arXiv:2204.11987v1 [math.CO] 25 Apr 2022

INHERENT SYMMETRIES OF GRAPHS, PATHS, AND TRAVELING SALESPERSON PROBLEMS

DONALD G. SAARI

Abstract. Without imposing restrictions on a weighted graph's arc lengths, symmetry structures cannot be expected. But, they exist. To find them, the graphs are decomposed into a component that dictates all closed path properties (e.g., shortest and longest paths), and a superfluous component that can be removed. The simpler remaining graph exposes inherent symmetry structures that form the basis for all closed path properties. For certain asymmetric problems, the symmetry is that of three-cycles; for the general undirected setting it is a type of four-cycles; for general directed problems with asymmetric costs, it is a product of three and four cycles. Everything extends immediately to incomplete graphs.

1. Introduction

With applications ranging from the design of microchips to the positioning of telescopes [2], understanding properties of a weighted graph's paths and closed paths has attained importance beyond mathematics. What complicates this analysis is that transmission costs between nodes typically include factors that differ from what is needed to determine path properties; e.g., they may reflect the problem's topography or congestions of various types. These portions of a graph's entries add nothing to the analysis, but they contribute to the complexity of these concerns.

The mathematical structure of graphs developed here separates a complete weighted graph into two unique components. The first, with best possible reduced degrees of freedom, has all of the information needed to develop the particulars of paths and closed paths. The remaining component is dismissed because it adds no value; it just complicates both the analysis of paths and the behavior of algorithms. This approach resembles (and is motivated by) a game theory decomposition [3] where one game component has only (and all) information needed to find all pure and mixed Nash strategic properties; another component captures coordination, cooperation, etc.

Three classes of graphs are examined. The first is a directed, asymmetric setting measuring differences from the average cost between vertices. The second and third are, respectively, the standard undirected symmetric cost and directed asymmetric cost settings. Symmetry structures for these classes differ; e.g., the symmetry structures for the excess cost graphs (Sect. 2) are three-cycles. Symmetries for the standard symmetric case (Sect. 3) are a form of four-cycles. Symmetries for the general asymmetric costs (Sect. 4) are a product of three and four cycles.

To describe the basic theme in terms of the first class, it turns out that these graphs can be embedded in the space of asymmetric paired comparisons. A "decision theory" decomposition divides this space into a linear subspace characterized by a strong form of transitivity and its normal bundle consisting of cycles [4]. Voting methods seek linear orders, so the cyclic components create complexities and paradoxical outcomes (e.g., Arrow's Impossibility Theorem [1]). Projecting

My thanks to George Hazelrigg for our several discussions. This work is part of a National Science Foundation project under NSF Award Number CMMI-1923164.

the data to the transitive subspace eliminates these difficulties and simplifies the analysis [4]. But cycles, not linear orders, are central for closed paths, so in this setting the transitive components are what obscure the analysis. Projecting the data to the cyclic subspace lowers the degrees of freedom, removes trouble-causing components, and uncovers the system's inherent three-cycle symmetry. Here, a A, B, C cycle's costs of going from A to B, B to C, and C to A are identical.

In general, each class of graphs is decomposed into a component characterized by a behavior that masks the closed path structures and a component that has only (and all) of the relevant closed path information. Everything extends to incomplete graphs. Most proofs are in Sect. 6.

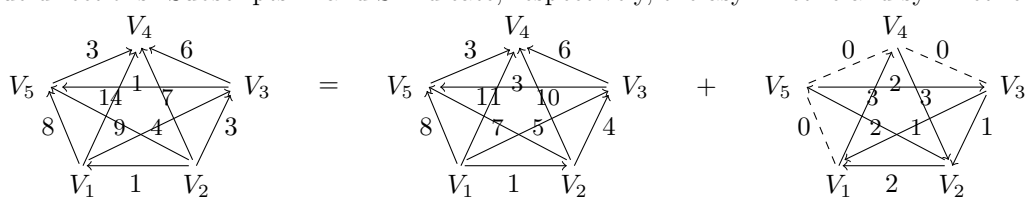
2. Asymmetric excess costs

Reimbursing a salesperson for the average cost of traveling between cities creates an incentive to find routes with below average costs. For notation, it takes 40 minutes to walk from home, H, to campus, C; returning uphill requires 50 minutes, so the average is 45. The "excess cost function" registers differences from the average where $C \xrightarrow{5} H$ represents both $C \xrightarrow{5} H$ and $H \xrightarrow{-5} C$.

Graphs in the space of asymmetric weighted, n-vertex graphs (with no loops) considered here, $\mathbb{G}^n_A$, are complete (i.e., each pair of vertices is connected with paths) and

$$V_j \xrightarrow{x} V_k \text{ represents both } V_j \xrightarrow{x} V_k \text{ and } V_k \xrightarrow{-x} V_j. \tag{1}$$

To simplify the graphs, only an arc's positive cost direction need be represented; this is because moving counter to an arrow represents a "below average" cost (Eq. 1). With this choice, V_2 in Fig. 1a is a "source" as all positive value directions point away; it is a "sink" with the negative value directions. Conversely, V_4 is a sink with positive value directions and a source for negative value directions. Subscripts A and S indicate, respectively, the asymmetric and symmetric cases.



a. Graph $\mathcal{G}^5_A$ **b.** Closed path independent; $\mathcal{G}^5_{A,cpi}$ **c.** Cyclic component; $\mathcal{G}^5_{A,cyclic}$

