}{3}$  and  $0$ . This means that the eigenvalues of  $\mathcal{L}^+(\mathbf{G})$  are  $\frac{1}{2}, \frac{3}{5}, 1, 1, 3$  and  $0$ . Since  $\text{Loops}(\mathbf{G}) = 4$  the result of Theorem 21 is that

$$g(\mathbf{G}_\infty) = \frac{109}{405} \simeq 0.269135.$$

A quick numerical experiment shows that for the graph  $\mathbf{G}$  in Figure 3,  $g(\mathbf{G}_{10}) \simeq 0.270064$  (sample average over 1 million trials), which shows that Theorem 21 is a quite good estimate even for small values of  $n$ . In fact,  $g(\mathbf{G}_2) \simeq 0.261687$  (sample average over 1 million trials), so the estimate is useful even for  $n = 2$ .

**Definition 22.** *The relative contraction factor of  $\mathbf{G}_1$  and  $\mathbf{G}_2$  is given by*

$$g(\mathbf{G}_1, \mathbf{G}_2) = \frac{\mathcal{E}(R_g^2; \mathbf{G}_1)}{\mathcal{E}(R_g^2; \mathbf{G}_2)} = \frac{g(\mathbf{G}_1)}{g(\mathbf{G}_2)}.$$

We now consider the example of the family of tricyclic and tetracyclic topological polymers shown in Figure 4. We chose these structure graphs because these polymers were actually synthesized by Suzuki, Yamamoto, and Tezuka [18] in 2014. To compare our simple model to a more detailed simulation, we constructed random configurations with topological constraints and self-avoiding effects via LAMMPS with the TSUBAME supercomputer at the Tokyo Institute of Technology. As is standard for simulations in the Kremer-Grest model, we used a repulsive Lennard-Jones (LJ) potential to model steric effects and a finitely extensive and nonlinear elastic (FENE) potential to model the bonds between atoms.

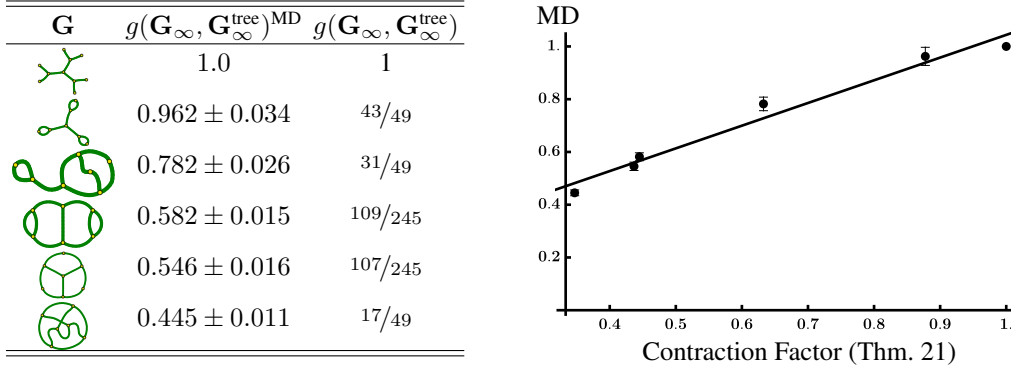


FIG. 4: The table above left shows six tricyclic and tetracyclic topological polymers synthesized by Suzuki, Yamamoto, and Tezuka [18] with the asymptotic relative contraction factor estimated by molecular dynamics simulation using (19) and the asymptotic relative contraction factor predicted by Theorem 21. The plot above right shows the molecular dynamics results ( $y$ -axis) versus the theoretical relative contraction factors ( $x$ -axis). The linear fit between molecular dynamics results and Theorem 21 has an  $R^2$  (coefficient of determination) of 0.966.

