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

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

(Dated: October 11, 2022)

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, which should be useful for predicting the g-factors of topological polymer models with excluded volume.

1. INTRODUCTION

We consider a classical model of polymers, discussed by James, Guth, and Flory [16, 19, 20] and called *phantom network theory*, in which the polymer molecule consists 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 [27]. 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 [29, 31], 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.[1] Suppose $\mathbf{G}$ has $\mathbf{v} = \mathbf{v}(\mathbf{G})$ vertices and $\mathbf{e} = \mathbf{e}(\mathbf{G})$

---

[*]Mathematics Department, University of Georgia, Athens, GA, USA

[†]Department of Physics, Ochanomizu University, Tokyo 112–8610, Japan

[‡]Department of Mathematics, Colorado State University, Fort Collins, CO, USA

[1] The direction picked for each edge is arbitrary and won't affect the theory. The directions just need to be consistent
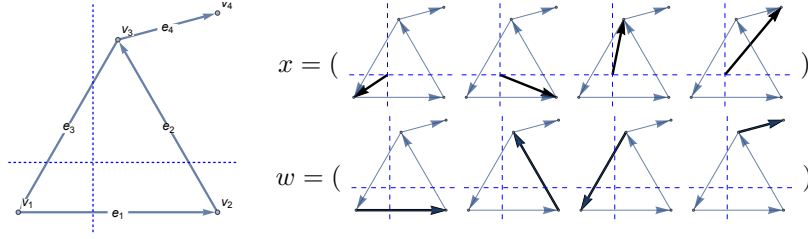
FIG. 1: A particular graph embedding in $\mathbb{R}^2$, along with the components of its vertex vector $x \in (\mathbb{R}^2)^4$ and edge vector $w \in (\mathbb{R}^2)^4$.

edges. 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 $x^k \in \mathbb{R}^{\mathbf{v}}$ is the vector of k-th coordinates of all vertex positions. An *edge vector* w for $\mathbf{G}$ is a $w \in (\mathbb{R}^d)^{\mathbf{e}}$ where $w_j \in \mathbb{R}^d$ is the displacement along edge e_j and $w^k \in \mathbb{R}^{\mathbf{e}}$ 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 $\mathbf{v} \times \mathbf{e}$ *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 $x^k \in \mathbb{R}^{\mathbf{v}}$ may be a vertex vector, only $w^k \in \operatorname{im} B^T$ 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 $w^k \in \operatorname{im} B^T$. 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 $\operatorname{im} B^T \subset \mathbb{R}^{\mathbf{e}}$.

We have previously shown [4, 5] that Gaussian TCRWs are exactly Gaussian phantom networks. We also proved that

Theorem 3 ([4]). *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 $x^k \in \mathbb{R}^{\mathbf{v}}$ are sampled independently from multivariate normals with mean 0 and covariance matrix $L(\mathbf{G})^+$.*

---

throughout any particular set of calculations.

In this paper, we will study the distribution of squared distances $\|x_i - x_j\|^2$ between vertices in Gaussian TCRWs, and hence in phantom networks. We will first note that the expectation $\mathcal{E}(\|x_i - x_j\|^2; \mathbf{G})$ is equal to the resistance distance [18, 23, 25] 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 Gaussian TCRW is given in terms of the Kirchhoff index (or quasi-Wiener index) of $\mathbf{G}$, denoted $\mathrm{Kf}(\mathbf{G})$ and defined in Section 2.

In practice, most of the graphs used to model topological polymers are constructed from a multigraph $\mathbf{G}$ by replacing each edge in $\mathbf{G}$ by a chain of n edges for some large n. Calling the resulting graph $\mathbf{G}_n$, we are interested in determining the asymptotics of the Kirchhoff index $\mathrm{Kf}(\mathbf{G}_n)$ as $n \to \infty$.

Theorem 4. *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*

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

where $\mathrm{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$ is the cycle rank of $\mathbf{G}$ and $\mathrm{Kf}^*(\mathbf{G})$ is the "degree-Kirchhoff index" introduced by Chen and Zhang [10]:

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

The quantity r_{ij} is the resistance distance between vertices v_i and v_j; see Section 2 for the precise definition.