Figure 1. Decomposition

Figure 1 depicts the general approach whereby graph $\mathcal{G}^5_A$ is uniquely decomposed into a "closed path independent" component $\mathcal{G}^5_{A,cpi}$ (defined in Def. 1) and a cyclic component $\mathcal{G}^5_{A,cyclic}$ to have $\mathcal{G}^5_A = \mathcal{G}^5_{A,cpi} + \mathcal{G}^5_{A,cyclic}$. The goal is to achieve this decomposition for any $\mathcal{G}^n_A \in \mathbb{G}^n_A$ to obtain

$$\mathcal{G}^n_A = \mathcal{G}^n_{A,cpi} + \mathcal{G}^n_{A,cyclic}. \tag{2}$$

Definition 1. *Graph $\mathcal{G}^n_{A,cpi} \in \mathbb{G}^n_A$ is "closed path independent" (cpi) iff all closed paths have length zero. A graph is strongly transitive iff path lengths of a triplet $\{V_i, V_j, V_k\}$ satisfy*

$$V_i \xrightarrow{x} V_j \xrightarrow{y} V_k = V_i \xrightarrow{z=x+y} V_k. \tag{3}$$

Both Eq. 3 paths start at V_i and end at V_k, so equality designates equal path lengths. This equation modifies the concept of "strong transitivity" developed for decision theory [4].

Theorem 1. *A graph is strongly transitive iff it is cpi. Strongly transitive graphs (equivalently, cpi graphs) with n vertices define a $(n-1)$-dimensional linear subspace $\mathbb{ST}^n_A \subset \mathbb{G}^n_A$.*

The Fig. 1b graph is strongly transitive and cpi. To check for strong transitivity, select any triplet, say $\{V_1, V_3, V_5\}$, and determine whether this triangle's leg lengths, $V_1 \xrightarrow{5} V_3, V_3 \xrightarrow{3} V_5$, and $V_1 \xrightarrow{8} V_5$, satisfy the *triangle equality* Eq. 3, which they do. To equate strong transitivity with cpi, reversing $V_1 \xrightarrow{8} V_5$ defines the closed path $V_1 \xrightarrow{5} V_3 \xrightarrow{3} V_5 \xrightarrow{-8} V_1$ with zero length.

While the proof of Thm. 1 is in Sect. 6, proving that $\mathbb{ST}^n_A$ is a linear subspace is a common exercise. For the dimensionality assertion, strong transitivity ensures that the $V_i \to V_j$ arc length equals the $V_i \xrightarrow{x} V_1 \xrightarrow{y} V_j$ length, where the $V_i \to V_j$ path is diverted to pass through V_1. As all arc lengths for $\mathcal{G}^n_{A,cpi} \in \mathbb{ST}^n_A$ are determined by the $\{V_1 \to V_k\}_{k=2}^n$ arc lengths, $\mathbb{ST}^n_A$ has dimension $(n-1)$.

A standard induction argument applied to Eq. 3 proves the following result.

Corollary 1. *For $\mathcal{G}^n_{A,cpi} \in \mathbb{ST}^n_A$, any path starting at V_i and ending at V_j has length equal to the $V_i \to V_j$ arc that connects the endpoints.*

A Fig. 1b example of Cor. 1 is where the 4 length of $V_2 \xrightarrow{7} V_5 \xrightarrow{3} V_4 \xrightarrow{-10} V_2 \xrightarrow{10} V_6 \xrightarrow{-6} V_3$, where vertices can be revisited, equals the arc length connecting the endpoints $V_2 \xrightarrow{4} V_3$.

2.1. **Cyclic Normal Bundle.** As Thm. 3 will assert, the $\mathcal{G}^n_{A,cpi} \in \mathbb{ST}^n_A$ component of $\mathcal{G}^n_A$ (see Eq. 2) blurs the closed path properties of $\mathcal{G}^n_A$. Thus, path properties must be based on the structure of $\mathbb{ST}^n_A$'s normal bundle. The $\mathbb{G}^n_A$ and $\mathbb{ST}^n_A$ dimensions are $\binom{n}{2}$ and $(n-1)$, so $\mathbb{ST}^n_A$'s normal subspace, $\mathbb{C}^n_A$, has dimension $\binom{n-1}{2}$. As described next, $\mathbb{C}^n_A$ consists of cyclic actions.

Theorem 2. [4] *For $n \geq 3$, the linear subspace orthogonal to $\mathbb{ST}^n_A$, $\mathbb{C}^n_A$, has dimension $\binom{n-1}{2}$. A basis for $\mathbb{C}^n_A$, which consists of three-cycles with equal costs between successive vertices, is*

$$\{V_1 \xrightarrow{1} V_j \xrightarrow{1} V_k \xrightarrow{1} V_1\}_{1<j<k\leq n}. \tag{4}$$

The Eq. 4 cycles are anchored at one vertex, so the following offers a more general choice.

Corollary 2. *If $\mathcal{CB}^n_A$ has $\binom{n-1}{2}$ three-cycles where each arc in a three-cycle has length 1 and each three-cycle has one arc that is not in any other $\mathcal{CB}^n_A$ three-cycle, then $\mathcal{CB}^n_A$ is a $\mathbb{C}^n_A$ basis.*

According to Thm. 2, the $\mathcal{G}^n_{A,cyclic} \in \mathbb{C}^n_A$ structure is governed by three-cycles. To motivate their Eq. 4 form, strong transitivity requires $V_1 \xrightarrow{x} V_j \xrightarrow{y} V_k = V_1 \xrightarrow{z} V_k$, which defines the equation $x+y-z=0$. This equation has the normal vector $(1,1,-1)$, which, when expressed in a path form, is $V_1 \xrightarrow{1} V_j \xrightarrow{1} V_k \xrightarrow{1} V_1$, or an Eq. 4 three-cycle. It follows that the multiple of a three-cycle measures how this triplet's $\mathcal{G}^n_A$ data portion deviates from the triplet's cpi "sameness."

This discussion leads to the following central result where Eq. 5 asserts that the Eq. 2 goal has been realized. The theorem's concluding statement is crucial for what follows in this section.

