PHYLOGENETIC NETWORKS AS CIRCUITS WITH RESISTANCE DISTANCE

STEFAN FORCEY AND DREW SCALZO

ABSTRACT. Phylogenetic networks are notoriously difficult to reconstruct. Here we suggest that it can be useful to view unknown genetic distance along edges in phylogenetic networks as analogous to unknown resistance in electric circuits. This *resistance distance*, well known in graph theory, turns out to have nice mathematical properties which allow the precise reconstruction of networks. Specifically we show that the resistance distance for a weighted 1-nested network is Kalmanson, and that the unique associated circular split network fully represents the splits of the original phylogenetic network (or circuit). In fact, this full representation corresponds to a face of the balanced minimal evolution polytope for level-1 networks. Thus the unweighted class of the original network can be reconstructed by either the greedy algorithm neighbor-net or by linear programming over a balanced minimal evolution polytope. We begin study of 2-nested networks with both minimum path and resistance distance, and include some counting results for 2-nested networks.

1. INTRODUCTION

Consider an electrical circuit: a network made of wires joining resistors in parallel and in sequence, with some portion hidden inside a opaque box. It is not always possible to determine that portion by testing the visible leads. However, we prove here that if the hidden portion has a particular form made of connected cycles, and we can test the resistance between all the pairs of leads, then the lengths and connected structure of the cycles in the circuit are uniquely determined. The mathematics used to recover that circuit is more typically found in work on phylogenetic networks.

Modeling heredity as the flow of genetic information suggests that mutations in DNA might be analogous to resistance in an electrical circuit. The weights of edges in a phylogenetic network can represent genetic distances: if we have the genomes of the two endpoints of an edge then we can use a model of mutation rates to calculate a real number distance. For several edges that form the unique path between two taxon-labeled leaves, the total distance is the sum of those edge weights. Paths between leaves are only unique if the network is a tree. When paths between are not unique, one option is to take the distance to be that of the minimum length path. This option may correspond to a parsimonious approach—assuming the least complicated history. This minimum path length distance is studied for instance in [12].
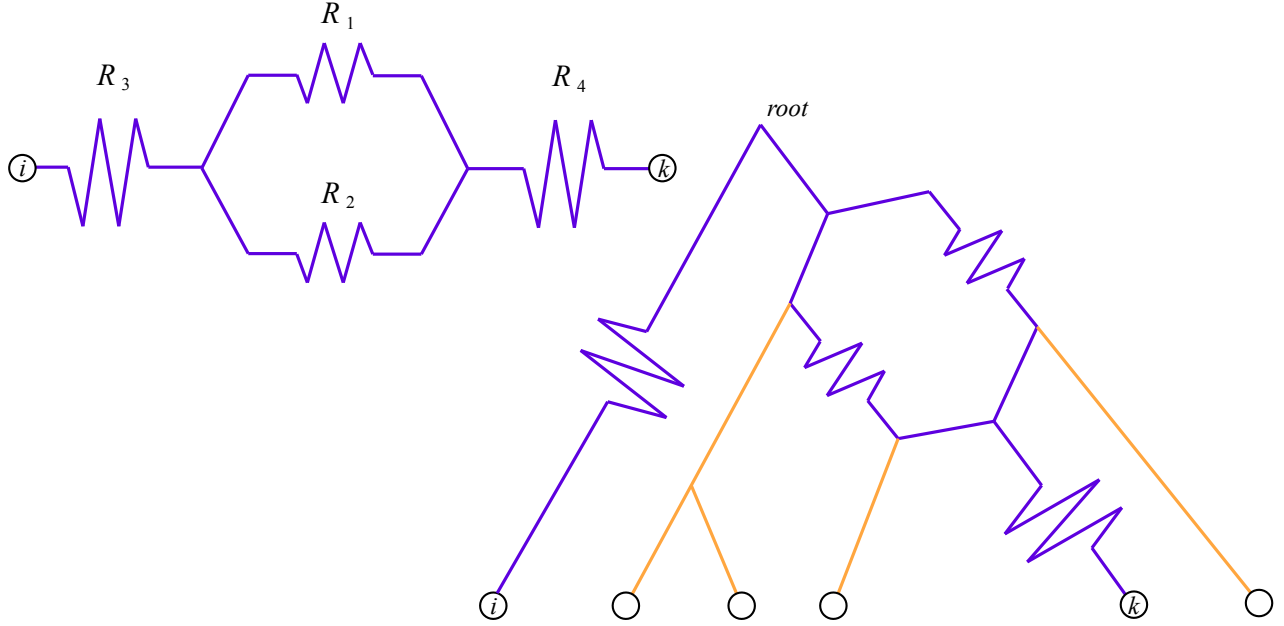
FIGURE 1. The total resistance from i to k is $R_{i,k} = \dfrac{R_1 R_2}{R_1 + R_2} + R_3 + R_4$. On the left is the circuit itself, on the right we see it as a pairwise circuit within a phylogenetic network. Here we have chosen i to be an outgroup, so the network is rooted at the top and the downward direction is forward in time.

Instead, however, the weight of an edge could represent the loss of information. Dividing and rejoining of edges illustrates events such as speciation, recombination, or hybridization. If the genetic information of an ancestor genome can be shared among descendants, and then collaboratively recovered upon hybridization, then a different metric than minimum distance may be appropriate. Here we consider weighted phylogenetic networks with the *resistance distance*, or *resistance metric.* The distance between two leaves of the network is found by considering the edge weights as electrical resistance, obeying Ohm's law. The metric resistance distance for all nodes (not only leaves) of graph is introduced in [15], and studied closely in subsequent papers such as [19] and [20]. To study graphs, the resistance of each edge is often assumed to have unit value, but the definitions allow any weight. We review the definitions in Section 2.1.3.

In [4], [5] and [6], the authors study circular planar graphs with boundary nodes that are analogous to the leaves of our phylogenetic networks. They consider resistance values (or conductivity) on the edges. They prove that complete information about the linear map which transforms electric current values at each boundary node to electric current values at all the edges can be used in some cases to recover the resistance values. In our applications there is no way to know the complete map of boundary currents to edge currents. However, we seek only to recover the graphical structure of the network, not the original edge weights.

In [10], the authors consider the entire set of resistance distances (again using unit values for edges), between any pair of nodes (not only leaves.) They show that using this metric is

useful for discovering Hamiltonian cycles via algorithms for the Travelling Salesman problem. There is a close connection to our applications, since the algorithm neighbor-net can be used as a greedy approach to the Travelling Salesman problem as shown in [16].

1.1. **Main Results and Overview.** In Section 2 we start by reviewing Ohm's law and resistance distance. Then we review the relevant definitions of mathematical phylogenetics, many taken from other sources to help make this paper self-contained. Section 3 contains some new results on 2-nested networks with regard to the minimum path distance. In Section 4 we state and prove the main results for 1-nested phylogenetic networks N. The upshot is that when the distances between taxa are effective resistances based on unknown connections, then using well known methods we can recover an unweighted circular split network, which gives us the precise class of (unweighted) 1-nested phylogenetic network. Specifically, this recovery is via the (greedy) algorithm neighbor-net as decribed in Theorem 4.3 or linear programming; see Theorem 5.5.

Several features of the resistance distance seem exactly suited to phylogenetic networks with weighted edges. First, from Theorem 4.1, the resistance distance of 1-nested phylogenetic networks is Kalmanson, allowing the circular split network to be uniquely reconstructed from the measured distances. Second, from Theorem 4.2, that reconstructed circular split network always displays precisely the same splits as the original network. As a consequence, the trivial splits which are the traditional final edges to the leaves of a phylogenetic network are automatically guaranteed to be represented in the split network—this is a condition beyond the basic Kalmanson condition. Finally, triangular subgraphs are interchangeable with three-edge stars when measuring resistance distance. This is known as the Y-Δ transform, pictured in Figure 2. The Y-Δ equivalence mirrors the fact that triangles in a phylogenetic network, when attached via bridges to the rest of the network, are indistinguishable from degree-three tree-like vertices by the linear functionals used for balanced minimal evolution. As well, the split networks are bipartite, so triangle free.

In Section 5 we review the balanced minimal evolution polytopes, and show how our results can be interpreted geometrically, in Theorems 5.2 and 5.5. In Section 6 we point out some interesting counterexamples and limiting cases, and conjecture about how to extend our results to more complicated networks. Finally in Section 7 we consider qualifications of experimental distance measurements in phylogenetics that would give justification for assuming the resistance analogy to be valid in practice.

2. Definitions and cited results

The classic Ohm and Kirchoff equations include: $R = V/I$ and $I = I_1 + I_2$. The first depends on the conductive material—it must be experimentally verified. It relates the resistance in a circuit to the constant voltage drop over the circuit and the constant current in all of the circuit. The second states that total current must equal the sum of circuit-parallel portions of that current after a branching in the circuit. Together, these rules imply the law for total resistance R_T for a pair of circuit-parallel resistances R_1, R_2. We have $R_T = R_1 R_2/(R_1 + R_2)$, which we refer to as Ohm's law for parallel resistance. Also, the voltage drop over a closed circuit must equal the total voltage: this implies that resistors in series are summed to find the total resistance. We illustrate the basic calculation of a total resistance in Figure 1. We illustrate the implied Y-Δ equivalence in Figure 2.
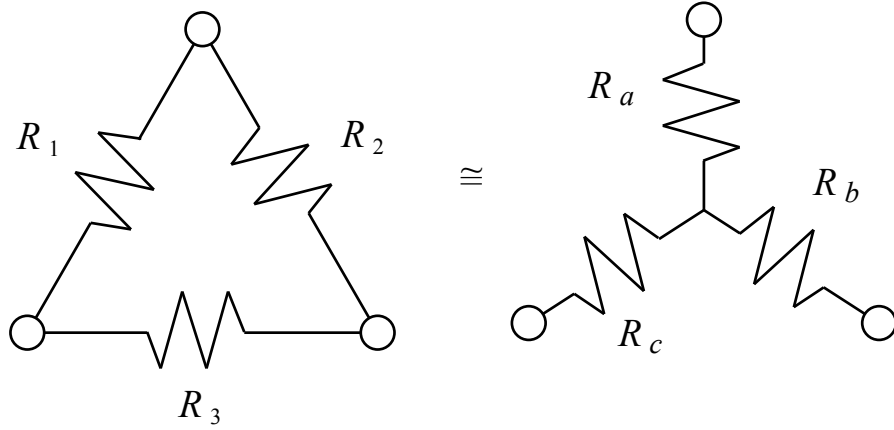
FIGURE 2. The two networks shown here have identical resistance between any two corresponding pairs of nodes at the three corners. Here $R_a = \dfrac{R_1 R_2}{R_1 + R_2 + R_3}, R_b = \dfrac{R_2 R_3}{R_1 + R_2 + R_3}$, and $R_c = \dfrac{R_1 R_3}{R_1 + R_2 + R_3}$.

2.1. **Phylogenetic definitions.** Many of the definitions and notes here are repeated from [12] for reference.

A *split* $A|B$ is a bipartition of $[n] = \{1, \ldots, n\}$. That is, A and B are non-empty disjoint subsets whose union is $[n]$. The two parts of a split are often called *clades*. If one clade of a split has only a single element, we call that split *trivial.* A *split system* is a set s of splits of $[n]$ which contains all the trivial splits. We say a split system s *refines* another split system s' when $s \supset s'$.

In this paper all graphs are simple (no multi-edges) and connected.

Definition 2.1. An *(unrooted) phylogenetic network* on $[n]$ is a simple connected graph with:
  i. Labeled leaves: n degree-1 vertices, labeled bijectively with the elements of $[n]$,
  ii. Unlabeled nodes: all these must have degree larger than 2.

A split $A|B$ is *displayed* by such a phylogenetic network N when there is (at least) one subset of edges of N whose deletion (keeping all nodes) results in two connected components with A and B their respective sets of labeled leaves. We call that collection of edges a *minimal cut* for the split when it contains no proper subset producing the same split. A *bridge* is a single edge which displays a split. A *trivial bridge* displays a trivial split. A *phylogenetic tree* is a cycle-free network, so every edge is a bridge. Figure 3 shows examples of splits displayed, for the trees and their two generalizations described here: phylogenetic networks and split networks. Recall that a cycle in a graph is a path of edges that does not revisit any nodes except for the node at which it starts and ends. The following is defined in [13]:

Definition 2.2. An unrooted phylogenetic network N is called *1-nested* when each edge of N is contained in at most one cycle, and N is triangle-free—all cycles are of length greater than 3 edges.