Along the way, we will prove some independently interesting results about resistances measured between points along edges in a resistor network.

Given the connection between Kirchhoff index and expected radius of gyration, this tells us something about the asymptotics of radius of gyration as we subdivide edges of a graph, corresponding to the thermodynamic limit in statistical mechanics as the system size goes to infinity. This result is most naturally expressed in terms of 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.

Theorem 5. *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*

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

Here $g(\mathbf{G}_n)$ is the g-factor of $\mathbf{G}_n$, $\mathrm{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$ 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 $\mathcal{L}(\mathbf{G})$.

We will 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. We suggest that the linear fit should be useful for estimating the g-factor of real topological polymers from the calculations with the ideal chain model. Recall that the estimates of the g-factor evaluated in molecular dynamics are for real topological polymer models with excluded volume, and hence are quite expensive to compute, while our coefficients are simple to compute.

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 [27].

The asymptotic contraction factors themselves could be useful for studying the mean-square radius of gyration for various real topological polymer models with excluded volume [32, 38]. It has been shown that the ratios among the estimates of the mean-square radius of gyration for real topological polymer models with different graphs and those of ideal topological polymer models with the same set of graphs are almost the same if the functionality (i.e., the degree) at each vertex is limited up to three [32]. Thus, even the asymptotic value of the mean-square radius of gyration for an ideal topological polymer model with graph $\mathbf{G}_n$ can be useful to estimate that of a real topological polymer model with the corresponding graph.

There is another interesting approach to the spectrum of the graph Laplacian for a topological polymer whose edges are subdivided into n pieces. A method for reducing the graph Laplacian of the Gaussian topological polymer of graph $\mathbf{G}_n$ which is obtained by replacing each chain of a given graph $\mathbf{G}$ by a chain of n edges was derived in [15]. In order to evaluate the g-factor by the method, one has to evaluate all the eigenvalues of the reduced matrix, which is not practical, in general. However, some information on the spectrum of eigenvalues can be investigated through it.

2. THE MULTIGRAPH LAPLACIAN AND RESISTOR NETWORKS

We commented above that the matrix $L = BB^T$ was known as the graph Laplacian of $\mathbf{G}$, and noted in Theorem 3 that the pseudoinverse L^+ gave the matrix of covariances of vertex positions for the ensemble of conformations of the polymer in solution. To understand these covariances better, we now summarize 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* $\deg v_i$ 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 adjacency matrix A is defined by $A_{ij} = \#$ one edge paths connecting v_i to v_j.

Every loop edge $v_i \to v_i$ 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

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

For functions $p \colon \{v_1, \ldots, v_\mathbf{v}\} \to \mathbb{R}$ 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

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

where v_j is *adjacent* to v_i in $\mathbf{G}$ if there is an edge $v_i \to v_j$ or $v_j \to v_i$ in $\mathbf{G}$. Note that for each loop edge $v_i \to v_i$ 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 $p \colon \{v_1, \ldots, v_\mathbf{v}\} \to \mathbb{R}$ 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 $e_k = v_i \to v_j$,

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

For functions $u \colon \{e_1, \ldots, e_\mathbf{e}\} \to \mathbb{R}$ 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

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

This means that the graph Laplacian $L = -\operatorname{grad}\operatorname{div}$. This agrees with the sign convention in Riemannian geometry [8], but is opposite from the sign convention in mathematical physics [12].

If the graph $\mathbf{G}$ is a conductive network, then a function $p \colon \{v_1, \ldots, v_\mathbf{v}\} \to \mathbb{R}$ is called a *potential*. Along every edge of $\mathbf{G}$, the difference in potential $\operatorname{grad} p = U$ 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 $\operatorname{div} I = 0$. 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

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

Note that the right hand side is in $(\ker L)^\perp = \operatorname{im} L^T = \operatorname{im} L$, 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

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

In graph theory [23], this value is called the *resistance distance* from v_i to v_j in **G**. Although electrical resistance seems to have little to do with random positions of vertices, the resistance distance plays a natural role in the theory of polymers; in Proposition 7 we will see that dr_{ij} is the expectation of the squared distance between monomers i and j.