We took conformations at every  $5\tau$  steps, where the relaxation time  $\tau$  was defined by the number of time steps at which the correlation between conformations was  $1/e$ . This relaxation time depends on the graph type of polymers and increases as the number of atoms grows. For instance, for the bipartite complete graph  $K_{3,3}$  of 447 atoms  $\tau = 3.2 \times 10^5$ . For the same graph with 897 atoms,  $\tau = 1.6 \times 10^6$  steps. For the alpha graph with comparable numbers of atoms relaxation took longer, requiring  $\tau = 5.3 \times 10^5$  for a 448 atom graph and  $2.0 \times 10^6$  steps for an 898-atom graph.

For polymer chains which have excluded volume, the expected squared radius of gyration should be proportional to  $\mathbf{v}^{2 \times 0.588}$  for large  $\mathbf{v}$ . Thus we fitted the results of the molecular dynamics data to  $\mathcal{E}(R_g^2; \mathbf{G}_n) = C_{\mathbf{G}} \mathbf{v}(\mathbf{G}_n)^{1.176} + \Delta_{\mathbf{G}}$ . We measured the quality of the fit to this model by  $\chi^2/DF$  (degrees of freedom), and obtained values  $\leq 1.1$  for all the graph types tested. We were then able to compute asymptotic contraction factors for our molecular dynamics data relative to the tree graph  $\mathbf{G}^{\text{tree}}$  by taking

$$g(\mathbf{G}_\infty, \mathbf{G}_\infty^{\text{tree}})^{\text{MD}} = \frac{C_{\mathbf{G}}}{C_{\text{tree}}}. \quad (19)$$

The resulting estimates of  $g(\mathbf{G}_\infty, \mathbf{G}_\infty^{\text{tree}})^{\text{MD}}$  are given in the table at left in Figure 4, which also compares these results to the predictions of Theorem 21. We see that the extremely simple computation in the theorem, which requires only finding the eigenvalues of a small matrix, is very successful at predicting the results of molecular dynamics simulations which consumed dozens of hours of supercomputer time.

## 8. CONCLUSION AND FUTURE DIRECTIONS

The Gaussian model of topologically constrained random walks is conceptually quite simple: the key idea is that each coordinate vector  $w^k$  of the edge vector  $w$  is sampled from a multivariate Gaussian supported on the linear subspace  $\text{im } B^T$  of  $\mathbb{R}^e$ . However, we have seen that despite its simplicity, the model has a rich mathematical theory and significant explanatory power.

There are many more interesting questions associated to this model. First, it would obviously be interesting to make the model more realistic by incorporating some notion of excluded volume. We note that this can't change the situation too much (at least for large configurations) as our molecular dynamics simulation did include these effects, but Theorem 21 still explained the simulation data quite well. Second, it would be interesting to predict the expected value of other quantities used for proxies in SEC, such as the mean width or hydrodynamic radius (cf. [22]). However, we think that one of the most promising future directions for study is the prediction of the properties of topological polymers whose underlying graph type is random, such as rubber or collagen. We have taken great care above to build a theory which applies to multigraphs for precisely this reason, as multiple and loop edges can't be excluded from most random graph models without significant difficulty.

## 9. ACKNOWLEDGMENTS

We have benefitted tremendously from discussions with various colleagues and friends, and we would especially like to thank Yasuyuki Tezuka and Satoshi Honda for conversations about topological polymers and Fan Chung for introducing us to spectral graph theory. We are also grateful to the Japan Science and Technology Agency, which sponsored conferences at Ochanomizu University and the Tokyo Institute of Technology that catalyzed the present work, and to 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) for their support.

- 
- [1] Jason Cantarella, Tetsuo Deguchi, Clayton Shonkwiler, and Erica Uehara. Gaussian random embeddings of multigraphs. Preprint, 2020, [arXiv:2001.11709](https://arxiv.org/abs/2001.11709) [cond-mat.stat-mech].
  - [2] Ángeles Carmona, Margarida Mitjana, and Enric Monsó. Effective resistances and Kirchhoff index in subdivision networks. *Linear and Multilinear Algebra*, 65(9):1–15, 2017.
  - [3] Edward F. Casassa. Some statistical properties of flexible ring polymers. *Journal of Polymer Science: Part A*, 3(2):605–614, 1965.
  - [4] Isaac Chavel. *Eigenvalues in Riemannian Geometry*. Academic Press, Orlando, 1984.