As in [12], we consider two 1-nested networks to be *split-equivalent* if they display the same set of splits. In fact, two split-equivalent 1-nested networks are related by the collapse (or
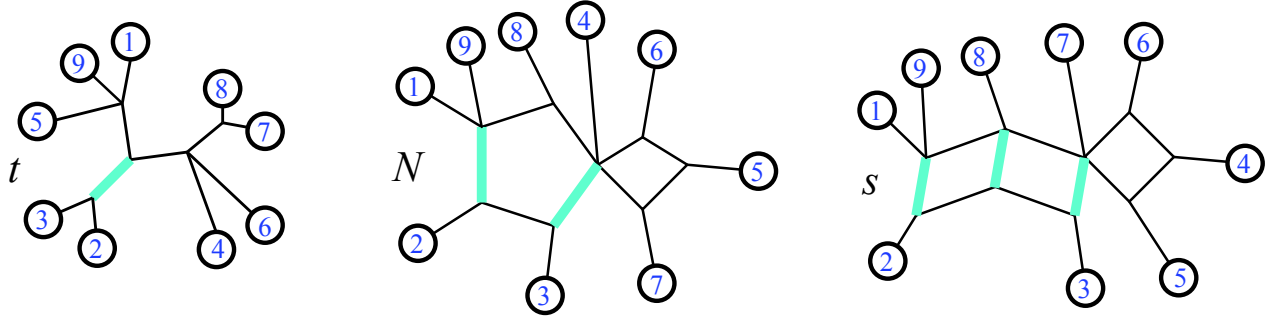
FIGURE 3. In a phylogenetic tree t, on the left, splits are always single edges. The highlighted edge is the split $\{2,3\}|\{1,4,5,6,7,8,9\}$. That same split is a pair of edges making a minimal cut in the 1-nested phylogenetic network N, center. Finally on the right, that same split is a set of parallel edges in a circular split network s.
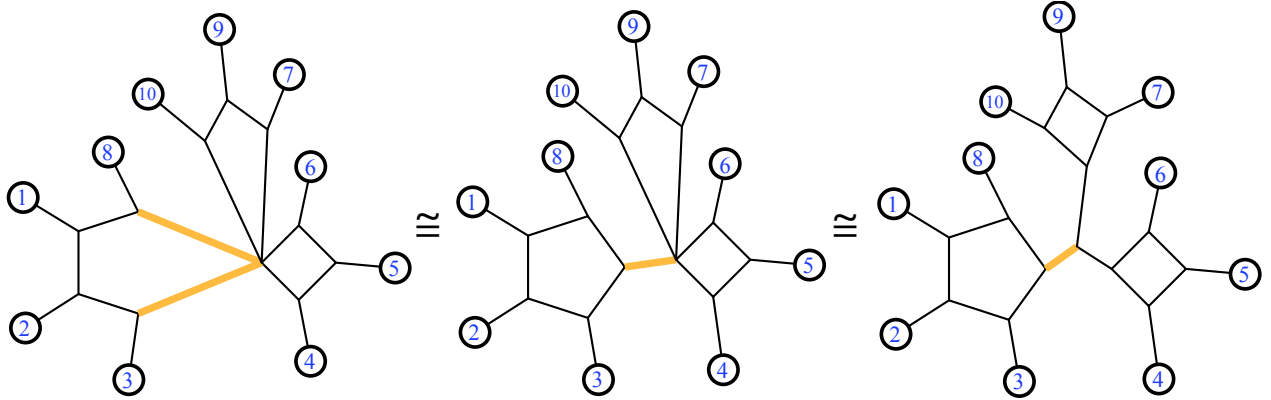

FIGURE 4. A trio of equivalent 1-nested phylogenetic networks, all representing the same set of splits. The highlighted edges display the same split in each network.

growth) of specific edges: any non-trivial bridge directly attached to a cycle can be collapsed to see a split-equivalent network. See Figure 4 for examples.

A *binary* phylogenetic network is one in which the unlabeled nodes each have degree 3.

We also will consider another generalization of a phylogenetic tree, in which each split corresponds to a unique set of edges.

Definition 2.3. A *split network* displaying a split system s on $[n]$ is an embedding in Euclidean space of a simple connected graph, also called s, with the following:

    i. Labeled leaves: n degree 1 nodes bijectively labeled by $[n]$,

    ii. Unlabeled nodes with degree larger than 1,

    iii. The set of edges partitioned into classes called *split-classes*, one split-class for each split $A|B$ in the system. It is required that for any two leaves: the set of edges on a shortest path between them intersects each split-class in at most one edge, and that the set of splits thus traversed is the same for any shortest path between those two leaves.
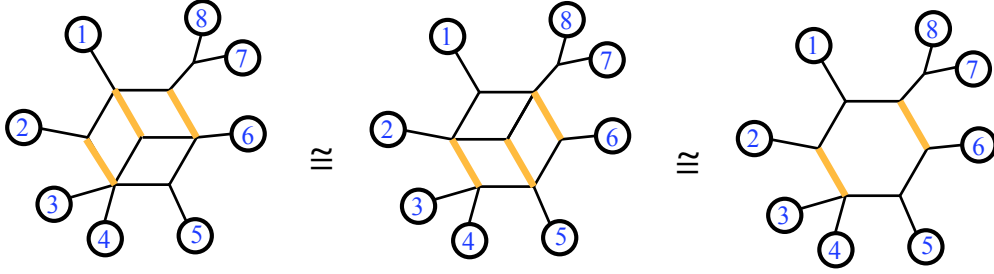
FIGURE 5. A trio of equivalent split networks, all three representing the same set of splits. The highlighted edges display the same split in each network. The third is the invariant exterior subgraph of all three.

iv. The split-class of edges corresponding to a split comprises a minimal cut displaying that split: deletion of those edges (keeping all nodes) results in two connected components with respective labeled leaves the two parts of that split.

The resulting graph will be bipartite. Typically each class of edges is embedded as a set of equal length parallel line segments. (Note: here parallel means geometrically parallel.) Alternate definitions use colors; the edges in a split-class are colored alike, as in, [7], [18]. A split-class of size one is a bridge. The requirements on split-classes imply that no two edges meet in a degree-2 node.

Several different split networks may often be drawn for the same split system, but we consider them equivalent as long as they represent the same splits. Note that in contrast to unrooted phylogenetic networks, the only sort of minimal cut of a split network that is said to display a split is one of the classes of geometrically parallel edges. Other cuts are ignored.

Definition 2.4. A *circular split system* is a split system which allows the embedding of a representative split network in the plane, with the labeled nodes all on the exterior, and thus arranged in a circular order.

Twisting the diagram around a bridge (reflecting one side through the line of the bridge), or around a cut-point node, does not change the list of splits. Any cyclic order of the leaves allowing an embedding of a split network in the plane is said to be *consistent* with that system.

Two circular split networks are equivalent if they display the same set of splits. An equivalence is shown in Figure 5.

The following lemma is from [12], included here for the terminology that will be useful in the next section.

Lemma 2.5. *Given a circular split network s, the nodes and edges adjacent to the exterior of the graph are a subgraph which is invariant: that is, this* exterior subgraph *will be identical to the exterior subgraph of any circular split network representing the same set of splits as s.*

For example see Figure 5. The exterior subgraph will be a series of cycles of even length, connected by cut-point nodes, nontrivial bridges, and trivial bridges to the leaves. In fact the exterior subgraph of s is a circular split network itself, displaying the same system as s.
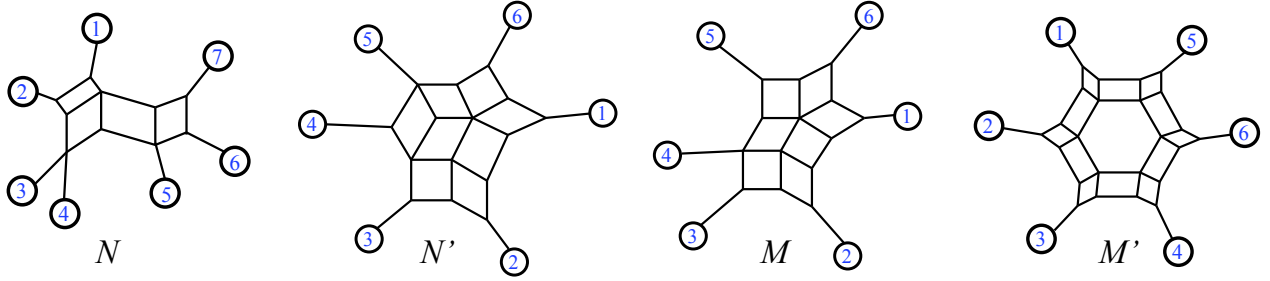
FIGURE 6. Two outer-path circular split networks on the left, N and N', have all shortest paths represented by exterior paths. Two non-outer-path circular split networks on the right, M and M' have representations which include shortcuts: for instance the path from leaf 1 to 4 in M and the path from 2 to 6 in M'.

(Typically, however, more interior edges are shown since parallelograms can help make the splits visually identifiable.)

Inntroduced in [12] is a subclass of circular split networks.

Definition 2.6. An *outer-path circular split system* is a split system whose representative circular split networks have shortest paths between pairs of leaves which can all be chosen to lie on the exterior of the diagram, that is, using only edges adjacent to the exterior.

Since the shortest paths are all the same length, this implies that outer-path split networks have no *shortcut*, that is, there is no path between leaves through the interior of the diagram that is strictly shorter than any path on the exterior. For examples, see Figure 6.

2.1.1. *Functions.*

Definition 2.7. For a 1-nested phylogenetic network N we define $\Sigma(N)$ to be the circular split system made up of the splits displayed by N.

In [13] it is shown that $\Sigma(N)$ is indeed a circular split system, since it can be represented by a circular split network, also referred to as $\Sigma(N)$.

Since $N \cong N'$ precisely when they display the same set of splits, Σ is well defined.

An algorithm for drawing a nice representing network of $\Sigma(N)$ is also presented in [13]. First, cycles of length 4 are each replaced by a parallelogram. For $m \geq 5$, each m-cycle is replaced by an *m-marguerite*: a collection of exactly $m^2 - 4m$ parallelograms arranged in a circle, each sharing sides with two neighbors, specifically organized as follows: each node of the original m-cycle is replaced by a rhombus, and then each edge of the cycle is replaced by $m - 5$ parallelograms in a row. The rows are attached to the rhombi along adjacent edges of each rhombus, so that the whole arrangement has $m(m-5)$ sides on the interior of the original m-cycle, and $m(m-3)$ sides on the exterior. Bridges are attached to the m remaining degree-2 vertices, one at each of the rhombi that replaced the original m nodes of the cycle.

Examples of representations of $\Sigma(N)$ are seen in Figure 7. Note that since the bridges in a split network are invariant, every representation of $\Sigma(N)$ will have the same bridges: these will match the maximal set of bridges of any representation of N. The range of $Sigma(N)$ will be referred to as the *faithfully phylogenetic* circular split networks.
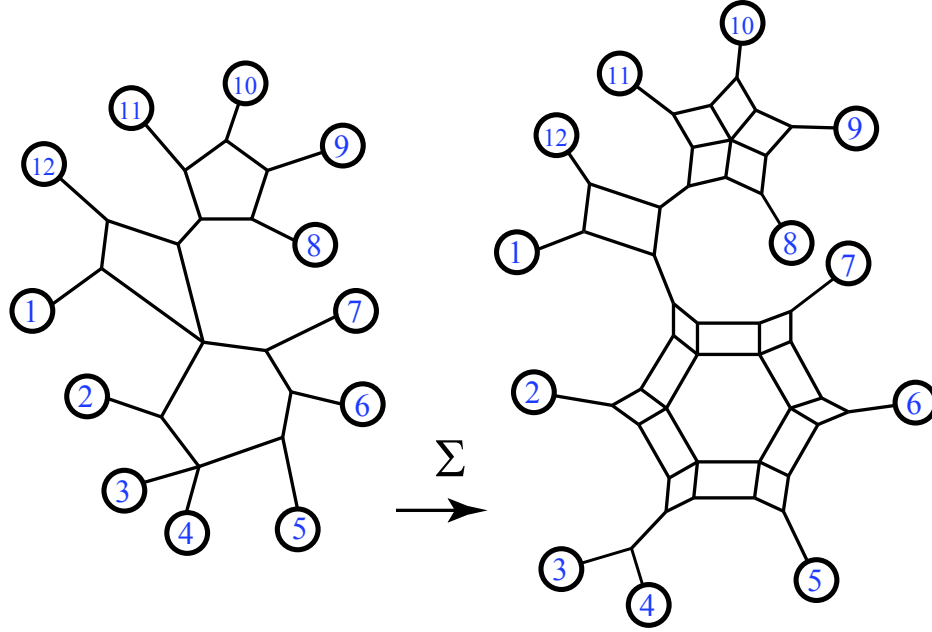
FIGURE 7. Example of the function Σ.

Next we define a function that takes a circular split network to a 1-nested phylogenetic network. The function is shown to exist in [13], and described on the split networks which are images of the function Σ. In [8] we define the general function L as follows:

Definition 2.8. Recall that the nodes and edges adjacent to the exterior of a circular split network are an invariant subgraph for the split system. Define $L(s)$ to be the *smoothed exterior subgraph* of s.

In other words, we construct the network $L(s)$ from a split system s by beginning with a split network diagram of s and considering the diagram as a planar drawing of its underlying planar graph, with leaves on the exterior. Then 1) delete all the edges that are not adjacent to the exterior of that graph, and 2) smooth away any resulting degree-2 nodes—delete the node but join the two adjacent edges to make one edge.