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 σ_i then $A^+ := V\Sigma^+ U^T$, where Σ^+ is a diagonal matrix with entries

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

When A is symmetric with nonnegative eigenvalues (for example, when $A = L$ is a graph Laplacian), then the singular value decomposition is just the spectral decomposition $A = U\Sigma U^T$ and the singular values are exactly the eigenvalues.

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 $Ax = b$, meaning it minimizes $\|Ax - b\|$. Explicitly, this means that $x \in \ker A^\perp$ and Ax is the orthogonal projection of b to $\operatorname{im} A$. It is also true that $A^+ A$ is orthogonal projection to $\ker A^\perp$ and AA^+ is orthogonal projection to $\operatorname{im} A$.

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

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

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 $\operatorname{im} L$, the solution p^i exactly solves the slightly different Poisson problem

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

We can now prove

Proposition 6. *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

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

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

The importance of this proposition is that the vertex covariances $(L^+)_{ij}$ are related to each other in a highly structured way: the columns of the covariance matrix L^+ are solutions to Poisson problems and hence obey a difference equation. This insight is a powerful method in probability theory [13] and will be the key tool we use to prove our results.

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 ith 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 $(\ker L)^\perp$). □

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

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

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 [1–3, 18, 21–23, 37].

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 7. *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 Gaussian TCRW 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 $3j(n-j)/n$ [39]. 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

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

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

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

This recovers an asymptotic result of Deguchi and Uehara [33] and, independently, of Zhu, Wang, Li, and Wang [38]. 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 8. The radius of gyration of a TCRW with vertex vector x in $\mathbb{R}^d$ is given by

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

The expected radius of gyration of a Gaussian 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. [28, 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 [14, eq. 18a]:

Theorem 9. *For any Gaussian TCRW in $\mathbb{R}^d$ with multigraph* $\mathbf{G}$, *we have*

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

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

Proof. Substituting the result of Proposition 7 into Definition 8, we get

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

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. $\qquad\square$

We note that the connection between Kirchhoff index and the trace of L^+ is well-known [18], 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 9. First, consider the path graph with $\mathbf{v}$ vertices $v_1, \ldots, v_{\mathbf{v}}$. The resistance between vertex v_i and v_j is simply $|i - j|$. Summing, we get

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

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

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

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 $j \in 0, \ldots, \mathbf{v} - 1$. Only $\lambda_0 = 0$, so we have to sum $\sum_{j=1}^{\mathbf{v}-1} \csc^2(\frac{\pi j}{\mathbf{v}}) = \frac{1}{3}(\mathbf{v}^2 - 1)$ [26, 4.4.6.5, p. 644]. Zhang et al. [37] 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 $\frac{d\mathbf{v}}{12}$ [7, 24] 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.

Proposition 10. *[25] If $p(t) = \det(L - tI) = t^{\mathbf{v}} + c_1 t^{\mathbf{v}-1} + \cdots + c_{\mathbf{v}-1} t + c_{\mathbf{v}}$ is the characteristic polynomial of the graph Laplacian $L(\mathbf{G})$,[2] then $\mathcal{E}(R_g^2; \mathbf{G}) = -\frac{d}{\mathbf{v}} \frac{c_{\mathbf{v}-2}}{c_{\mathbf{v}-1}}$.*

We note that Proposition 10 has the perhaps surprising corollary:

Corollary 11. *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 [35]. 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 12. 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.