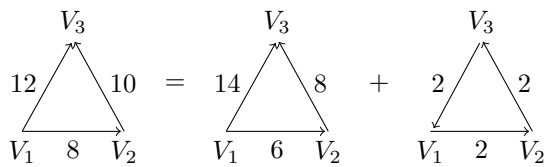
Theorem 3. *Space $\mathbb{G}^n_A$ is divided into a linear subspace $\mathbb{ST}^n_A$ and its orthogonal complement $\mathbb{C}^n_A$. For $\mathcal{G}^n_A \in \mathbb{G}^n_A$, there are unique $\mathcal{G}^n_{A,cpi} \in \mathbb{ST}^n_A$ and $\mathcal{G}^n_{A,cyclic} \in \mathbb{C}^n_A$ so that*

$$\mathcal{G}^n_A = \mathcal{G}^n_{A,cpi} + \mathcal{G}^n_{A,cyclic}; \tag{5}$$

$\mathcal{G}^n_{A,cpi}$ and $\mathcal{G}^n_{A,cyclic}$ are, respectively, the orthogonal projections of $\mathcal{G}^n_A$ to $\mathbb{ST}^n_A$ and to $\mathbb{C}^n_A$. The length of a closed path in $\mathcal{G}^n_A$ equals its $\mathcal{G}^n_{A,cyclic}$ length.

The critical last statement is a consequence of the linear form of Eq. 5, which requires the length of a path in $\mathcal{G}^n_A$ to equal the sum of its lengths in $\mathcal{G}^n_{A,cpi}$ and in $\mathcal{G}^n_{A,cyclic}$. By design, the length of a closed path in $\mathcal{G}^n_{A,cpi}$ is zero. Namely, $\mathcal{G}^n_{A,cpi}$ extracts those portions of $\mathcal{G}^n_A$ entries that have

nothing substantive to contribute to closed path lengths. It now follows that the path's lengths in $\mathcal{G}_A^n$ and in $\mathcal{G}_{A,cyclic}^n$ must agree. In turn, this means that *all relevant closed path information for $\mathcal{G}_A^n$ is encoded in the three-cycles of $\mathcal{G}_{A,cyclic}^n$*. Stated differently, $\mathcal{G}_A^n \in \mathbb{G}_A^n$ has an inherent three-cycle symmetry structure displayed by $\mathcal{G}_{A,cyclic}^n$ but camouflaged by $\mathcal{G}_{A,cpi}^n$.



a. $\mathcal{G}_A^3$ **b.** $\mathcal{G}_{A,cpi}^3$ **c.** $\mathcal{G}_{A,cyclic}^3$

Figure 2. Interpreting $\mathcal{G}_{A,cyclic}^n$

To expand on the comment that the $\mathcal{G}_{A,cpi}^n$ entries contribute *nothing substantive* about closed path lengths, notice that computing the $V_1 \xrightarrow{8} V_2 \xrightarrow{10} V_3 \xrightarrow{-12} V_1$ length of 6 in Fig. 2a involves a subtraction cancelation. To appreciate this structure, let the optimal (but unknown) cancelled values be u, v, and w from, respectively, arcs $\widehat{V_1V_2}$, $\widehat{V_2V_3}$, and $\widehat{V_3V_1}$. That is, $(8-u)+(10-v)+(-12-w)=6$ where the cancelled values define the equation $u+v+w=0$, which corresponds to a zero-length closed path. For $n>3$, this cancellation applies to all triplets, so these extracted values define a $\mathbb{ST}_A^n$ graph. The optimal choice of removed terms comes from the $\mathbb{ST}_A^n$ graph that most closely resembles $\mathcal{G}_A^n$, which is its orthogonal projection $\mathcal{G}_{A,cpi}^n$ (Thm. 3). Indeed, with Fig. 2, the $\mathcal{G}_{A,cpi}^3$ component extracts $u+v+w=6+8-14=0$. What remains are portions of arc entries that, without further need of modification, are relevant for computing path lengths. These terms define the $\mathcal{G}_{A,cyclic}^3$ graph with its $V_1 \xrightarrow{2} V_2 \xrightarrow{2} V_3 \xrightarrow{2} V_1$ closed cycle that directly provides the path length of 6. Not only do the superfluous $\mathcal{G}_{A,cpi}^n$ terms complicate computations, but, as discussed below, they can sidetrack optimization approaches such as the greedy algorithm.

The same behavior holds in general; e.g., the zero length of each triplet in Fig. 1b identifies the optimal subtraction cancelations for computing $\mathcal{G}_A^5$ path lengths. As the cyclic $\mathcal{G}_{A,cyclic}^n$ captures the germane portions for path-length considerations, when searching for the longest or shortest $\mathcal{G}_A^n$ paths, ignore $\mathcal{G}_A^n$ and analyze only the simpler $\mathcal{G}_{A,cyclic}^n$. (Some $\mathcal{G}_{A,cyclic}^n$ arcs belong to several cycles; e.g., the $V_2 \xrightarrow{2} V_1$ leg in Fig. 1c is the sum of this arc's length in two cycles.)

A slight modification of Thm. 3 describes the length of any connected path.

Corollary 3. *The length of a path in $\mathcal{G}_A^n$ that connects V_j with V_k is the length of this path in $\mathcal{G}_{A,cyclic}^n$ plus the length of the $V_j \to V_k$ arc in $\mathcal{G}_{A,cpi}^n$.*

To illustrate, the length of 12 for path $V_1 \xrightarrow{14} V_4 \xrightarrow{-6} V_3 \xrightarrow{1} V_5 \xrightarrow{3} V_4$ in $\mathcal{G}_A^5$ (Fig. 1a), which can meet vertices multiple times, equals the easier computed length of 1 for $V_1 \xrightarrow{3} V_4 \xrightarrow{0} V_3 \xrightarrow{-2} V_5 \xrightarrow{0} V_4$ in $\mathcal{G}_{A,cyclic}^5$ plus 11 from the $V_1 \xrightarrow{11} V_4$ arc length in $\mathcal{G}_{A,cpi}^5$. The $3-2$ subtraction manifests the algebraic arrangement of the triplets.