Note that by its construction, L preserves bridges and cut-point nodes. When restricted to phylogenetic trees, the functions L and Σ are both the identity. In general however, we have the following from [12]:

Theorem 2.9.

    1) L and Σ *form a Galois connection between unweighted 1-nested phylogenetic networks and circular split networks, with ordering by containment of splits, in which L is the lower and Σ the upper adjoint.*

    2) *The Galois connection via L and Σ is a reflection (but not a poset isomorphism).*

    3) *As corollaries from Galois theory* [11] *we see that L is surjective, (but Σ is not surjective) and that $L \circ \Sigma$ is the identity map.*

2.1.2. *Weights and metrics.* Weighted phylogenetic trees have non-negative real number assigned to their branches, often representing the genetic distance between the two nodes. A
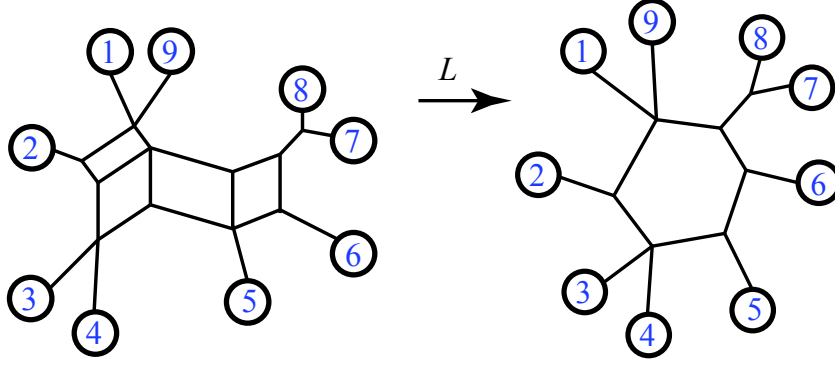
FIGURE 8. Example of the function L.

weight of 0 can mean the edge is collapsed, and the resulting space of trees, called BHV_n, is studied in [2]. Now we may generalize weighted trees with weighted networks in two distinct ways: by assigning non-negative real numbers to splits or to edges.

Definition 2.10. A *weighted phylogenetic network* N has non-negative real numbers assigned to its edges, described by a weight function w_N.

Definition 2.11. A *weighted split network* s has non-negative weights assigned to each split, by a weight function w_s. Equivalently, every edge in a geometrically parallel class of s has the same weight.

Definition 2.12. For a weighted phylogenetic network N, or a weighted split network s, we denote by $\overline{N}$, respectively $\overline{s}$, the unweighted networks found by forgetting the weights.

A pairwise distance function assigns a non-negative real number to each pair of values from $[n]$. We call the lexicographically listed outputs for distinct pairs a *distance vector* $\mathbf{d}$, with entries denoted $d_{ij} = \mathbf{d}(i,j) = \mathbf{d}(j,i)$ for each pair of taxa $i \neq j \in [n]$ (also known as a dissimilarity matrix, or discrete metric when obeying the metric axioms.)

Definition 2.13. When the distance vector is *Kalmanson*, or *circular decomposable* it means there exists a cyclic order of $[n]$ such that for any subsequence (i,j,k,l) of that order, $\mathbf{d}$ obeys this condition:

$$\max\{d_{ij} + d_{kl}, d_{jk} + d_{il}\} \leq d_{ik} + d_{jl}.$$

Definition 2.14. Given a weighted split system s on $[n]$ we can derive a metric $\mathbf{d}_s$ on $[n]$,

$$\mathbf{d}_s(i,j) = \sum_{i \in A, j \in B} w_s(A|B)$$

where the sum is over all splits of s with i in one part and j in the other. The metric is often referred to as the distance vector $\mathbf{d}_s$.

It is well known that Kalmanson metrics are represented uniquely by weighted circular split networks. Specifically, from [18] we have the following:

Lemma 2.15. *A distance vector $\mathbf{d}$ is Kalmanson with respect to a circular order c if and only if $\mathbf{d} = \mathbf{d}_s$ for s a unique weighted circular split system s, (not necessarily containing all trivial splits) with each split $A|B$ of s having both parts contiguous in that circular order c.*

Definition 2.16. We define the *minimum path distance* vector $\mathbf{d}_N$ for a weighted 1-nested phylogenetic network N, where

$$\mathbf{d}_N(i,j) = \min_p \{\sum_{e \in p} w_N(e) \mid p \text{ is a path connecting } i, j\}$$

where the minimum is over paths p from leaf i to leaf j, and each sum is over edges in one of those paths. Examples are calculated in Figures 9 and 11.

2.1.3. *Resistance distance.* Isolating sections of circuit-parallel paths between two leaves allows the Ohm relations, together with the Y-Δ transformation, to be used to find the effective resistance between those leaves. A simplifying fact is that the resistance between two leaves only depends on the resistances of edges that are in paths between those leaves. (We use the term *pairwise circuit* P_{ij} to refer to the edges that are in any path between leaves i, j. For example see Figure 18.)

There is a well-known alternate method for calculating effective resistances. As defined in [15], the resistance distance matrix for a graph G with n total vertices (leaves and non-leaf nodes) is given by:

$$\Omega_{ij} = \Gamma_{ii}^{-1} + \Gamma_{jj}^{-1} - 2\Gamma_{ij}^{-1}$$

where $\Gamma = L + 1/n$, the Laplacian matrix of G plus the $n \times n$ matrix with $1/n$ on the main diagonal, 0 elsewhere.

Our resistance distance for phylogenetic networks uses entries of the matrix Ω.

Definition 2.17. We define the *resistance distance* vector $\mathbf{d}_N^R$ for a weighted phylogenetic network N, where $\mathbf{d}_N^R(i,j)$ is the resistance distance on the graph between leaves i and j. That is, $\mathbf{d}_N^R(i,j) = \Omega_{ij}$ for leaves i and j. The distance can also be calculated using the basic relations of Ohm's law. Examples of the resistance distance vector are in Figures 10, 21 and 23.

2.1.4. *Weighted Functions.* Now we define functions between the weighted split networks and the weighted phylogenetic networks. As previously explained in [8], we begin by extending the function L to a weighted version L_w.

Definition 2.18. For a weighted circular split network s we define $L_w(s)$ to be the 1-nested phylogenetic network $L(\overline{s})$ (the smoothed exterior subgraph of the unweighted version of s), with weighted edges. The weight of an edge in the image is found by summing the weights of splits which contribute to that edge. Let $p_s(e)$ be the set of splits $A|B$ of s, such that $A|B$ is represented by edges in s one of which is used to form the edge e in $L(s)$. If w_s is the weight function on s then the weight function on $L_w(s)$ is:

$$w_{L_w(s)}(e) = \sum_{A|B \in p_s(e)} w_s(A|B).$$

By this definition we have the following (from [12]):

Lemma 2.19. $\overline{L_w(s)} = L(\overline{s})$.

For an example of L_w see Figure 12.

From [12], we have the fact that the minimum path distance is Kalmanson for planar networks. Therefore, we can make the following:
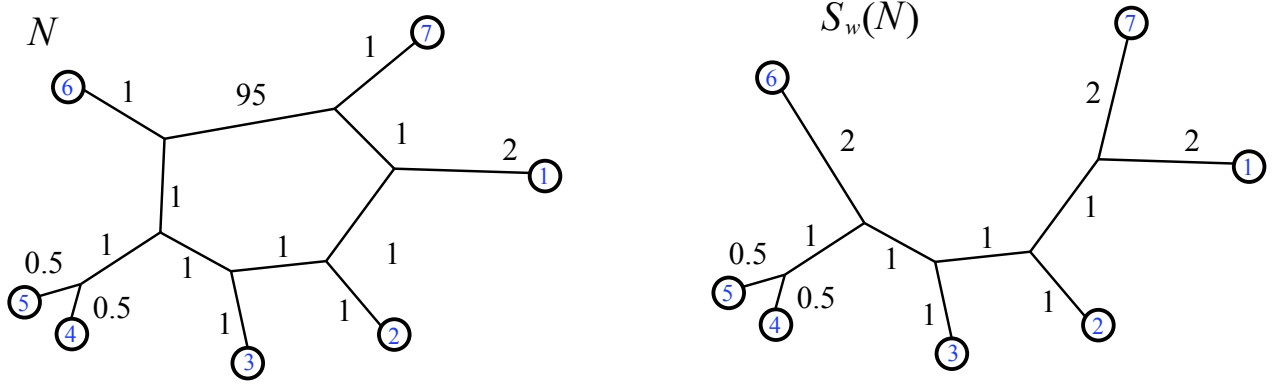
FIGURE 9. Example of the action of S_w. Here

$$\mathbf{d}_N = (4, 5, 6.5, 6.5, \; 7, \; 4,$$
$$3, 4.5, 4.5, \; 5, \; 4,$$
$$3.5, 3.5, \; 4, \; 5,$$
$$1, 3.5, 6.5,$$
$$3.5, 6.5,$$
$$7)$$

.

Definition 2.20. Given a weighted unrooted phylogenetic network N that can be drawn on the plane with leaves on the exterior, we define $S_w(N)$ to be the unique weighted circular split network with the same minimum path distance vector as N. That is, $\mathbf{d}_N = \mathbf{d}_{S_w(N)}$. This image is calculable, for instance, as the circular split network $S_w(N) = \mathcal{N}(\mathbf{d}_N)$, where $\mathcal{N}$ is the neighbor-net algorithm defined by [3].

For an example see Figure 9. Another example of S_w, on a 2-nested network, is in Figure **??**. When we restrict to the domain of weighted circular split networks arising from weighted 1-nested networks, the codomain of S_w is the outer-path circular split networks, and the distance vector is preserved by the map L_w. Specifically from [12] we have:

Lemma 2.21. *For any weighted 1-nested phylogenetic network N, if $s = S_w(N)$ then s is outer-path and thus $\mathbf{d}_{L_w(s)} = \mathbf{d}_s$.*

S_w is defined using the minimum path distance metric. Similarly, since we will see that the resistance distance is Kalmanson in Theorem 4.1, we can make the following definition using resistance distance.

Definition 2.22. For a weighted 1-nested phylogenetic network N we define $R_w(N)$ to be the unique weighted circular split network corresponding to the resistance distance $\mathbf{d}_N^R$. The algorithm neighbor-net is guaranteed to produce $R_w(N)$ using input $\mathbf{d}_N^R$.

The function R_w is shown by example in Figure 10. For another example, on a 2-nested network that happens to be Kalmanson, see Figure 23.
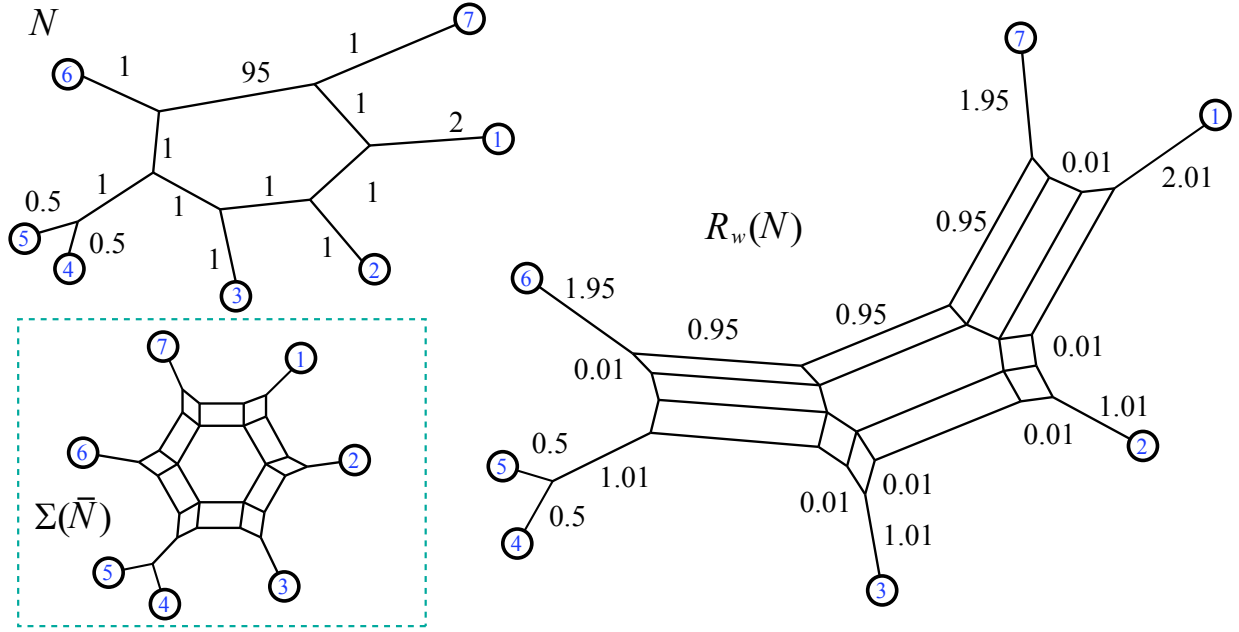
FIGURE 10. Example of the function R_w which takes a weighted phylogenetic network and outputs the split network associated to its resistance distance. Here

$$\mathbf{d}_N^R = (3.99, 4.96, 6.41, 6.41, 6.84, 3.99,$$
$$2.99, 4.46, 4.46, 4.91, 3.96,$$
$$3.49, 3.49, 3.96, 4.91,$$
$$1, \ 3.49, 6.34,$$
$$3.49, 6.34,$$
$$6.75)$$

.