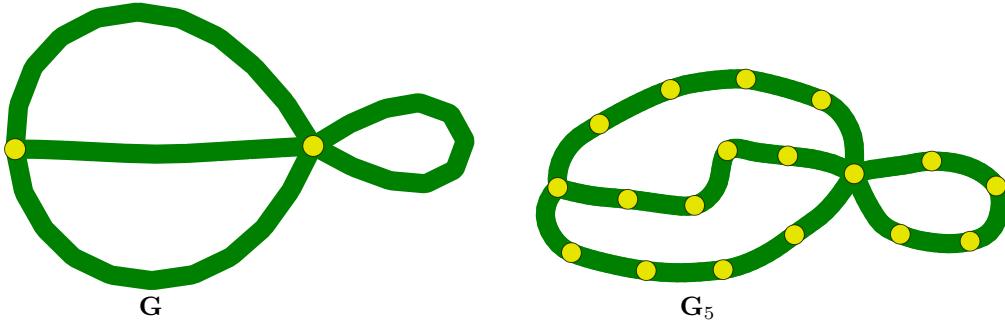[2] Note that $c_{\mathbf{v}} = 0$ for any graph Laplacian because $\lambda = 0$ is an eigenvalue.

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.

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

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

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

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

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 13. 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 on the asymptotics of radius of gyration:

Theorem 14. *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 Gaussian TCRW with underlying graph* $\mathbf{G}_n$ *embedded in* $\mathbb{R}^d$ *obeys*

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

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 2×2 matrix $\left(\begin{smallmatrix} 1 & -1 \\ -1 & 1 \end{smallmatrix} \right)$. A computation reveals that $\mathcal{L}^+ = \frac{1}{4}\mathcal{L}$ and $\operatorname{tr}\mathcal{L}^+ = \frac{1}{2}$. Thus Theorem 14 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 1×1 matrix whose single entry is zero, so $\operatorname{tr}\mathcal{L}^+ = 0$. Thus Theorem 14 also recovers the coefficient $\frac{d}{12}$ from (10).

We will now develop several tools leading up to the proof of Theorem 14. 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 [6].

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 $v_0, \ldots, v_n$. 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 15. *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 16. *Suppose $a \to b$ is an edge of a connected multigraph $\mathbf{G}$ subdivided in $\mathbf{G}_n$ by $v_0, \ldots, v_n$. The Poisson problem*

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

has minimum norm solution

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

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

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 15. The Laplacian may be checked directly at those points using (1). Since all solutions of Poisson problems on a connected graph **G** differ by a constant there must be some C which yields the minimum norm solution. $\qquad\square$

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 17. *Suppose that $a \to b$ and $c \to d$ are any edges in **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 $a \to b$ in $\mathbf{G}_n$ and r is k edges from c along the subdivision of $c \to d$ 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 $-\frac{1}{\mathbf{v}}$ 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 $\{c, d, r\}$. $\qquad\square$

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

We now use the decomposition formula of Proposition 17 to compute the resistance between vertices in the middle of subdivided edges in terms of resistances between their endpoints.

Proposition 18. *Suppose that $a \to b$ and $c \to d$ 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 $a \to b$ in $\mathbf{G}_n$ and r is k edges from c along the subdivision of $c \to d$ 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 $a \to b$ and $c \to d$ are different edges,

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

while if $a \to b = c \to d$, 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 [6] and Example 6 in [9] 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 18 – 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 $Lp^v(w) = -\frac{1}{\mathbf{v}}$. Thus, if $w \notin \mathcal{S}$, $(Lp^\mu)(w) = -\frac{1}{\mathbf{v}} \sum_{v \in \mathcal{S}} \mu_v = 0$.

This means that Lemma 15 tells us that p^μ is linear on $a \to b$ and $c \to d$. Thus we can evaluate p^μ 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 μ_c and μ_d,

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

Now (3) tells us that $r_{qr} = p^{qr}(q) - p^{qr}(r)$. Further, using the definition of p^μ, Proposition 17 can be rephrased as $p^{qr} = p^{ajb} - p^{ckd} + p^\mu$. Using the definition of p^μ 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 $a \to b$ and $c \to d$ are different. It remains only to prove that

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

Using (12) in Proposition 16, we can compute that

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

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

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

$\square$

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

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

Definition 19. We call a vertex $v \in \mathbf{G}_n$ an *interior* vertex if it is one of the vertices $v_1, \ldots, v_{n-1}$ along the subdivision in $\mathbf{G}_n$ of an edge $a \to b$ 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 $a \to b$ will be $\mathcal{I}(a \to b)$. We call vertices that are not interior *boundary* vertices, and note that they are also vertices of $\mathbf{G}$.