Proof: The length of a path in $\mathcal{G}_A^n$ equals the sum of its lengths in $\mathcal{G}_{A,cpi}^n$ and $\mathcal{G}_{A,cyclic}^n$. According to Cor. 1, the length of a connected path in $\mathcal{G}_{A,cpi}^n$ starting at V_j and ending at V_k equals the length of the arc $V_j \to V_k$ connecting the endpoints. This completes the proof. □

In general, it is easier to extract $\mathcal{G}_A^n$ path properties from $\mathcal{G}_{A,cyclic}^n$ than from $\mathcal{G}_A^n$; e.g., even the flawed greedy algorithm (GA) shows that $V_1 \xrightarrow{3} V_4 \xrightarrow{3} V_2 \xrightarrow{2} V_5 \xrightarrow{2} V_3 \xrightarrow{1} V_1$ of length 11 is the longest $\mathcal{G}_{A,cyclic}^5$ (Fig. 1c) Hamiltonian path. According to Eq. 1, its reversal (length -11) is the

shortest. This solves the $\mathcal{G}_A^5$ TSP problem because, as Thm. 3 asserts, these two $\mathcal{G}_{A,cyclic}^5$ paths identify, respectively, the longest and shortest $\mathcal{G}_A^5$ Hamiltonian paths and their lengths. But GA[1] fails with $\mathcal{G}_A^5$ primarily because the $\mathcal{G}_{A,cpi}^5$ values, which cancel when calculating lengths, divert the algorithm. To see this, using the GA to search for the longest Hamiltonian path of $\mathcal{G}_A^5$ yields the incorrect $V_1 \xrightarrow{14} V_4 \xrightarrow{-3} V_5 \xrightarrow{-1} V_3 \xrightarrow{-3} V_2 \xrightarrow{1} V_1$. Applying GA to $\mathcal{G}_{A,cpi}^5$ generates the same path, which underscores the fact that $\mathcal{G}_{A,cpi}^5$ is the source of this problem.

For $\mathbb{G}_A^n$, the tasks of finding the longest and shortest $\mathcal{G}_{A,cyclic}^n$ Hamiltonian paths coincide. This is because the reversal of one is the other.

Corollary 4. *If the length of a path in $\mathcal{G}_{A,cyclic}^n$ is x, then the length of its reversal is $-x$.*

2.2. **Decomposition.** To compute $\mathcal{G}_{A,cpi}^n$ and $\mathcal{G}_{A,cyclic}^n$, recall that $\mathcal{G}_{A,cpi}^n \in \mathbb{ST}_A^n$ is the orthogonal projection of $\mathcal{G}_A^n$. Terms from this projection are described next; for added details see [4].

The projection is a linear algebra exercise. To connect graphs with vectors, let $\mathbf{d}_A^n \in \mathbb{R}_A^{\binom{n}{2}}$ be

$$\mathbf{d}_A^n = (d_{1,2}, d_{1,3}, \ldots, d_{1,n}; d_{2,3}, \ldots, d_{2,n}; d_{3,4}, \ldots; d_{n-1,n}), \text{ where } d_{i,j} = -d_{j,i}; \tag{6}$$

the semicolons designate where the first subscript changes. To identify $\mathbb{G}_A^n$ with $\mathbb{R}_A^{\binom{n}{2}}$, let $d_{i,j}$ be the arc length $V_i \xrightarrow{d_{i,j}} V_j$. As $V_i \xrightarrow{d_{i,j}} V_j$ equals $V_j \xrightarrow{-d_{i,j}} V_i$, it follows that $d_{j,i} = -d_{i,j}$. Let $\mathbb{ST}_A^n \subset \mathbb{G}_A^n$ also denote the $(n-1)$-dimensional (strongly transitive) subspace of $\mathbb{R}_A^{\binom{n}{2}}$ where each triplet $\{i, j, k\}$ satisfies $d_{i,j} + d_{j,k} = d_{i,k}$. With these identifications, structures of $\mathbb{R}_A^{\binom{n}{2}}$ and $\mathbb{G}_A^n$ can be described interchangeably.

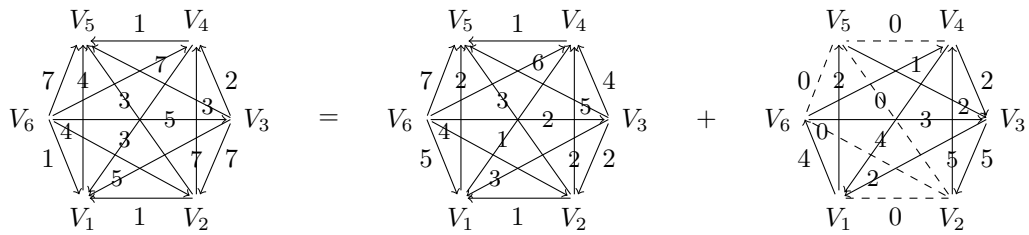
Definition 2. *For vertex V_j of $\mathcal{G}_A^n \in \mathbb{G}_A^n$, let $\mathcal{S}_A(V_j)$ be $\frac{1}{n}$ times the sum of the arc lengths leaving vertex V_j, $j = 1, \ldots, n$.*

Theorem 4. [4] *For $\mathcal{G}_A^n \in \mathbb{G}_A^n$, the $\mathcal{G}_{A,cpi}^n$ path length from V_i to V_j is $d_{i,j} = \mathcal{S}_A(V_i) - \mathcal{S}_A(V_j)$, $i, j \in \{1, \ldots, n\}$. Each of $\mathcal{G}_A^n$ and $\mathcal{G}_{A,cpi}^n$ satisfy $\sum_{j=1}^n \mathcal{S}_A(V_j) = 0$. Graph $\mathcal{G}_{A,cyclic}^n$ is given by $\mathcal{G}_{A,cyclic}^n = \mathcal{G}_A^n - \mathcal{G}_{A,cpi}^n$. All vertices of $\mathcal{G}_{A,cyclic}^n$ satisfy the stronger $\mathcal{S}_A(V_j) = 0$. Conversely, if all vertices of $\mathcal{G}_A^n \in \mathbb{G}_A^n$ satisfy $\mathcal{S}_A(V_j) = 0$, then $\mathcal{G}_A^n \in \mathbb{C}_A^n$*