When restricted to phylogenetic trees, the functions L_w and S_w are both the identity, and $S_w = R_w$. When restricted to weighted 1-nested phylogentic networks and outer-path circular split networks we have the following from [12]:

Theorem 2.23.

1) L_w and S_w form a Galois connection between weighted 1-nested phylogenetic networks and outer-path circular split networks, ordered by containment of splits when distance vectors match, in which L_w is the upper and S_w the lower adjoint.
2) The Galois connection via L_w and S_w is a coreflection (but not a poset isomorphism).
3) L_w is injective but not surjective, S_w is surjective but not injective, and $S_w \circ L_w$ is the identity map.

3. 2-NESTED NETWORKS

In this section, we examine functions between 1-nested and 2-nested networks, and circular split networks. Towards the end of [13], the authors ask: is it possible to characterize split

systems induced by more complex uprooted networks such as 2-nested networks (i.e., networks obtained from 1-nested networks by adding a chord to a cycle)? We interpret this question to be about the result of applying S_w and begin by noting two interesting cases.

3.1. Minimum path distance on 2-nested networks.

Definition 3.1. If every edge of N is part of at most two cycles, we call it a 2-nested network. By this definition, 2-nested networks contain 1-nested networks as a subset, which in turn contain 0-nested networks, which are phylogenetic trees. By strict k-nested networks we mean k-nested but not $(k-1)$-nested.

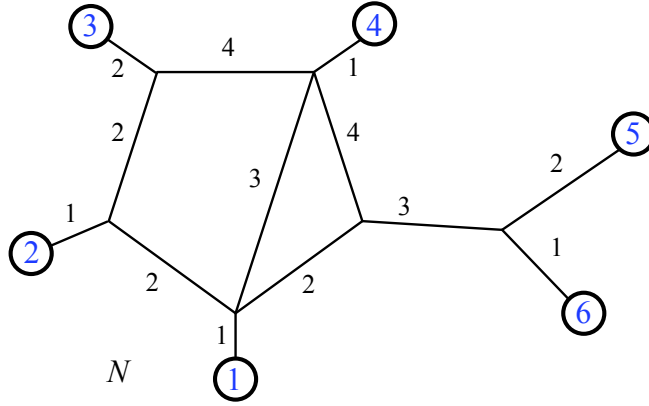A weighted 2-nested network is shown in Figure 11, with its minimum path distance vector.

FIGURE 11. The minimum path distance vector for the weighted 2-nested network N is $\mathbf{d}_N = (4, 7, 5, 8, 7, 5, 7, 10, 9, 7, 13, 12, 10, 9, 3)$. Note that $\mathbf{d}_{14} = 5$, for example, referring to the shortest distance between leaves 1 and 4.

Note that weighted 2-nested networks often have images under S_w that are not outer-path circular split networks. For instance see Figure 12. Thus applying S_w and then L_w in sequence will produce a weighted 1-nested network that has a different distance vector than the original.

Theorem 3.2. *For every weighted 1-nested network M, there exists some (not unique) weighted 2-nested network N such that the minimum path distance vectors coincide:* $\mathbf{d}_M = \mathbf{d}_N$.

Proof. Consider a 1-nested network M with positive values for its edges and a 2-nested network N that has the same exterior subgraph. Let N also have the same positive values for its exterior edges, but a positive value for its internal chord large enough such that on paths of least distance the internal chord of the 2-nested network is never used. Therefore both networks will have the same distance vector $\mathbf{d}_M = \mathbf{d}_N$. $\square$

3.2. Counting 2-nested networks.
We begin counting the total number of unweighted binary, triangle free, 2-nested networks. The numbers of unweighted binary, triangle free, 2-nested networks exist with n leaves are: 6, 120, 2790,... for $n = 4, 5, 6$.
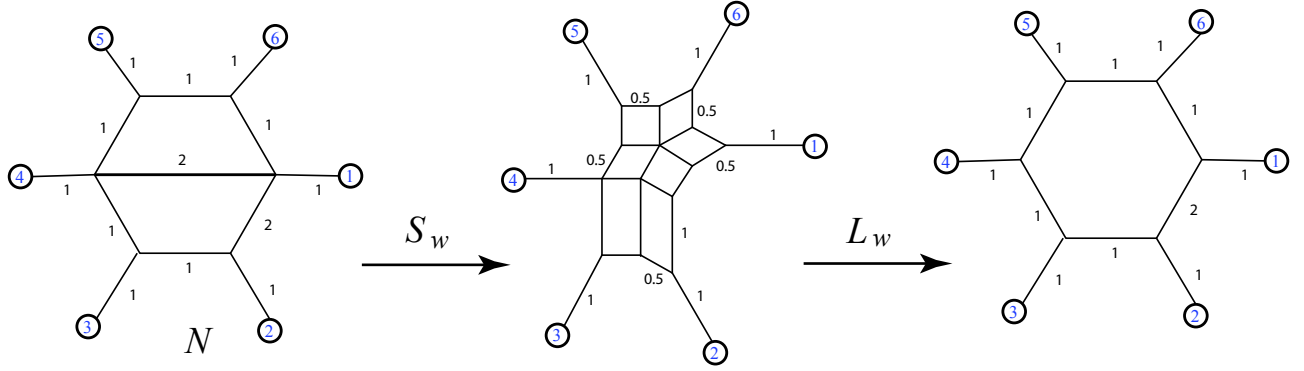
FIGURE 12. Here the output of $S_w(N)$ is a non-outer-path circular split network, and its image under L_w has a distance vector that does not match the original: for instance $\mathbf{d}_N(1,4) = 4$ but the distance from one to 4 in $L_w(S_w(N))$ is 5.

First, consider structures with 4 leaves ($n = 4$). We start by considering the unlabeled pictures, and then count the ways to assign the values $1, \ldots, 4$ to the leaves. In fact, we can simplify further by finding the unlabelled 1-nested networks and showing the potential locations of chords simultaneously in each picture. There is one such unlabeled picture for $n = 4$ as shown in Figure 13, with two possible internal chords. There are $\frac{3!}{2}$ ways to arrange the leaves before choosing a chord. Therefore, the total number of unweighted binary triangle-free 2-nested networks with $n = 4$ leaves is $(2)\frac{3!}{2} = 6$.

For $n = 5$ the possible internal structures are shown in Figure 13. There are 5 possible internal chords for one structure, and 2 possible internal chords for the other. The number of ways to arrange the leaves of the first structure is $n!$, and the second structure is $(n - 1)!$ (since the first is not rotationally symmetric.) However, rearranging the leaves clockwise and counterclockwise yield the same rearrangement, so we must then divide by 2 to eliminate half of the arrangements garnered from the counting of those leaves. Finally, if there were a bridge connecting any components of the structure, simply divide by 2 for the twisting around that bridge. The counting for each $n = 5$ structure in Figure 13 is as follows:

$$\frac{5(2)}{2}\frac{4!}{2} = 60,$$

$$\frac{4(1)}{2}\frac{5!}{2}\frac{1}{2} = 60.$$

The total number of networks for $n = 5$ is $= 60 + 60 = 120$.

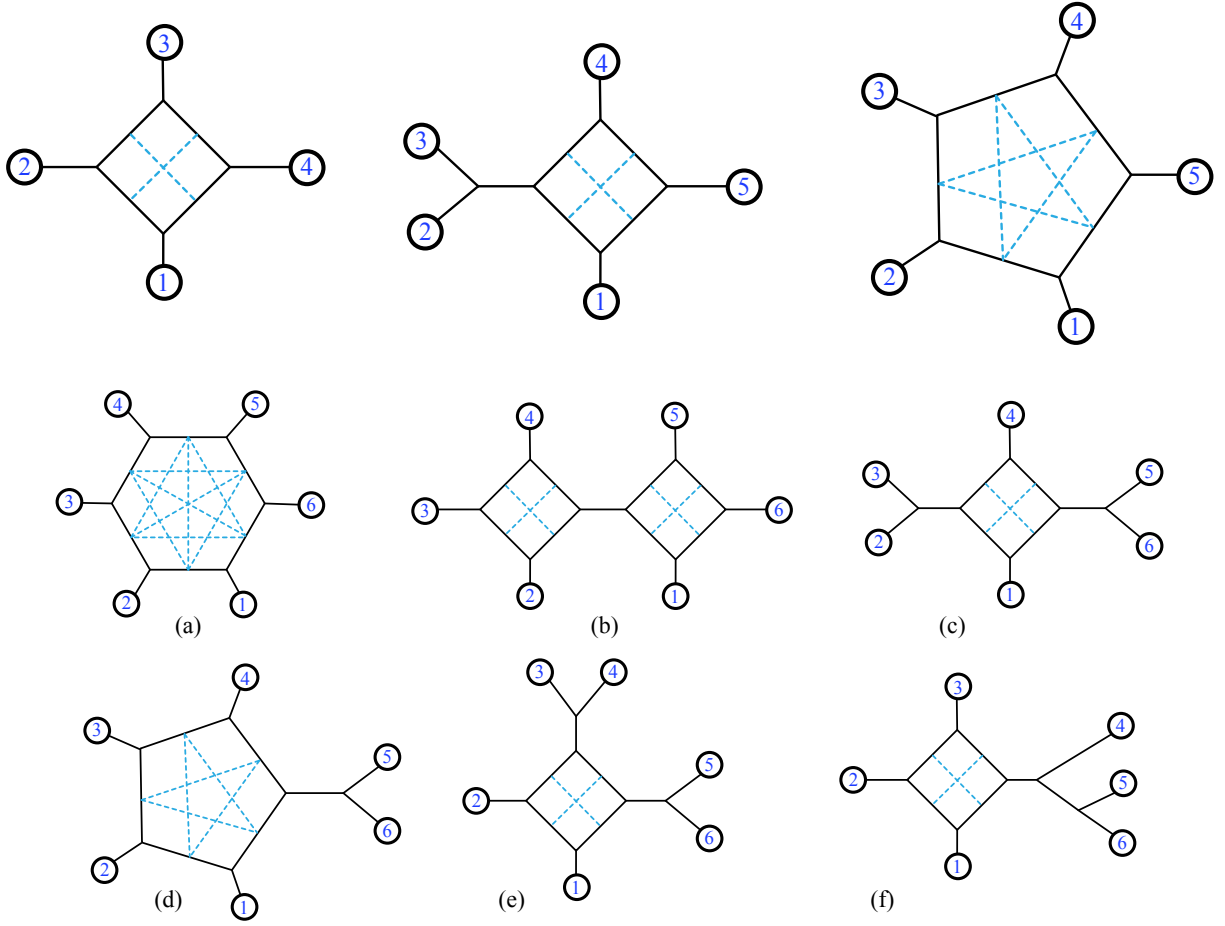For $n = 6$ the counting for each structure is as follows (from a to f as pictured in Figure 13):

FIGURE 13. For $n = 4$, there is only one exterior structure with two internal chords possible (as seen by the dotted lines above). For $n = 5$, there exist two exterior structures. For $n = 6$ there are 6 such structures, labelled a-f.

$$\text{(a)} \qquad \frac{6(3)}{2}\frac{5!}{2} = 540,$$

$$\text{(b)} \qquad (2)(2)\frac{4(1)}{2}\frac{6!}{2}\frac{1}{2}\frac{1}{2} = 720,$$

$$\text{(c)} \qquad \frac{4(1)}{2}\frac{6!}{2}\frac{1}{4}\frac{1}{2} = 90,$$

$$\text{(d)} \qquad \frac{5(2)}{2}\frac{6!}{2}\frac{1}{2} = 900,$$

$$\text{(e)} \qquad \frac{4(1)}{2}\frac{6!}{2}\frac{1}{4}\frac{1}{2} = 180,$$

$$\text{(f)} \qquad \frac{4(1)}{2}(6!)\frac{1}{4} = 360.$$

The total number of networks for $n = 6$ is $= 540 + 720 + 90 + 900 + 180 + 360 = 2790$.

Notice for (f), reading the labels clockwise is not equivalent to reading them counterclockwise due the tree structures. This means we just consider $6!$ and not $\frac{6!}{2}$.

4. KALMANSON NETWORKS

The main result in this section is that the resistance metric is Kalmanson for 1-nested phylogenetic networks, and that the unique associated split network has the same exterior form as the original 1-nested phylogenetic network. First we show that $\mathbf{d}_N^R$ obeys the Kalmanson condition: there exists a circular ordering of $[n]$ such that for all $i < j < k < l$ in that ordering,

$$\max\{\mathbf{d}_N(i,j) + \mathbf{d}_N(k,l), \mathbf{d}_N(j,k) + \mathbf{d}_N(i,l)\} \leq \mathbf{d}_N(i,k) + \mathbf{d}_N(j,l).$$

Theorem 4.1. *Given a 1-nested phylogenetic network N with positive weighted edges and n leaves, the resistance metric on its leaves is Kalmanson.*

Proof. The cyclic order that we need to exist in order to demonstrate the Kalmanson property is found by choosing any cyclic order of $[n]$ consistent with N.