It is now clear that

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

We are ready to prove Theorem 4, which is really the heart of Theorem 14:

Theorem 4. 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

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

where $\mathrm{Loops}(\mathbf{G}) = \mathbf{e}(\mathbf{G}) - \mathbf{v}(\mathbf{G}) + 1$ is the cycle rank of $\mathbf{G}$ and $\mathrm{Kf}^*(\mathbf{G})$ is the "degree-Kirchhoff index" introduced by Chen and Zhang [10]:

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

We note that this result is compatible with the result of iteratively applying the subdivision operator given by Theorem 3.4 of [36], 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 $\lim_{n \to \infty} \frac{1}{n^3} \mathrm{Kf}(\mathbf{G}_n) = \lim_{k \to \infty} \frac{1}{(2^k)^3} \mathrm{Kf}(\mathbf{G}_{2^k})$ until we know that both limits exist. Further, [36] doesn't identify the number of loops as part of the result and depends on a result of [9] 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 18 we see that if $a \to b$ and $c \to d$ are different edges of $\mathbf{G}$, then

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

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

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

The last step follows from Foster's theorem, originally proved by Foster [17] for graphs and generalized to weighted graphs by Tetali [30]. 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

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

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

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

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

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

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

$$
\sum_{\substack{a \to b \\ \text{edge of } \mathbf{G}}} \left(-2r_{ab} + \sum_{r \in \mathbf{G}} (\deg_{\mathbf{G}} r)(r_{ar}^{\mathbf{G}} + r_{br}^{\mathbf{G}}) \right) = -2(\mathbf{v}(\mathbf{G}) - 1) + \sum_{q,r \in \mathbf{G}} (\deg_{\mathbf{G}} q \deg_{\mathbf{G}} r) r_{qr}
$$

$$
= -2(\mathbf{v}(\mathbf{G}) - 1) + 2 \operatorname{Kf}^*(\mathbf{G}).
$$

Putting all of this together, we can conclude that

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

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

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

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

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

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

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

Using Proposition 18 and summing directly, one finds that

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

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

6. PROOF OF THEOREM 14

We are now ready to prove Theorem 14.

Proof. Recall from Theorem 9 that

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

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

$$\lim_{n\to\infty} \frac{1}{n}\mathcal{E}\left(R_g^2; \mathbf{G}_n\right) = \lim_{n\to\infty} \frac{d}{n\mathbf{v}(\mathbf{G}_n)^2}\,\mathrm{Kf}(\mathbf{G}_n) = \frac{d}{\mathbf{e}(\mathbf{G})^2} \lim_{n\to\infty} \frac{1}{n^3}\,\mathrm{Kf}(\mathbf{G}_n).$$

Computing the last term with Theorem 4, we see that

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

A result of Chen and Zhang [11], generalized to weighted graphs and hence to multigraphs by Chen [9, p. 1694], is that the degree-Kirchhoff index is related to the normalized graph Laplacian by

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

which completes the proof. $\qquad\qquad\square$

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. [28, 1.48]).

Definition 20. The *contraction factor* g of a TCRW 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 14 with the formula (8) for the expected radius of gyration of the path graph yields Theorem 5, which we restate for convenience.

Theorem 5. 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 Gaussian TCRW with underlying graph $\mathbf{G}_n$ embedded in $\mathbb{R}^d$ obeys