The concluding statement follows from Thm. 4's first sentence. This is because $\mathcal{S}_A(V_j) = 0$ for all vertices requires all $d_{i,j}$ legs of $\mathcal{G}_{A,cpi}^n$ to equal zero. As $\mathcal{G}_{A,cpi}^n = 0$, $\mathcal{G}_A^n$ equals its $\mathcal{G}_{A,cyclic}^n$ component.

With $\mathcal{G}_A^6$ (Fig. 3a), the $S_A(V_j)$ values (called 'Borda Values' in [4]) are $\mathcal{S}_A(V_1) = \frac{1}{6}(-1 - 5 - 3 + 4 - 1) = -1, \mathcal{S}_A(V_2) = 0, \mathcal{S}_A(V_3) = 2, \mathcal{S}_A(V_4) = -2, \mathcal{S}_A(V_5) = -3, \mathcal{S}_A(V_6) = 4$. Thus (Thm. 4), the Fig. 3b values for $\mathcal{G}_{A,cpi}^6$ are $d_{1,2} = \mathcal{S}_A(V_1) - \mathcal{S}_A(V_2) = -1, d_{1,3} = 3, d_{1,4} = 1, d_{1,5} = 2, d_{1,6} = -5, d_{2,3} = -2, d_{2,4} = 2, d_{2,5} = 3, d_{2,6} = -4, d_{3,4} = 4, d_{3,5} = 5, d_{3,6} = -2, d_{4,5} = 1, d_{4,6} = -6, d_{5,6} = -7$. Graph $\mathcal{G}_{A,cyclic}^6$ follows from the equality $\mathcal{G}_{A,cyclic}^6 = \mathcal{G}_A^6 - \mathcal{G}_{A,cpi}^6$; this defines Fig. 3c.

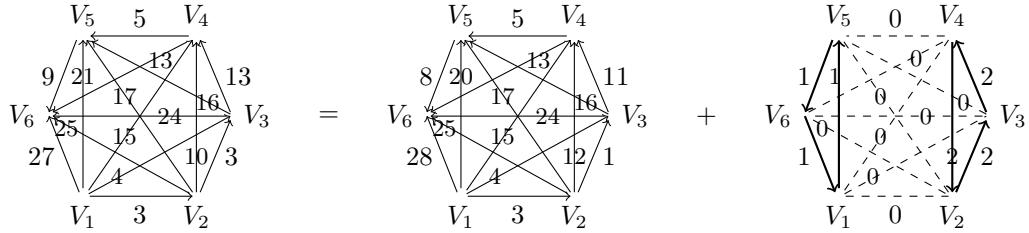
[1]Because its limitations and failings are well known, the greedy algorithm (GA) is used to illustrate advantages of the decomposition. Other GA difficulties, caused by the algebra of cycles, are indicated with Fig. 12.



a. $\mathcal{G}_A^6$ **b.** $\mathcal{G}_{A,cpi}^6$ **c.** $\mathcal{G}_{A,cyclic}^6$

Figure 3. Decomposition of a $\mathcal{G}_A^6$

Notice how the redundant $\mathcal{G}_{A,cpi}^6$ (Fig. 3b) dominates the $\mathcal{G}_A^6$ structure even though the simpler $\mathcal{G}_{A,cyclic}^6$ (Fig. 3c) is what determines all of $\mathcal{G}_A^6$'s closed path properties. This must be expected because, according to the orthogonal projection construction, $\mathcal{G}_{A,cpi}^n$ is the $\mathbb{ST}_A^n$ graph that most closely resembles $\mathcal{G}_A^n$. As true with Fig. 2, a feature of this $\mathcal{G}_{A,cpi}^6$ and $\mathcal{G}_A^6$ similarity is that $\mathcal{G}_{A,cpi}^6$ collects terms involved in subtraction/cancellations when computing $\mathcal{G}_A^6$ path lengths; the remaining $\mathcal{G}_{A,cyclic}^6$ entries are the relevant portions for determining path properties.



a. $\mathcal{G}_A^6$ **b.** $\mathcal{G}_{A,cpi}^6$ **c.** $\mathcal{G}_{A,cyclic}^6$

Figure 4. Advantages of $\mathcal{G}_{A,cyclic}^n$

Figure 4 illustrates how seriously the $\mathcal{G}_{A,cpi}^n$ terms can cloud a path analysis. While path properties are determined by the extremely simple $\mathcal{G}_{A,cyclic}^6$ (Fig. 4c), this clarity is not obvious from $\mathcal{G}_A^6$ (Fig. 4a). The reason is that $\mathcal{G}_A^6$ more closely resembles the associated $\mathcal{G}_{A,cpi}^6$. As developed next, this is a general phenomenon.

Definition 3. *Two graphs $\mathcal{G}_{A,1}^n, \mathcal{G}_{A,2}^n \in \mathbb{G}_A^n$ are "closed path equivalent" iff $\mathcal{G}_{A,1,cpi}^n = \mathcal{G}_{A,2,cpi}^n$.*

Corollary 5. *The "closed path equivalent" relationship is an equivalence relation. Two graphs are equivalent iff their difference is a graph in $\mathbb{ST}_A^n$. Thus, an equivalence class of this relationship is the sum of a $\mathcal{G}_{A,cyclic}^n$ and the $(n-1)$-dimensional linear subspace $\mathbb{ST}_A^n$.*

According to Cor. 5, multiple (actually, most) choices of a $\mathcal{G}_{A,cpi}^n$ from the vast offerings of the $(n-1)$-dimensional $\mathbb{ST}_A^n$ dictate the form of $\mathcal{G}_A^n$ and obscure the relevant $\mathcal{G}_{A,cyclic}^n$.