There are three basic cases to consider. Begin by noting that for each pair of the four leaves i, j, k, l there is a sub-graph, called the *pairwise circuit*, for instance P_{ik}, made of all the edges which are part of any path between those two leaves. The pairwise circuit will contain perhaps some cycles—it will in fact be a series of cycles connected by paths. We are especially interested in the intersection I of the two "crossing" pair circuits, $I = P_{ik} \cap P_{jl}$.

Case 1: The intersection I is a single cycle. Here the four leaves i, j, k, l have pairwise circuits that reach the cycle I at four different nodes. Notice that any of the two pairwise circuits summed in the Kalmanson condition will include all four of the smaller pairwise circuits from each of the four leaves i, j, k, l to the node of I closest to that respective leaf. We'll call those closest nodes v_i, v_j, v_k, v_l. Thus the only differences between the four sums in the Kalmanson condition arise from the different contributions of the cycle I. We call the cumulative edge weights between those nodes a from v_i to v_j and b, c, d similarly, following the cyclic order. For instance, Figure 14, b is the sum of the weights on edges of I between the nodes v_i and v_j.

Thus the sum $\mathbf{d}_N^R(i,j) + \mathbf{d}_N^R(k,l)$ contains the terms

$$\frac{b(a+d+c)}{a+b+c+d} + \frac{d(a+b+c)}{a+b+c+d}.$$

The sum $\mathbf{d}_N^R(j,k) + \mathbf{d}_N^R(i,l)$ contains

$$\frac{c(a+b+d)}{a+b+c+d} + \frac{a(b+c+d)}{a+b+c+d}.$$

While the sum $\mathbf{d}_N^R(i,k) + \mathbf{d}_N^R(j,l)$ contains

$$\frac{(a+d)(b+c)}{a+b+c+d} + \frac{(a+b)(c+d)}{a+b+c+d}.$$

All the edge weights are positive, and the denominators of all the terms are the same. Clearly the latter sum, when expanded, has a numerator larger than either of the first two.

Case 2: The intersection I is a series of cycles but has at least two cycles. In this case there are two possible ways that the inequality breaks down, depending on which pair of consecutive leaves (ij or jk) reach I at the same end of I. In Figure 15 below we choose ij to do so, on the left-hand cycle, but the other option is similar. Checking this case can be done visually for the equality: $\mathbf{d}_N^R(i,k) + \mathbf{d}_N^R(j,l) = \mathbf{d}_N^R(i,l) + \mathbf{d}_N^R(j,k)$ since the two sums end up using precisely

$$\mathbf{d}_{ik} + \mathbf{d}_{jl} > \mathbf{d}_{ij} + \mathbf{d}_{kl}$$

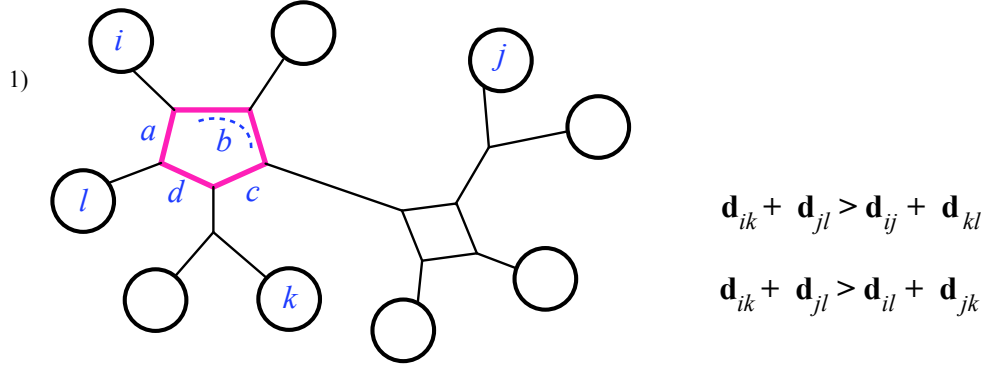$$\mathbf{d}_{ik} + \mathbf{d}_{jl} > \mathbf{d}_{il} + \mathbf{d}_{jk}$$

FIGURE 14. Case 1 of Theorem 4.1: the highlighted edges are the intersection I of the pairwise circuits between leaves i, k and j, l.

the same effective resistances. That is, both $\mathbf{d}_N^R(i, k) + \mathbf{d}_N^R(j, l)$ and $\mathbf{d}_N^R(i, l) + \mathbf{d}_N^R(j, k)$ contain the terms:

$$\frac{c(a+b)}{a+b+c} + \frac{a(b+c)}{a+b+c} + \frac{x(w+y)}{w+x+y} + \frac{w(x+y)}{w+x+y}$$

.

The inequality $\mathbf{d}_N^R(i, k) + \mathbf{d}_N^R(j, l) > \mathbf{d}_N^R(i, j) + \mathbf{d}_N^R(k, l)$ (for the subcase where again i, j reach the same end of I) is easily checked. Here the larger sum contains more terms than the smaller (from the parts of I not in the pairwise circuits for i, j and k, l). As well, when the smaller sum has terms with denominator matching a term in the larger, the numerator is indeed larger in the latter. For instance, in Figure 15, the sum $\mathbf{d}_N^R(i, j) + \mathbf{d}_N^R(k, l)$ has terms:

$$\frac{b(a+c)}{a+b+c} + \frac{y(w+x)}{w+x+y}$$

. The numerator here is exceeded by the terms in $\mathbf{d}_N^R(i, k) + \mathbf{d}_N^R(j, l)$ as just listed above. Finally, notice that there are sub-cases of Case 2 in which the smaller sum will have fewer or no terms at all matching the denominators of terms in the larger sum; these occur when I includes a path at one end or at both ends. See Figure 16 for example.

Case 3: The intersection I is a path. In this case it is quickly verified that the Kalmanson inequality is satisfied as an equality. See Figure 17 for example.

$\square$

The fact that effective resistance distance is a Kalmanson metric immediately suggests that it would be a good candidate for modelling weighted phylogenetic networks. First there is the intuition from experience that if two pathways of heredity exist, the ancestor individual or species will have more in common with the extant individual or species. Thus mutations in the genetic code play the role of resistors to the flow of information.

Secondly, Kalmanson metrics are known to be the only example for which each metric is represented uniquely by a split network. In the case of the resistance distance, the associated unique split network has an additional advantage: it is guaranteed to represent faithfully every split displayed by the original 1-nested network.
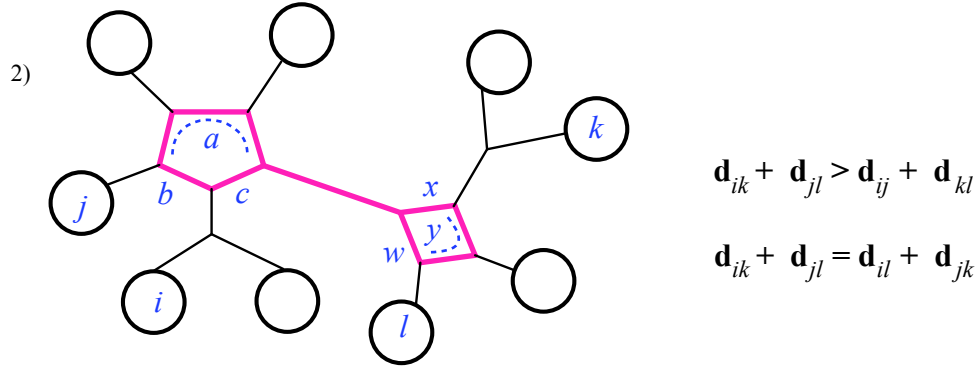
FIGURE 15. Case 2 of Theorem 4.1: the highlighted edges are the intersection I of the pairwise circuits between leaves i, k and j, l.
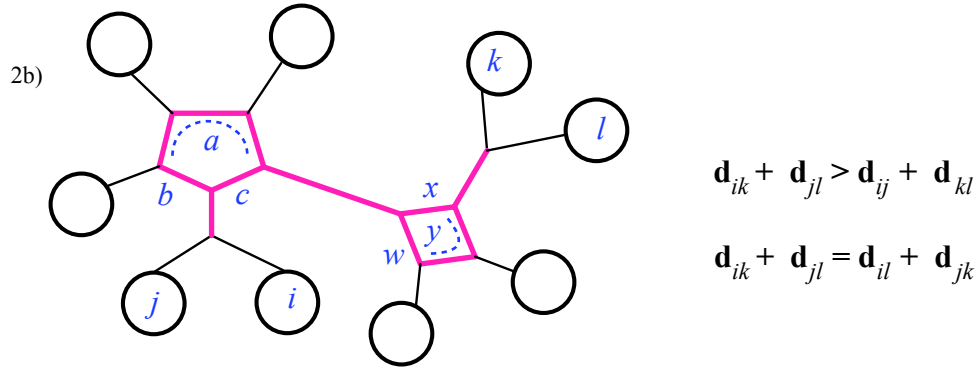
$$\mathbf{d}_{ik} + \mathbf{d}_{jl} > \mathbf{d}_{ij} + \mathbf{d}_{kl}$$

$$\mathbf{d}_{ik} + \mathbf{d}_{jl} = \mathbf{d}_{il} + \mathbf{d}_{jk}$$

FIGURE 16. The highlighted edges are the intersection I of the pairwise circuits between leaves i, k and j, l.

$$\mathbf{d}_{ik} + \mathbf{d}_{jl} > \mathbf{d}_{ij} + \mathbf{d}_{kl}$$

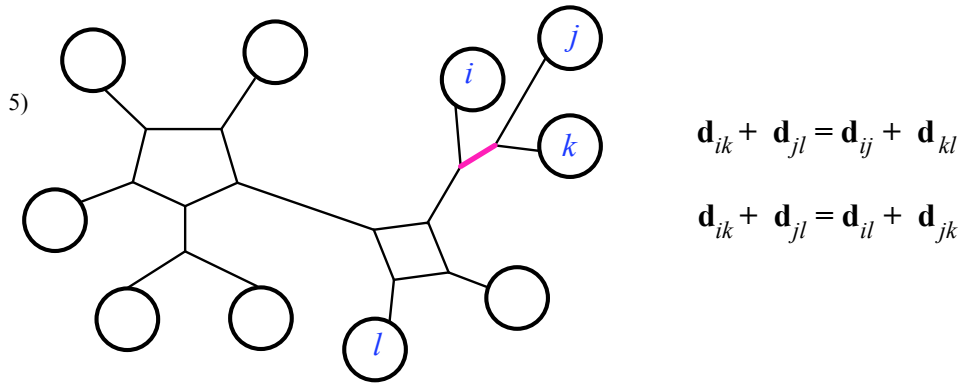$$\mathbf{d}_{ik} + \mathbf{d}_{jl} = \mathbf{d}_{il} + \mathbf{d}_{jk}$$

FIGURE 17. Case 3 of Theorem 4.1: the highlighted edge is the intersection I of the pairwise circuits between leaves i, k and j, l.

$$\mathbf{d}_{ik} + \mathbf{d}_{jl} = \mathbf{d}_{ij} + \mathbf{d}_{kl}$$

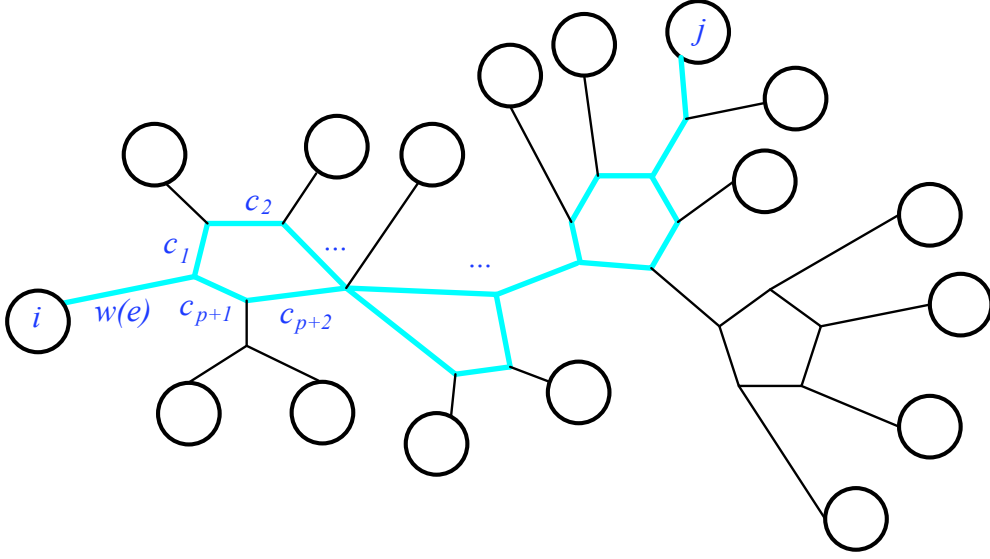$$\mathbf{d}_{ik} + \mathbf{d}_{jl} = \mathbf{d}_{il} + \mathbf{d}_{jk}$$

FIGURE 18. The highlighted subgraph is the pairwise circuit P_{ij}.