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

We may now use Theorem 5 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 6×6 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 $\operatorname{Loops}(\mathbf{G}) = 4$ the result of Theorem 5 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 5 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 $n = 2$.
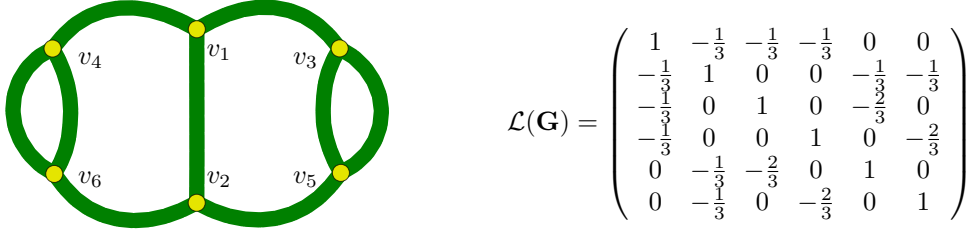
$$\mathcal{L}(\mathbf{G}) = \begin{pmatrix} 1 & -\frac{1}{3} & -\frac{1}{3} & -\frac{1}{3} & 0 & 0 \\ -\frac{1}{3} & 1 & 0 & 0 & -\frac{1}{3} & -\frac{1}{3} \\ -\frac{1}{3} & 0 & 1 & 0 & -\frac{2}{3} & 0 \\ -\frac{1}{3} & 0 & 0 & 1 & 0 & -\frac{2}{3} \\ 0 & -\frac{1}{3} & -\frac{2}{3} & 0 & 1 & 0 \\ 0 & -\frac{1}{3} & 0 & -\frac{2}{3} & 0 & 1 \end{pmatrix}$$

FIG. 3: A multigraph topological polymer synthesized in the Tezuka lab [29] together with its normalized graph Laplacian $\mathcal{L}(\mathbf{G})$.

Definition 21. 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 [29] 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 potential to model steric effects and a finitely extensive and nonlinear elastic (FENE) potential to model the bonds between atoms. The potential of the monomer A in the Kremer–Grest model is given by:

$$U_{bond} = -15.0 R_0 \ln \left[1 - \left(\frac{r_{AB}}{R_0} \right)^2 \right]$$

$$U_{ex} = \begin{cases} 4 \left((r_{AB})^{-12} - (r_{AB})^{-6} \right) & \left(r_{AB} \leq 2^{1/6} \right) \\ 0 & \left(r_{AB} > 2^{1/6} \right) \end{cases}$$

Here, $r_{AB} = |r_A - r_B|$ is the distance between monomer A and B, and R_0 is 1.5. Note that the force derived from U_{bond} is exerted only if A and B have a chemical link.

The trajectory of monomers is derived from solving the Langevin equation numerically:

$$m\ddot{r}_A = -\sum_B \left(\frac{dU_{bond}}{dr_A} + \frac{dU_{ex}}{dr_A} \right) + f_{friction} + f_{random}$$

$$f_{friction} = -\frac{m}{\delta} \dot{r}_A, \qquad f_{random} \propto \sqrt{\frac{m k_B T}{\delta \Delta t}}$$

Most of the quantities were unitless. We set $k_B T = 1.0$, $m = 1.0$, $\delta = 2.0$, $\Delta t = 0.01$. The Verlet algorithm was used for the time integrator. The monomers moved in a box under periodic boundary conditions. The box dimensions of xyz-coordinates are typically 5 or more times longer than the average of the gyration radius of the monomer chains. Thus, the monomers moved in dilute solution with small monomer density.

We took conformations at every 5τ steps, where the relaxation time τ 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 2.0×10^6 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 χ^2/DF (degrees of freedom), and obtained values ≤ 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

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

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 5. 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.

Although the fit in Figure 4 is purely empirical, it is good enough that we might hope that a similar correlation will also hold for other topological polymers with complex graphs. Since estimates of the g-factor are obtained in molecular dynamics with excluded volume, while the contraction factors are derived from the ideal chain model of topological polymers with no excluded volume, this would give a dramatic reduction in computational complexity of estimating g-factors. One reason for caution is that in all of our examples the functionality at each vertex is no more than three, so the mean-square radius of gyration may not be much enhanced by the excluded volume effect [32].