2.3. **Structure of triplets.** Analyzing closed path properties of a $\mathcal{G}_A^n$ involves the algebraic structure of the $\mathcal{G}_{A,cyclic}^n$ three-cycles. The following shows how to identify $\mathcal{G}_{A,cyclic}^n$'s three-cycles.

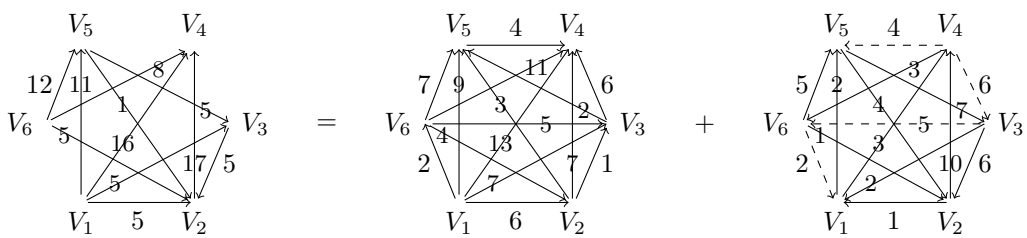
Theorem 5. *For $\mathcal{G}_{A,cyclic}^n$, only one three-cycle of a Cor. 2 basis has a $\widehat{V_jV_k}$ arc. The cycle's multiple is the $V_j \xrightarrow{d_{j,k}} V_k$ weight in $\mathcal{G}_{A,cyclic}^n$.*

Proof: Arc $V_j \xrightarrow{d_{j,k}} V_k$ in $\mathcal{G}^n_{A,cyclic}$ appears only in $V_s \xrightarrow{x} V_j \xrightarrow{x} V_k \xrightarrow{x} V_s$ of the $\mathcal{CB}^n_A$ basis. As all weights in a three-cycle agree, this is the cycle's multiple. □

To illustrate, Fig. 1c has the three three-cycles $V_1 \xrightarrow{3} V_4 \xrightarrow{3} V_2 \xrightarrow{3} V_1$, $V_1 \xrightarrow{1} V_2 \xrightarrow{1} V_3 \xrightarrow{1} V_1$ and $V_2 \xrightarrow{2} V_5 \xrightarrow{2} V_3 \xrightarrow{2} V_2$. Figure 3c has the four three-cycles $V_2 \xrightarrow{5} V_4 \xrightarrow{5} V_3 \xrightarrow{5} V_2, V_1 \xrightarrow{4} V_6 \xrightarrow{4} V_4 \xrightarrow{4} V_1, V_3 \xrightarrow{3} V_4 \xrightarrow{3} V_6 \xrightarrow{3} V_3$, and $V_1 \xrightarrow{2} V_5 \xrightarrow{2} V_3 \xrightarrow{2} V_1$. When finding Hamiltonian paths, to avoid prematurely returning to a vertex, at most two arcs of a three-cycle can be used. With this caveat, the GA can succeed with $\mathcal{G}^n_{A,cyclic}$ settings where it would fail for $\mathcal{G}^n_A$. For instance, the GA delivers the longest Fig. 3c Hamiltonian path $V_3 \xrightarrow{5} V_2 \xrightarrow{5} V_4 \xrightarrow{4} V_1 \xrightarrow{4} V_6 \xrightarrow{0} V_5 \xrightarrow{2} V_3$ with length 20, which is its $\mathcal{G}^6_A$ (Fig. 3a) path length (Thm. 3). For $\mathcal{G}^n_A \in \mathbb{G}^n_A$, its shortest Hamiltonian route reverses the longest.

A vertex that is a source or sink imposes an obstacle in finding optimal paths. For instance, if V_j is a sink for negative cost directions, as is V_6 in Fig. 3a, all ways to leave this vertex require using a positive cost direction. Fortunately, general properties of $\mathcal{G}^n_{A,cyclic} \in \mathbb{C}^n_A$ can be obtained via the $\mathbb{C}^n_A$ basis (Thm. 2, Cor. 2); sample conclusions are in Thm. 6. The first assertion identifies two well behaved settings. The second comment asserts that although sources and sinks are not unusual in $\mathcal{G}^n_A$, which cause subtraction cancellations, they never arise in $\mathcal{G}^n_{A,cyclic}$. According to the theorem's last comment, sources and sinks appear in $\mathcal{G}^n_A$ only because $\mathcal{G}^n_{A,cpi}$ almost always has them. Thus if $\mathcal{G}^n_A$ has a source and/or sink, expect that its structure is dominated by $\mathcal{G}^n_{A,cpi}$.

Theorem 6. *For $\mathcal{G}^n_{A,cyclic} \in \mathbb{C}^n_A$, $n = 4, 5$, its longest and shortest Hamiltonian paths have, respectively, all non-negative arc costs and non-positive arc costs. For $n \geq 4$, while $\mathcal{G}^n_A$ can have sinks and/or sources, this is impossible for a $\mathcal{G}^n_{A,cyclic}$. In contrast, if all positive cost directions of $\mathcal{G}^n_{A,cpi}$ are non-zero, then both the positive and negative cost directions have a sink and a source.*



a. $\mathcal{G}^6_A$ **b.** $\tilde{\mathcal{G}}^6_{A,cpi}$ **c.** $\tilde{\mathcal{G}}^6_{A,cyclic}$