Theorem 4.2. *Given a 1-nested phylogenetic network N with positive weighted edges and n leaves, and letting $\mathbf{d}_N^R$ be the resistance metric on the n leaves, then the unique associated split network $\mathcal{N}(\mathbf{d}_N^R)$ displays precisely the same splits as displayed by N.*

Proof. A split $A|B$ displayed by N, is displayed by a single bridge e with weight $w(e)$, a pair of edges both in the same cycle c with respective weights a_c and x_c, or is displayed in more than one way. Let the weight of a specific display of a split in N be $w(e)$ in the first case and $(a_c x_c)/z_c$ in the second case, where z_c is the sum of all the weights in the cycle. We claim: if the split $A|B$ in $\Sigma(\overline{N})$ is assigned the sum of the weights of all distinct displays of that split as displayed in N, then the resulting distance metric $\mathbf{d}$ from the weighted split network thus constructed is indeed $\mathbf{d}_N^R$. Therefore we will conclude, since Theorem 4.1 shows that $\mathbf{d}_N^R$ is Kalmanson, that the weighted split network thus constructed is equal to the unique split network corresponding to $\mathbf{d}_N^R$, as found for instance by the algorithm neighbor-net.

First we check that the claim holds. Consider the pairwise circuit P_{ij} in N for a given pair i, j of leaves. It will be a series of paths and cycles, as seen for example in Figure 18. Thus each cycle c in P_{ij} will be split into two circuit-parallel paths p_c and q_c of respective lengths p, q. Both paths begin and end at the two nodes where that cycle is attached to the rest of the series. Now the resistance distance $\mathbf{d}_N^R(i, j)$ will be the sum of the weights of the (non-circuit-parallel) paths, and of the effective resistances of the circuit-parallel paths. Specifically, every weighted edge of P_{ij} not in a cycle will contribute its weight to the sum, and every weighted edge in a cycle of P_{ij} will appear in one of two factors in the numerator of the term giving the effective resistance from those circuit-parallel paths. We see that

$$\mathbf{d}_N^R(i,j) = \sum_{e \in P_{ij}} w(e) + \sum_{c \in P_{ij}} \frac{(c_1 + \cdots + c_p)(c_{p+1} + \cdots + c_{p+q})}{c_1 + \cdots + c_{p+q}} = \sum_{e \in P_{ij}} w(e) + \sum_{c \in P_{ij}} \sum_{\substack{c_m \in p_c \\ c_r \in q_c}} \frac{c_m c_r}{z_c}$$

where $c_1, \ldots, c_p$ and $c_{p+1}, \ldots, c_{p+q}$ are the weights of the circuit-parallel paths of cycle $c \in P_{ij}$, with $z_c = c_1 + \cdots + c_{p+q}$ being the total weight of c. That is, we expand the numerator of each term from a cycle. Now, the distance metric corresponding to the weighted split network we constructed using $\Sigma(\overline{N})$ has distance

$$\mathbf{d}(i,j) = \sum_{\substack{A|B \in N \\ i \in A, j \in B}} w(A|B)$$

Now splits in N, and thus in $\Sigma(\overline{N})$, which separate leaves i, j are precisely those displayed by a bridge in P_{ij} or by a pair of circuit-parallel edges in a cycle of P_{ij}. Thus using the weights for splits (as stated above):

$$w(A|B) = \sum_{A|B \text{ disp. by } e} w(e) + \sum_{\substack{A|B \text{ disp. by} \\ c_m, c_r \in c}} \frac{c_m c_r}{z_c}$$

in the split metric, gives us the desired claim: $\mathbf{d} = \mathbf{d}_N^R$.

Then we conclude that since the associated weighted split network to the original Kalmanson metric $\mathbf{d}_N^R$ is the unique such network such that the split metric equals the original Kalmanson metric, then $\mathcal{N}(\mathbf{d}_N^R)$ will have precisely the splits of N and thus of $\Sigma(\overline{N})$. $\square$

Remark: The fact that we can take a weighted 1-nested phylogenetic network N and build a weighted circular split network s which has the same metric, $\mathbf{d}_s = \mathbf{d}_N^R$, implies another proof that the resistance distance is Kalmanson. Since the circular split network is planar, and the split metric on it is the same as the minimum path network on it, that metric is guaranteed to be Kalmanson. However, our original proof is nice in that we see which of the inequalities are strict, and which are actually equalities.

The first important implication of these theorems is that the resistance distance on any 1-nested phylogenetic network N is precisely represented by a unique circular split network $\mathcal{N}(\mathbf{d}_N^R)$. Exactly all the splits displayed by the original N are present in $\mathcal{N}(\mathbf{d}_N^R)$. Thus the function L applied to the unweighted version of $\mathcal{N}(\mathbf{d}_N^R)$ returns the unweighted version of N itself.

Theorem 4.3. *Given weighted 1-nested N, we have that $\overline{\mathcal{N}(\mathbf{d}_N^R)} = \Sigma(\overline{N})$. Thus $L(\overline{\mathcal{N}(\mathbf{d}_N^R)}) = \overline{N}$.*

Proof. The first equality follows directly from Theorem 4.2, since neighbor-net is guranteed to output the splits of the unique circular split network associated to the Kalmanson metric given by the resistance distance, which is indeed all the splits displayed by the network N. Then from [12], we have the second equality since $L \circ \Sigma$ is shown there to be the identity map. $\square$

The first application implied by this result is that when using neighbor-net on a measured distance matrix, if we assume that it reflects a resistance distance, we can always recover the form of the original network. The weights of splits in the result of neighbor net are interesting, they are in fact terms in the expansion of the calculated resistance distance. However, the first advantage we see is that the original unweighted phylogenetic network can be directly recovered by taking the exterior of the result of neighbor-net.

As an alternative to neighbor-net, there are polytopes which can serve as the domain for linear programming that finds the best-fit 1-nested phylogenetic network for a measured distance matrix.

5. Implications for Polytopes

In [8] we described for each n a sequence of polytopes that interpolate between the well-known Symmetric Travelling Salesman Polytope ($STSP(n)$) and the Balanced Minimum Evolution Polytope ($BME(n)$). The new polytopes are called the level-1 network polytopes $BME(n, k)$ for $0 \leq k \leq n - 3$. Each is of dimension $\binom{n}{2} - n$. After scaling, all of their vertices are located at barycenters of the faces of $STSP(n)$, and each $BME(n, k)$ is nested inside of $BME(n, j)$ for $j \leq k$. In this nested polytope picture, the largest is $BME(n, 0)$ which is (a scaled version of) $STSP(n)$ and the smallest is $BME(n, n - 3) = BME(n)$. In [12] we looked at implications of the Galois connections studied there for these polytopes, and illustrated a facet with PC-trees. Here we repeat some of the same introductory definitions and remarks, add some information relevant to resistance distance, and illustrate the same facet in Figure 19 but this time with circular split networks.

Definition 5.1. For a binary, 1-nested phylogenetic network N, (weighted or unweighted) the vector $\mathbf{x}(N)$ is defined to have lexicographically ordered components $x_{ij}(N)$ for each unordered pair of distinct leaves $i, j \in [n]$ as follows:

$$x_{ij}(N) = \begin{cases} 2^{k - b_{ij}} & \text{if there exists } c \text{ consistent with } N; \text{ with } i, j \text{ adjacent in } c, \\ 0 & \text{otherwise.} \end{cases}$$

where k is the number of bridges in N and b_{ij} is the number of bridges crossed on any path from i to j.

The convex hull of all the $\mathbf{x}(N)$ such that binary N has k nontrivial bridges is the level-1 network polytope $BME(n, k)$. As shown in [8], the vertices of $BME(n, k)$ are precisely the vectors $\mathbf{x}(N)$ for N binary with n leaves and k nontrivial bridges. In light of Theorems 4.1 and 4.2, we can characterize the vertices as follows:

Theorem 5.2. *The vertices of the polytope $BME(n, k)$ correspond to the unweighted binary 1-nested phylogenetic networks with k non-trivial bridges that are the images $L(\overline{R_w(N)})$ for weighted 1-nested networks N with n leaves.*

Proof. This is a direct result of the fact if N has its non-trivial bridges all with both nodes of degree three, then $L(\overline{R_w(N)})$ will be a binary 1-nested network, from Theorem 4.2. The function will introduce bridges that separate all cycles. $\square$

Also as shown in [8], an equivalent definition of the vector $\mathbf{x}(N)$ is the vector sum of the vertices of the $STSP(n)$ which correspond to cyclic orders consistent with N. The vertices of $STSP(n)$ are the incidence vectors $\mathbf{x}(c)$ for each cyclic order c of n, where the i, j component is 1 for i and j adjacent in the order c, 0 otherwise. This alternative definition may be applied to any 1-nested phylogenetic network, not just the binary ones.

Definition 5.3. For a 1-nested phylognetic network N, the vector $\mathbf{x}(N) = \sum_c \mathbf{x}(c)$ where the sum is over all cyclic orders of $[n]$ consistent with N.

Note that for phylogenetic trees t (with nodes of any degree), Definition 5.3 of $\mathbf{x}(t)$ agrees with the definition of the coefficient n_t in [17], in the proof of Theorem 4.2 of that paper.

Two polytopes are *nested* when one is contained in the other, with all vertices of the smaller on faces of the larger. In [8] it is shown that for any n the scaled polytopes $(2^{n-3-k})\text{BME}(n,k)$ are sequentially nested, from $k = 0$, the largest, to $k = n - 3$, the smallest. Each vertex of a smaller scaled polytope is at the barycenter of a face of $\text{BME}(n,0)$. Figure 19 shows a facet of BME(5,0) which corresponds to the split $\{1,2\}|\{3,4,5\}$. This facet is scaled so that scaled vertices of $\text{BME}(5,1)$ are seen as barycenters of faces.

In [12] it is shown that the minimum path distance vector for a 1-nested phylogenetic network may be seen as a linear functional, and that it is minimized over the $\text{BME}(n,k)$ polytope. Specifically, from [12],

Theorem 5.4. *Given any weighted 1-nested phylogenetic network N with n leaves, the product $\mathbf{x}(\hat{N}) \cdot \mathbf{d}_N$ is minimized over $BME(n,k)$ precisely for the unweighted binary networks $\hat{N}$ with k bridges such that $\overline{S_w(N)} \leq \Sigma(\hat{N})$.*

Now we can extend that result to resistance distances. In fact it becomes stronger: binary networks can be directly recovered even when they have long edges, since the action of R_w preserves all splits. Precisely, we have:

Theorem 5.5. *Minimizing $\mathbf{x}(N) \cdot \mathbf{d}_N^R$ is achieved at the face of $BME(n,k)$ with vertices $\mathbf{x}(\hat{N})$, for unweighted binary networks $\hat{N}$ with k bridges such that $\hat{N}$ refines $\overline{N}$.*

Proof. We claim that $\mathbf{x}(N) \cdot \mathbf{d}_N^R$ is the same as $\mathbf{x}(N) \cdot \mathbf{d}_N'$ for $N' = L_w(R_w(N))$. That is because the leaves which are adjacent in some circular order consistent with N and thus in $R_w(N)$ have distance between them which is the sum of the splits that separate them. Since those leaves are adjacent, the shortest path of splits between them will lie on the exterior of $R_w(N)$. In fact, for adjacent i, j an edge of a cycle on the path between them with weight a, contributes $\frac{a(b+c+d+...)}{a+b+c+d+...}$ to $\mathbf{d}_N^R(i,j)$, where the other edges of that cycle have weights $b, c, d, ...$. Bridges e between them contribute their weights $w(e)$. These values are the same as those for the splits displayed between i, j, seen in the proof of Theorem 4.2. Therefore:

$$\mathbf{x}(N) \cdot \mathbf{d}_N^R = \sum_c \mathbf{x}(c) \cdot \mathbf{d}_n^R$$

$$= \sum_c \mathbf{x}(c) \cdot \mathbf{d}_s, \text{ for } s = R_w(N)$$

$$= \sum_c \mathbf{x}(c) \cdot \mathbf{d}_{N'} \text{ for } N' = L_w(R_w(N))$$

$$= \mathbf{x}(N) \cdot \mathbf{d}_{N'}$$

We know from Theorems 8, 9, and 11 of [9] that for any weighted 1-nested phylogenetic network M with n leaves, the product $\mathbf{x}(\hat{M}) \cdot \mathbf{d}_M$ is minimized over $\text{BME}(n,k)$ precisely for binary networks $\hat{M}$ with k bridges such that $\overline{M} \leq \hat{M}$.