8. CONCLUSION AND FUTURE DIRECTIONS

Our approach to phantom network theory 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*
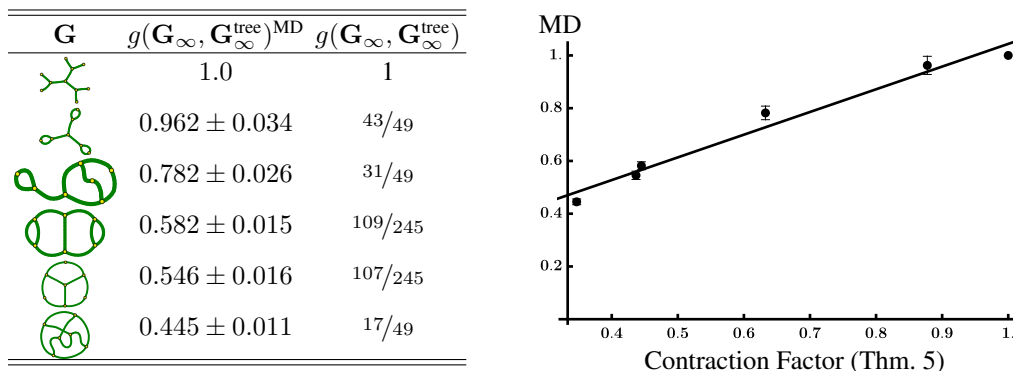
G	$g(\mathbf{G}_\infty, \mathbf{G}_\infty^{\mathrm{tree}})^{\mathrm{MD}}$	$g(\mathbf{G}_\infty, \mathbf{G}_\infty^{\mathrm{tree}})$
	1.0	1
	0.962 ± 0.034	$43/49$
	0.782 ± 0.026	$31/49$
	0.582 ± 0.015	$109/245$
	0.546 ± 0.016	$107/245$
	0.445 ± 0.011	$17/49$

FIG. 4: The table above left shows six tricyclic and tetracyclic topological polymers synthesized by Suzuki, Yamamoto, and Tezuka [29] with the asymptotic relative contraction factor estimated by molecular dynamics simulation using (19) and the asymptotic relative contraction factor predicted by Theorem 5. 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 5 has an R^2 (coefficient of determination) of 0.966.

the linear subspace $\operatorname{im} B^T$ of $\mathbb{R}^{\mathbf{e}}$. 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 5 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. [34]). 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.

---

[1] Darko Babić, Douglas J. Klein, István Lukovits, Sonja Nikolić, and Nenad Trinajstić. Resistance-distance matrix: A computational algorithm and its application. *International Journal of Quantum Chemistry*, 90(1):166–176, 2002.

[2] Ravindra B. Bapat, Ivan Gutman, and Wenjun Xiao. A simple method for computing resistance distance. *Zeitschrift für Naturforschung A*, 58(9–10):494–498, 2003.

[3] Danail Bonchev, Alexandru T. Balaban, Xiaoyu Liu, and Douglas J. Klein. Molecular cyclicity and centricity of polycyclic graphs. I. Cyclicity based on resistance distances or reciprocal distances. *International Journal of Quantum Chemistry*, 50(1):1–20, 1994.

[4] Jason Cantarella, Tetsuo Deguchi, Clayton Shonkwiler, and Erica Uehara. Gaussian random embeddings of multigraphs. Preprint, 2020, `arXiv:2001.11709 [cond-mat.stat-mech]`.

[5] Jason Cantarella, Tetsuo Deguchi, Clayton Shonkwiler, and Erica Uehara. Random graph embeddings with general edge potentials. Preprint, 2022, `arXiv:2205.09049 [cond-mat.stat-mech]`.

[6] Ángeles Carmona, Margarida Mitjana, and Enric Monsó. Effective resistances and Kirchhoff index in subdivision networks. *Linear and Multilinear Algebra*, 65(9):1–15, 2017.

[7] Edward F. Casassa. Some statistical properties of flexible ring polymers. *Journal of Polymer Science: Part A*, 3(2):605–614, 1965.

[8] Isaac Chavel. *Eigenvalues in Riemannian Geometry*. Academic Press, Orlando, 1984.

[9] Haiyan Chen. Random walks and the effective resistance sum rules. *Discrete Applied Mathematics*, 158(15):1691–1700, 2010.