Figure 5. An incomplete $\mathcal{G}^6_A$

2.4. **Incomplete graphs.** Incomplete graphs are similarly reduced. The approach is illustrated with $\mathcal{G}^6_A$ (Fig. 5a) where arcs $\widehat{V_1V_6}$, $\widehat{V_3V_4}$, $\widehat{V_3V_6}$, and $\widehat{V_4V_5}$ are excluded. To complete $\mathcal{G}^6_A$, add in the missing arcs with arbitrarily selected lengths. (For Fig. 5a, arcs of zero length were added.) Denote the completed graph by $\tilde{\mathcal{G}}^6_A$, and compute $\tilde{\mathcal{G}}^6_{A,cpi}$ (Fig. 5b) and $\tilde{\mathcal{G}}^6_{A,cyclic}$ (Fig. 5c). Closed path lengths in $\mathcal{G}^6_A$ agree with their lengths in $\tilde{\mathcal{G}}^6_{A,cyclic}$. (The four dashed arrows in Fig. 5c denote the forbidden arcs.) As the path $V_2 \xrightarrow{10} V_4 \xrightarrow{3} V_6. \xrightarrow{5} V_5 \xrightarrow{7} V_3 \xrightarrow{2} V_1 \xrightarrow{-1} V_2$ uses most of the largest allowed $\tilde{\mathcal{G}}^6_{A,cyclic}$ leg lengths, it is easy to show that its length of 26 is the longest $\tilde{\mathcal{G}}^6_{A,cyclic}$ Hamiltonian path. Thus, this is the longest $\mathcal{G}^6_A$ Hamiltonian path with the same length.

Theorem 7. *For an incomplete $\mathcal{G}^n_A$, replace all non-admissible arcs with arcs of arbitrary lengths to define $\tilde{\mathcal{G}}^n_A$; compute $\tilde{\mathcal{G}}^n_{A,cyclic}$. The length of a closed path in $\mathcal{G}^n_A$ equals its length in $\tilde{\mathcal{G}}^n_{A,cyclic}$.*

The length of an admissible path starting at V_j and ending at V_k in the incomplete $\mathcal{G}_A^n$ is its length in $\tilde{\mathcal{G}}_{A,cyclic}^n$ plus the length of the $V_j \to V_k$ arc in $\tilde{\mathcal{G}}_{A,cpi}^n$.

The concluding Thm. 7 assertion allows the $V_j \to V_k$ arc in $\tilde{\mathcal{G}}_{A,cpi}^n$ to be a forbidden $\mathcal{G}_A^n$ choice. As a Fig. 5 example, the $\mathcal{G}_A^6$ path $V_6 \xrightarrow{12} V_5 \xrightarrow{5} V_3$ has length 17. In $\tilde{\mathcal{G}}_{A,cyclic}^6$ this path $V_6 \xrightarrow{5} V_5 \xrightarrow{7} V_3$ has length 12, which is added to the 5 length of the banned $V_6 \xrightarrow{5} V_3$ arc in $\tilde{\mathcal{G}}_{A,cpi}^6$.

Proof: A closed path's length in $\mathcal{G}_A^n$ is the same in $\tilde{\mathcal{G}}_A^n$ and (by Thm. 3) in $\tilde{\mathcal{G}}_{A,cyclic}^n$.

A connected path's length in $\mathcal{G}_A^n$ is the same in $\tilde{\mathcal{G}}_A^n$, which equals the sum of its lengths in $\tilde{\mathcal{G}}_{A,cpi}^n$ and $\mathcal{G}_{A,cyclic}^n$. Its $\tilde{\mathcal{G}}_{A,cpi}^n$ length is that of its $V_j \to V_k$ arc, which completes the proof. □

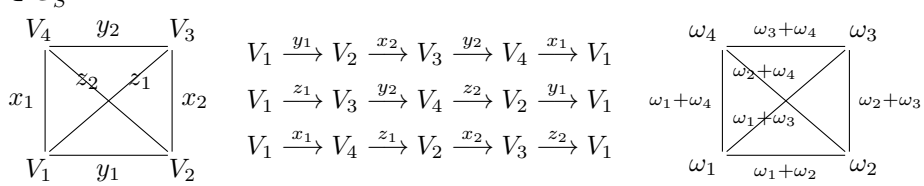
2.5. **Lower degrees of freedom.** The $\mathbb{C}_A^n$ graphs have all (and only) path information, which simplifies deriving closed path properties, computing lengths, and designing algorithms by using the (somewhat predictive) algebra of three-cycles. General path properties follow from the $\mathbb{C}_A^n$ basis, which characterizes all $\mathcal{G}_{A,cyclic}^n$ choices. By seeking general closed path properties, rather than just Hamiltonian circuits, $\mathbb{C}_A^n$ is a best possible component. This is because the simplest closed path is a triplet, so that three-cycle must be in $\mathbb{C}_A^n$.

3. Symmetric cost settings

Deriving the structure of $\mathbb{G}_S^n$—the space of n-vertex complete symmetric weighted (no loops) graphs—mimics what was done for $\mathbb{G}_A^n$. As with Eq. 2, the goal is to decompose a $\mathcal{G}_S^n \in \mathbb{G}_S^n$ as

$$\mathcal{G}_S^n = \mathcal{G}_{S,cpi}^n + \mathcal{G}_{S,cyclic}^n, \tag{7}$$

where the definition of $\mathcal{G}_{S,cpi}^n$ will captures terms that, at least initially, can be ignored when computing lengths of $\mathcal{G}_S^n$ closed paths. Thus, all central $\mathcal{G}_S^n$ path properties are based on the $\mathcal{G}_{S,cyclic}^n$ structure. The approach is to find the manifold of $\mathbb{G}_S^n$ graphs where closed paths have a fixed length; this manifold defines the $\mathcal{G}_{S,cpi}^n$ components. The normal bundle of this manifold measures deviations from "sameness" to capture what is needed to find closed path properties of a $\mathcal{G}_S^n \in \mathbb{G}_S^n$.