Thus in our case we have $\mathbf{x}(\hat{N}) \cdot \mathbf{d}_{N'}$ is minimized over $\text{BME}(n,k)$ precisely for the unweighted binary networks $\hat{N}$ with k bridges such that $\overline{N'} \leq \hat{N}$. Here, $\overline{(N')} = \overline{N}$, since $\overline{L_w(R_w(N))} = \overline{N}$. The inequality here is refinement. □
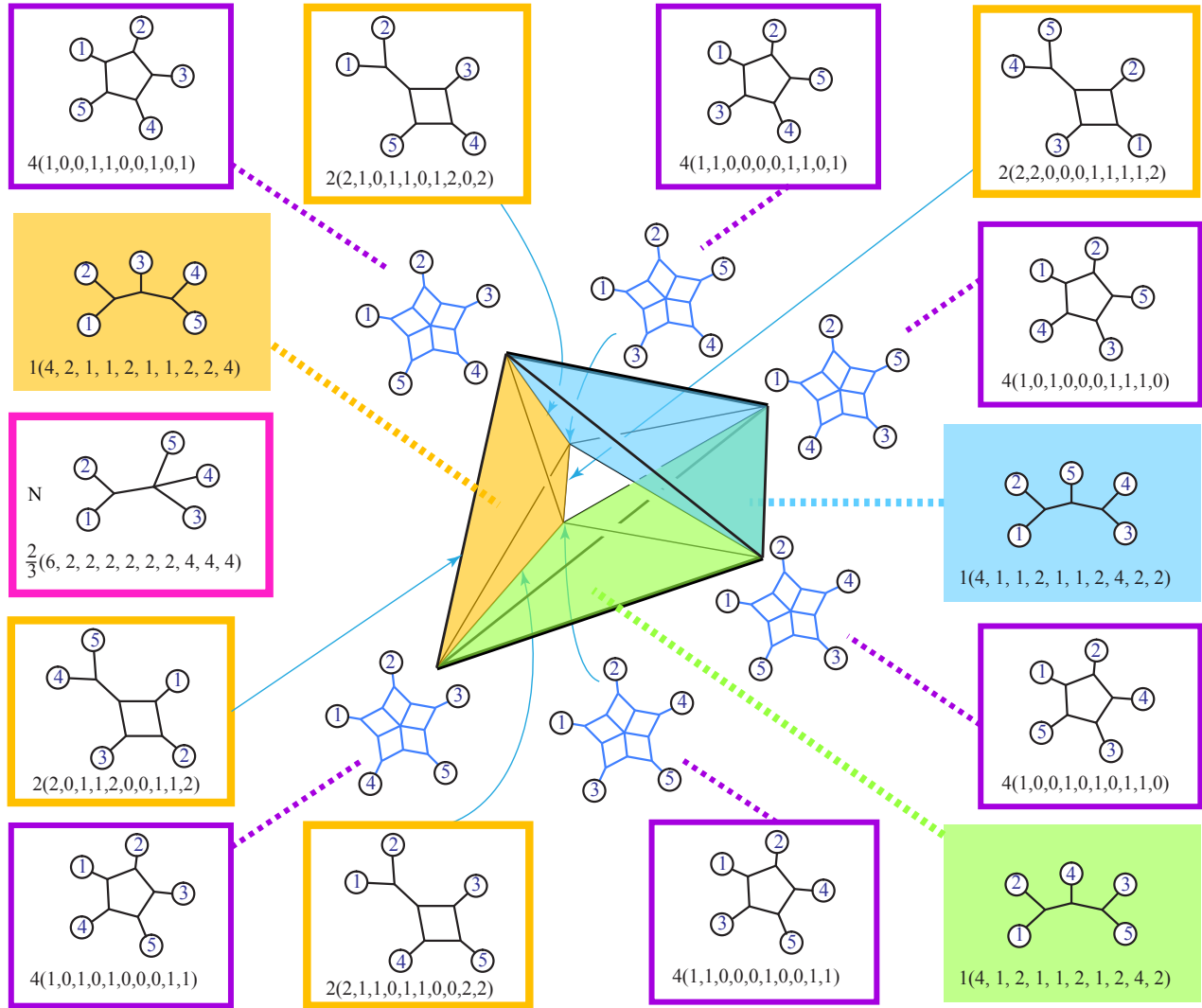
FIGURE 19. A scaled split facet of BME(5,0) = STSP(5). This 4D facet corresponds to the split $\{1, 2\}|\{3, 4, 5\}$ (pictured as the tree N, center-left), and is also known as a subtour elimination facet. Three of its tetrahedral subfaces correspond to networks, and are shaded. Vertices and some faces are labeled with both 1-nested networks and split networks. A (scaled version of) the vector $\mathbf{x}$ is shown beneath each network: the barycenter of the face represented by that network.

For example compare Figures 10 and 20. It is easily checked that although $\mathbf{d}_{N'} \neq \mathbf{d}_N^R$, we have $\mathbf{x}(N) \cdot \mathbf{d}_{N'} = \mathbf{x}(N) \cdot \mathbf{d}_N^R = 51.4$

The implication then is that using either linear programming on BME(n, 0) or neighbor-net, assuming that the resistance metric is valid, the resulting split network gives the true exterior form of the original 1-nested phylogenetic network.
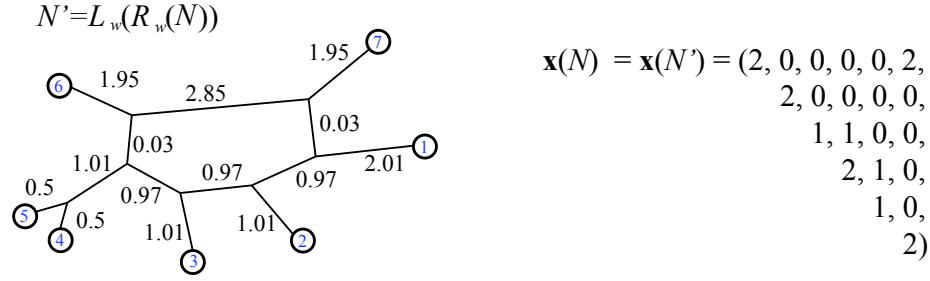
$N'=L_w(R_w(N))$

$\mathbf{x}(N) = \mathbf{x}(N') = (2, 0, 0, 0, 0, 2,$
$2, 0, 0, 0, 0,$
$1, 1, 0, 0,$
$2, 1, 0,$
$1, 0,$
$2)$

FIGURE 20. Using N from Figure 10, we find N' as in the proof of Theorem 5.5. Here

$$\mathbf{d}_N = (3.99, 4.96, 6.43, 6.43, 6.84, 3.99,$$
$$2.99, 4.46, 4.46, 4.93, 3.96,$$
$$3.49, 3.49, 3.96, 4.93,$$
$$1, \ 3.49, 6.34,$$
$$3.49, 6.34,$$
$$6.75)$$

.

6. COUNTEREXAMPLES AND CONJECTURES

6.1. **Indistinguishable weightings.** Resistance distance metrics on a 1-nested phylogenetic network are not in bijection with edge weightings, but the split-equivalence class is an invariant of those edge weights. That is, if two networks N and N' have the same resistance distance metric $\mathbf{d}_N^R = \mathbf{d}_{N'}^R$, this does not imply that $N = N'$, but it does imply that $\overline{N} = \overline{N'}$. The latter fact is implied by Lemma 4.1 and the theorems of Section 4, and we can see the former fact via counterexample. In Figure 21 we show two weighted phylogenetic networks with 4 leaves, called N and N'. Their resistance distances between leaves are identical:

$$\mathbf{d}_N^R = \mathbf{d}_{N'}^R = \left(\frac{122}{23}, \frac{178}{23}, \frac{108}{23}, \frac{198}{23}, \frac{168}{23}, \frac{176}{23} \right).$$

Note that we do see that $\overline{N} = \overline{N'}$. There are 7 split-classes of 1-nested phylogenetic networks on 4 leaves, and our theorems show that none of the other 6 classes can be given edge weights that yield this same resistance distance metric on four leaves.

6.2. **Non-Kalmanson networks.** Not all resistance distances are Kalmanson, even when restricted to phylogenetic networks. For a counterexample, consider the network N formed by having 6 leaves attached to the 6 vertices of the complete bipartite graph $K_{3,3}$, pictured in Figure 22. The resistance distance metric for complete bipartite graphs is found in [15]. Consider that $K_{m,n}$ is the graph join if two edgeless graphs: $K_{m,n} = \overline{K}_m + \overline{K}_n$ with unit weight for each edge. Then the resistance distance on $K_{m,n}$ is $2/n$ for vertices that have no edge between them (they are both the same color), and $(m + n - 1)/mn$ for vertices with an edge between them [15]. For our example N, let the two (same-colored) parts of the graph (3 nodes
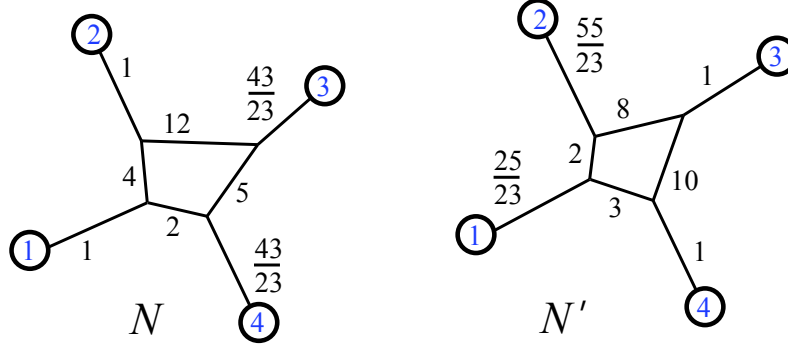
FIGURE 21. Two weighted phylogenetic networks with identical resistance distances for their leaves.
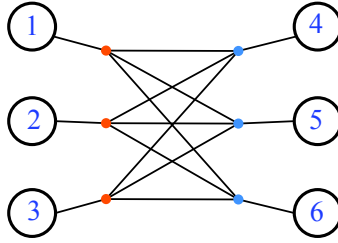

FIGURE 22. A phylogenetic network with non-Kalmanson resistance distance. All the edge lengths are 1.

each, say red and blue) be attached to the leaves $\{1, 2, 3\}$ and $\{4, 5, 6\}$ respectively. Letting each edge have weight 1, we find the resistance distance between any two leaves attached to the same colored part is $2 + 2/3 = 8/3$, while the distance between any two leaves, with one attached to each part, is $2 + 5/9 = 23/9$. In any circular order of the leaves, there will be a sub-sequence i, j, k, l where the first two leaves i, j are attached to the same color, and the second two k, l are both attached to the other color. Thus $\mathbf{d}_N^R(i, j) + \mathbf{d}_N^R(k, l) = 16/3 = 48/9$ which is larger than $\mathbf{d}_N^R(i, k) + \mathbf{d}_N^R(j, l) = 46/9$. This counterexample raises the question of necessary conditions for a network with resistance distance to be Kalmanson.

6.3. **Outer Planarity.** We conjecture that outer planarity is a sufficient condition for Kalmanson: that if a weighted pylogenetic network can be drawn in the plane with its leaves on the exterior that the resistance distance is Kalmanson. We note that it this condition is not necessary: it can be checked that the complete graph K_5 with unit edges has the Kalmanson property.

6.4. **Faithfully phylogenetic Kalmanson distance vectors.** Following the terminology in Definition 2.7, we call a Kalmanson distance vector $\mathbf{d}$ *faithfully phylogenetic* if the unique

$$\mathbf{d}_{N'}^R = \mathbf{d}_N^R = \mathbf{d}_s = \left(\frac{50}{15}, \frac{51}{15}, \frac{41}{15}, \frac{41}{15}, \frac{51}{15}, \frac{50}{15} \right)$$
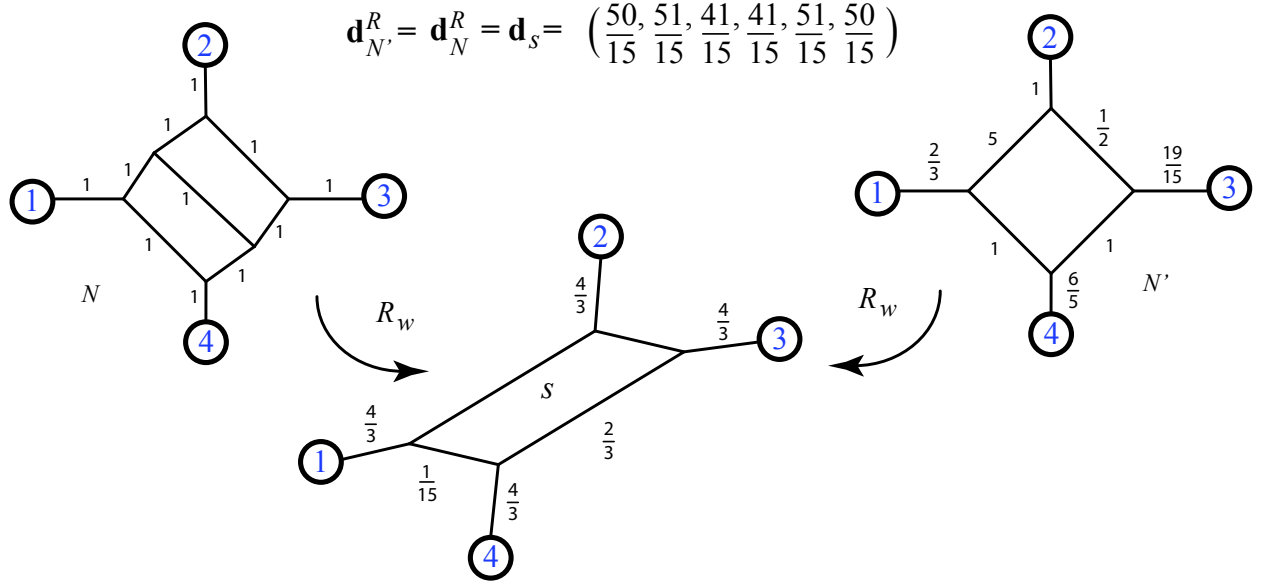
FIGURE 23. Two weighted phylogenetic networks with identical resistance distances for their leaves, and their common split network.

circular split network associated to $\mathbf{d}$ is in the range of Σ (after forgetting weights). We conjecture that faithfully phylogenetic Kalmanson distance vectors always arise from resistance distances. Specifically we conjecture that if $\mathbf{d}$ is faithfully phylogenetic, then $\mathbf{d} = \mathbf{d}_N^R$ for some weighted phylogenetic network N.