[10] Haiyan Chen and Fuji Zhang. Resistance distance and the normalized Laplacian spectrum. *Discrete Applied Mathematics*, 155(5):654–661, 2007.

[11] Haiyan Chen and Fuji Zhang. Resistance distance local rules. *Journal of Mathematical Chemistry*, 44(2):405–417, 2007.

[12] Richard Courant and David Hilbert. *Methods of Mathematical Physics Vol II: Partial Differential Equations*. Wiley–Interscience, New York–London, 1962.

[13] Peter G. Doyle and J. Laurie Snell. *Random Walks and Electric Networks*, volume 22 of *Carus Mathematical Monographs*. Mathematical Association of America, Washington, DC, 1984.

[14] Bruce E. Eichinger. Configuration statistics of Gaussian molecules. *Macromolecules*, 13(1):1–11, 1980.

[15] Bruce E. Eichinger and James E. Martin. Distribution functions for Gaussian molecules. II. Reduction of the Kirchhoff matrix for large molecules. *The Journal of Chemical Physics*, 69(10):4595–4599, 1978.

[16] Paul J. Flory. *Statistical Mechanics of Chain Molecules*. Interscience Publishers, New York, 1969.

[17] 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.

[18] 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.

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

[20] Hubert M. James and Eugene Guth. Theory of the elastic properties of rubber. *The Journal of Chemical Physics*, 11(10):455–481, 1943.

[21] Douglas J. Klein. Resistance-distance sum rules. *Croatica Chemica Acta*, 75(2):633–649, 2002.

[22] Douglas J. Klein, José Luis Palacios, Milan Randić, and Nenad Trinajstić. Random walks and chemical graph theory. *Journal of Chemical Information and Computer Sciences*, 44(5):1521–1525, 2004.

[23] Douglas J. Klein and Milan Randić. Resistance distance. *Journal of Mathematical Chemistry*, 12(1–4):81–95, 1993.

[24] Hendrik Anthony Kramers. The behavior of macromolecules in inhomogeneous flow. *The Journal of Chemical Physics*, 14(7):415–424, 1946.

[25] 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.

[26] Anatoliĭ Platonovich Prudnikov, Yurii Aleksandrovich Brychkov, and Oleg Igorevich Marichev. *Integrals and Series. Vol. 1*. Gordon & Breach Science Publishers, New York, 1986.

[27] Michael Rubinstein and Ralph H. Colby. *Polymer Physics*. Oxford University Press, 2003.

[28] Robert Stepto, Taihyun Chang, Pavel Kratochvíl, Michael Hess, Kazuyuki Horie, Takahiro Sato, and Jiří Vohlídal. 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.

[29] 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.

[30] Prasad Tetali. Random walks and the effective resistance of networks. *Journal of Theoretical Probability*, 4(1):101–109, 1991.

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

[32] Erica Uehara and Tetsuo Deguchi. Statistical and hydrodynamic properties of topological polymers for various graphs showing enhanced short-range correlation. *The Journal of Chemical Physics*, 145(16):164905, 2016.

[33] Erica Uehara and Tetsuo Deguchi. Statistical properties of multi-theta polymer chains. *Journal of Physics A: Mathematical and Theoretical*, 51(13):134001, 2018.

[34] 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.

[35] Yujun Yang. The Kirchhoff index of subdivisions of graphs. *Discrete Applied Mathematics*, 171:153–157, 2014.

[36] Yujun Yang and Douglas J. Klein. Resistance distance-based graph invariants of subdivisions and triangulations of graphs. *Discrete Applied Mathematics*, 181:260–274, 2015.

[37] Heping Zhang and Yujun Yang. Resistance distance and Kirchhoff index in circulant graphs. *International Journal of Quantum Chemistry*, 107(2):330–339, 2006.

[38] Lijuan Zhu, Xiaoyan Wang, Jianfeng Li, and Yanwei Wang. Radius of gyration, mean span, and geometric shrinking factors of bridged polycyclic ring polymers. *Macromolecular Theory and Simulations*, 25(5):482–496, 2016.

[39] 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.