- [5] Haiyan Chen. Random walks and the effective resistance sum rules. *Discrete Applied Mathematics*, 158(15):1691–1700, 2010.
- [6] Haiyan Chen and Fuji Zhang. Resistance distance and the normalized Laplacian spectrum. *Discrete Applied Mathematics*, 155(5):654–661, 2007.
- [7] Haiyan Chen and Fuji Zhang. Resistance distance local rules. *Journal of Mathematical Chemistry*, 44(2):405–417, 2007.
- [8] Richard Courant and David Hilbert. *Methods of Mathematical Physics Vol II: Partial Differential Equations*. Wiley–Interscience, New York–London, 1962.
- [9] Bruce E. Eichinger. Configuration statistics of Gaussian molecules. *Macromolecules*, 13(1):1–11, 1980.
- [10] Ronald M. Foster. The average impedance of an electrical network. In Staff of the Department of Aeronautical Engineering and Applied Mechanics of the Polytechnic Institute of Brooklyn, editors, *Reissner Anniversary Volume: Contributions to Applied Mechanics*, pages 333–340. J. W. Edwards, Ann Arbor, 1949.
- [11] Ivan Gutman and Bojan Mohar. The quasi-Wiener and the Kirchhoff indices coincide. *Journal of Chemical Information and Computer Sciences*, 36(5):982–985, 1996.
- [12] Douglas J. Klein and Milan Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1–4):81–95, 1993.
- [13] Hendrik Anthony Kramers. The behavior of macromolecules in inhomogeneous flow. *The Journal of Chemical Physics*, 14(7):415–424, 1946.
- [14] Bojan Mohar, Darko Babić, and Nenad Trinajstić. A novel definition of the Wiener index for trees. *Journal of Chemical Information and Modeling*, 33(1):153–154, 1993.
- [15] Anatoliĭ Platonovich Prudnikov, Yurii Aleksandrovich Brychkov, and Oleg Igorevich Marichev. *Integrals and Series. Vol. 1*. Gordon & Breach Science Publishers, New York, 1986.
- [16] Michael Rubinstein and Ralph H. Colby. *Polymer Physics*. Oxford University Press, 2003.
- [17] Robert Stepto, Taihyun Chang, Pavel Kratochvíl, Michael Hess, Kazuyuki Horie, Takahiro Sato, and Jiří Vohlřídál. Definitions of terms relating to individual macromolecules, macromolecular assemblies, polymer solutions, and amorphous bulk polymers (IUPAC Recommendations 2014). *Pure and Applied Chemistry*, 87(1):71–120, 2015.
- [18] 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.
- [19] Prasad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.
- [20] Yasuyuki Tezuka. Topological polymer chemistry designing complex macromolecular graph constructions. *Accounts of Chemical Research*, 50(11):2661–2672, 2017.
- [21] Erica Uehara and Tetsuo Deguchi. Statistical properties of multi-theta polymer chains. *Journal of Physics A: Mathematical and Theoretical*, 51(13):134001, 2018.
- [22] Yanwei Wang, Iwao Teraoka, Flemming Y. Hansen, Günther H. Peters, and Ole Hassager. A theoretical study of the separation principle in size exclusion chromatography. *Macromolecules*, 43(3):1651–1659, 2010.
- [23] Yujun Yang. The Kirchhoff index of subdivisions of graphs. *Discrete Applied Mathematics*, 171:153–157, 2014.
- [24] Yujun Yang and Douglas J. Klein. Resistance distance-based graph invariants of subdivisions and triangulations of graphs. *Discrete Applied Mathematics*, 181:260–274, 2015.
- [25] Heping Zhang and Yujun Yang. Resistance distance and Kirchhoff index in circulant graphs. *Internation-*

- tional Journal of Quantum Chemistry*, 107(2):330–339, 2006.
- [26] Bruno H. Zimm and Walter H. Stockmayer. The dimensions of chain molecules containing branches and rings. *The Journal of Chemical Physics*, 17(12):1301, 1949.