Note that not all Kalmanson distance vectors arise from resistance distances, simply due to the fact that not all circular split networks are in the range of Σ.

6.5. **2-nested Kalmanson networks.** An easier conjecture, perhaps, is that 2-nested phylogenetic networks have Kalmanson resistance distance. This appears to be the case, experimentally. For instance in Figure 23 we show a simple 2-nested network N whose resistance distance is clearly Kalmanson: in fact it is the same resistance distance as possessed by the shown 1-nested network.

6.6. **Indistinguishable weightings and invariants.** We conjecture that for every weighted 2-nested network there is a weighted 1-nested network with matching resistance distance. Again see Figure 23. However, in light of the above conjecture 6.3, we conjecture that the exterior shape of networks is an invariant of resistance distance: specifically that if any two outer planar networks N, N' have $\mathbf{d}_N^R = \mathbf{d}_{N'}^R$ then $L(\overline{R_w(N)}) = L(\overline{R_w(N')})$.

6.7. **Limiting case.** Consider when an edge in a cycle of N has a very large weight, or high resistance. As this weight grows, the limit of $L_w(R_w(N))$ approaches a network with that edge being deleted entirely. We see this by considering any two circuit-parallel paths with resistance R_1 and R_2 the first of which uses an edge with variable weight w (all other weights constant). Then letting $w \to \infty$ implies $R_1 \to \infty$ and thus $R_1 R_2/(R_1 + R_2)$ approaches R_2 by L'Hospital's rule. Thus as w goes to ∞ we see that the resistance distances using those circuit-parallel paths reduce to the path distances, and so the distance metric from that network approaches one
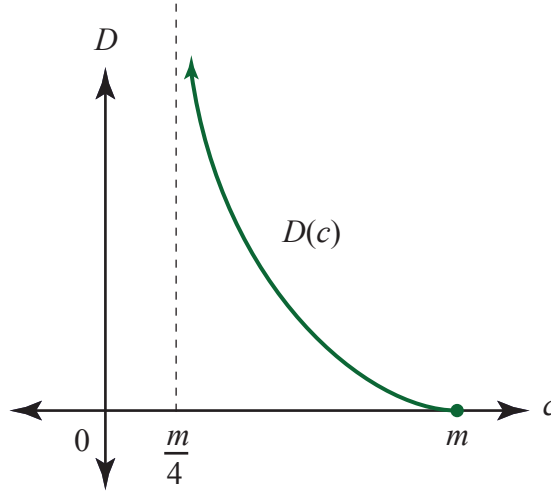
FIGURE 24. Calculated Jukes-Cantor distance D as a function of the number of matching sites c in aligned sequences of length m.

without that edge. This is similar to the way in which S_w, which uses the minimal path distance on N, serves to delete some edges as seen in Figure 10.

7. DISTANCE MEASURES

A question is raised about the mathematics which precedes the work described in this paper: what sort of measurement should actually yield the experimental resistance distances in a real example? What should play the role of attaching the ohmmeter to pairs of wires? Usually, DNA sequences of length m are aligned (a multi-step problem of its own) and then the number of disagreeing sites is counted. Let p be the proportion of disagreements to the length m of the sequence: $p = (m - c)/m$ where c is the number of correct, matching sites. Then there is a selection of mutation models, such as the simplest Jukes-Cantor model, which predict a distance D which is the expected total number of mutations. Experimentally we find that distance D as a function of the observed disagreements. Alternately we could choose D from the list of evolutionary models: for instance

$$D = K = -\frac{1}{2} \ln((1 - 2p - q)\sqrt{1 - 2q})$$

for Kimura's two parameter model. Or, alignment-free models such as the k-mer distance measures as described in [1].

Here, we would want a distance $D = R$ which is summed when in sequence but obeys the Ohm equations. The answer will depend both on the model of mutation we choose and the model of recombination we choose. For instance, $D = -\frac{3}{4} \ln(1 - \frac{4}{3}p)$ for the Jukes-Cantor model, as described in [14]. Rewriting using $p = (m - c)/c$ we have:

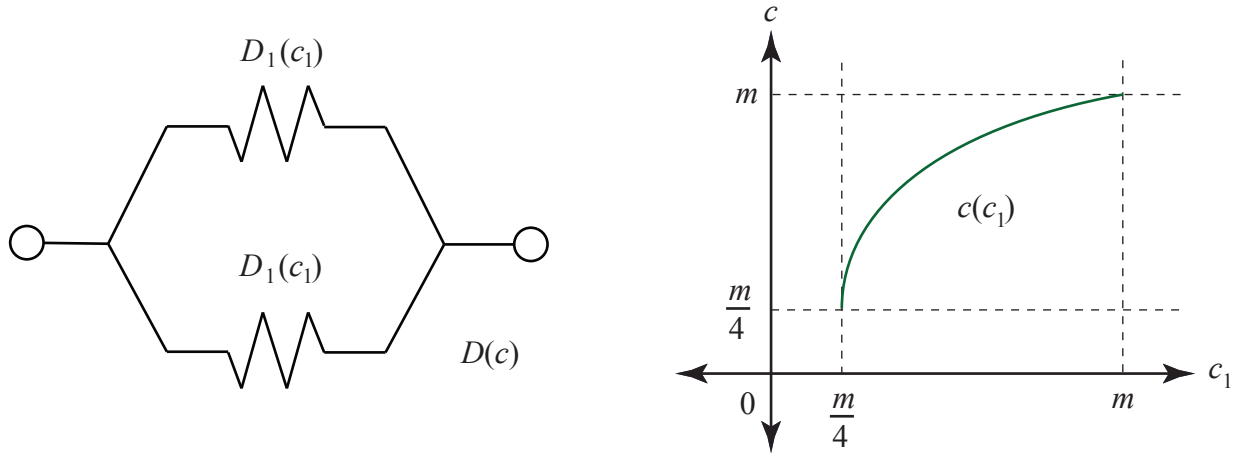$$D(c) = \frac{3}{4} \ln \left(\frac{3m}{4c - m} \right)$$

FIGURE 25. On the left is a simple parallel circuit with identical resistance on each branch. If the resistance is the Jukes-Cantor distance and obeys the Ohm laws, then the number c of matching sites at the end of the circuit will depend on the number c_1 of correct matching sites at the end of each branch before recombination.

D has the graph in Figure 24. The c-axis is explained by the fact that in the Jukes-Cantor model, mutations of the 4 nucleotides A, G, T, C can replace any letter with another—including a self replacement. This implies that the smallest number of matching sites is $\frac{m}{4}$, while the largest is m. We can use D for the resistance distance only if there is experimental evidence that for circuit-parallel paths we have $D = D_1 D_2/(D_1 + D_2)$, where $D_1(c_1)$ and $D_2(c_2)$ are the distances for each path, in expected numbers of mutations as a function of correct matching sites. There are certainly some features of D that look promising, including the shape of its graph above: resistance typically ranges from 0 to infinity. Assuming that the formula for D over the circuit-parallel paths does hold, when one of the circuit-parallel resistances is infinite: say $D_1 \to \infty$; then we see that $D \to D_2$. Similarly, as $c_1 \to m/4$, we have that c, the number of correct sites after recombination, approaches c_2.

When both branches have the same distance $D_1 = D_2$, and it obeys Ohm's law, we see the total resistance $D = D_1/2$. Using the formula for $D(c)$ and $D_1(c_1)$ and solving for c we get the following function, graphed in Figure 25:

$$c = \frac{m}{4} + \sqrt{3\left(\frac{m}{4}c_1 - \left(\frac{m}{4}\right)^2\right)}.$$

Thus as a first check the geneticist could compare two genomes and their hybrid genome with a common ancestor. When the two are close to the same distance from the common ancestor (both have c_1 matching sites), then the pair (c_1, c) for c the number of matches between the hybrid and the common ancestor might fit the parabola as seen in Figure 25. If that fit is achieved, then it would be reasonable to apply the theorems of this paper.

References

1. Elizabeth S. Allman, John A. Rhodes, and Seth Sullivant, *Statistically consistent k-mer methods for phylogenetic tree reconstruction*, J. Comput. Biol. **24** (2017), no. 2, 153–171. MR 3607847
2. L. Billera, S. Holmes, and K. Vogtmann, *Geometry of the space of phylogenetic trees*, Adv. in Appl. Math. **27** (2001), no. 4, 733–767. MR 1867931
3. David Bryant, Vincent Moulton, and Andreas Spillner, *Consistency of the neighbor-net algorithm*, Algorithms for Molecular Biology **2** (2007), no. 1, 8.
4. E. B. Curtis, D. Ingerman, and J. A. Morrow, *Circular planar graphs and resistor networks*, Linear Algebra Appl. **283** (1998), no. 1-3, 115–150. MR 1657214
5. Edward B. Curtis and James A. Morrow, *Determining the resistors in a network*, SIAM J. Appl. Math. **50** (1990), no. 3, 918–930. MR 1050922
6. _____, *The Dirichlet to Neumann map for a resistor network*, SIAM J. Appl. Math. **51** (1991), no. 4, 1011–1029. MR 1117430
7. Andreas Dress, Katharina T. Huber, Jacobus Koolen, Vincent Moulton, and Andreas Spillner, *Basic phylogenetic combinatorics*, Cambridge University Press, Cambridge, 2012. MR 2893879
8. Cassandra Durell and Stefan Forcey, *Level-1 phylogenetic networks and their balanced minimum evolution polytopes*, J. Math. Biol. **80** (2020), no. 5, 1235–1263. MR 4071414
9. _____, *Level-1 phylogenetic networks and their balanced minimum evolution polytopes*, J. Math. Biol. **80** (2020), no. 5, 1235–1263. MR 4071414
10. Vladimir Ejov, Jerzy A Filar, Michael Haythorpe, John F Roddick, and Serguei Rossomakhine, *A note on using the resistance-distance matrix to solve hamiltonian cycle problem*, arxiv:1902.10356 (2019).
11. M. Erné, J. Koslowski, A. Melton, and G. E. Strecker, *A primer on Galois connections*, Papers on general topology and applications (Madison, WI, 1991), Ann. New York Acad. Sci., vol. 704, New York Acad. Sci., New York, 1993, pp. 103–125. MR 1277847
12. Stefan Forcey and Drew Scalzo, *Galois connections for phylogenetic networks and their polytopes*, arXiv/abs/2004.11944 (2020).
13. P. Gambette, K. T. Huber, and G. E. Scholz, *Uprooted phylogenetic networks*, Bull. Math. Biol. **79** (2017), no. 9, 2022–2048. MR 3685182
14. T.H. Jukes and C.R. Cantor, *Evolution of protein molecules*, Mammalian protein metabolism (H. N. Munro, ed.), Academic Press, 1969, pp. 21–132.
15. D. J. Klein and M. Randić, *Resistance distance*, vol. 12, 1993, Applied graph theory and discrete mathematics in chemistry (Saskatoon, SK, 1991), pp. 81–95. MR 1219566
16. D. Levy and Lior Pachter, *The neighbor-net algorithm*, Advances in Applied Mathematics **47** (2011), 240–258.
17. Charles Semple and Mike Steel, *Cyclic permutations and evolutionary trees*, Adv. in Appl. Math. **32** (2004), no. 4, 669–680. MR 2053839 (2005g:05042)
18. Mike Steel, *Phylogeny—discrete and random processes in evolution*, CBMS-NSF Regional Conference Series in Applied Mathematics, vol. 89, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2016. MR 3601108
19. Yujun Yang and Douglas J. Klein, *Resistance distance-based graph invariants of subdivisions and triangulations of graphs*, Discrete Appl. Math. **181** (2015), 260–274. MR 3284531
20. _____, *Two-point resistances and random walks on stellated regular graphs*, J. Phys. A **52** (2019), no. 7, 075201, 18. MR 3916426

URL: http://www.math.uakron.edu/~sf34/