a. $\mathcal{G}_S^4$ **b.** Equal path lengths **c.** $\mathcal{G}_{S,cpi}^4$ representation

Figure 6. Closed path independence for $\mathbb{G}_S^4$

The $\mathbb{G}_S^n$ cpi definition and structure differs from that of $\mathbb{G}_A^n$; e.g., rectangles replace triangles. For instance, $\mathcal{G}_S^4 \in \mathbb{G}_S^4$ (Fig. 6a) is cpi iff the three Fig. 6b routes have equal length. With cancelations, this requires the three sums of the vertical, the horizontal, and the diagonal lengths to agree, or

$$x_1 + x_2 = y_1 + y_2 = z_1 + z_2. \tag{8}$$

Hamiltonian paths for $\mathcal{G}_S^4 \in \mathbb{G}_S^4$ combine two of the pairs of diagonals, vertical edges, or horizontal edges, so the two smallest Eq. 8 sums define the shortest path with length equal to this sum.[2] For

[2]Should the vertices define a triangle with one in the interior, the three pairs are defined by the triangle's three vertices. A pair is the arc from a vertex to the interior point and the triangle's leg that is opposite the vertex.

instance, if $x_1 + x_2 = 10$, $y_1 + y_2 = 20$, and $z_1 + z_2 = 30$, then the shortest Hamiltonian circuit traverses the perimeter and it has length 30..

All cpi graphs in $\mathbb{G}_S^4$ satisfy two independent equations (Eq. 8) in six variables. One solution has zero leg lengths, so all solutions (i.e., all cpi graphs $\mathcal{G}_{S,cpi}^4$) are characterized by Eq. 8's four-dimensional kernel. One choice uses weights $\{\omega_j\}_{j=1}^4$ where ω_j is assigned to vertex V_j, $j = 1, \ldots, 4$, to define the $\widehat{V_jV_k}$ length of $\omega_j + \omega_k$ (Fig. 6c). It follows from Fig. 6c that this choice satisfies Eq. 8. For $\mathbb{G}_S^4$, these are the 'closed path independent' graphs. The common path length depends on how often each vertex is visited; e.g., a $\mathcal{G}_{S,cpi}^4$ closed path that visits each of the three vertices $\{V_i\}_{i=1}^3$ twice has length $4\sum_{j=1}^3 \omega_j$; Hamiltonian paths in $\mathcal{G}_{S,cpi}^4$ have length $T(\mathcal{G}_{S,cpi}^4) = 2\sum_{j=1}^4 \omega_j$.

The above discussion centered on Fig. 6 extends to $n \geq 4$.

Definition 4. *A graph $\mathcal{G}_S^n \in \mathbb{G}_S^n$ is* 'closed path independent' *iff for any set of vertices, all closed paths that pass through each of these vertices once have the same length.*

Theorem 8. *For $n \geq 4$, a cpi graph $\mathcal{G}_{S,cpi}^n \in \mathbb{G}_S^n$ assigns weight ω_j to vertex V_j, $j = 1, \ldots, n$; the $\widehat{V_jV_k}$ length is $\omega_j + \omega_k$, $j \neq k$. A closed path passing once through the vertices $\{V_j\}_{j\in\mathcal{D}}$ has length $2\sum_{j\in\mathcal{D}} \omega_j$. A Hamiltonian path length in $\mathcal{G}_{S,cpi}^n$ is $2\sum_{j=1}^n \omega_j$.*

Determining the structure of $\mathcal{G}_{S,cpi}^n$ requires identifying $\mathbb{G}_S^n$ with $\mathbb{R}_S^{\binom{n}{2}}$. Here, $\mathbb{R}_S^{\binom{n}{2}}$ differs from $\mathbb{R}_A^{\binom{n}{2}}$ (Eq. 6) because in $\mathbb{G}_A^n$ (Sect. 2.2), $d_{i,j} = -d_{j,i}$, but in $\mathbb{G}_S^n$, $d_{i,j} = d_{j,i}$. Thus,

$$\mathbf{d}_S^n = (d_{1,2}, d_{1,3}, \ldots, d_{1,n}; d_{2,3}, \ldots, d_{2,n}; d_{3,4}, \ldots; d_{n-1,n}) \in \mathbb{R}_S^{\binom{n}{2}}, \text{ where } d_{i,j} = d_{j,i}; \tag{9}$$

Theorem 9. *The space of $\mathcal{G}_{S,cpi}^n$ graphs, denoted by $\mathbb{CPI}_S^n$, is a n-dimensional linear subspace of $\mathbb{G}_S^n$, or, equivalently, of $\mathbb{R}_S^{\binom{n}{2}}$. Let $\mathbf{B}_j^n \in \mathbb{R}_S^{\binom{n}{2}}$ be where $d_{j,k} = 1$ for all $k \neq j, k = 1, \ldots, n$; all other $d_{u,v} = 0$. A basis for $\mathbb{CPI}_S^n$ is $\{\mathbf{B}_j^n\}_{j=1}^n$.*

As $\mathcal{G}_{S,cpi}^n$ will identify components of $\mathcal{G}_S^n$ entries that cloud analyzing closed paths, the emphasis shifts to $\mathcal{G}_{S,cyclic}^n = \mathcal{G}_S^n - \mathcal{G}_{S,cpi}^n$. The dimensions of $\mathbb{G}_S^n$ and $\mathbb{CPI}_S^n$ are $\binom{n}{2}$ and n, so $\mathbb{CPI}_S^n$'s normal subspace, $\mathbb{C}_S^n$, has dimension $\frac{n(n-3)}{2}$. (The dimension of $\mathbb{C}_A^n$ is $\binom{n-1}{2} = \frac{n(n-3)}{2} + 1$.) Its four-cycle structure (Thm 10) identifies $\mathcal{G}_S^n$'s inherent symmetry.

Theorem 10. *Let vector $\mathbf{b}_{i,j,k,s}^n$ be where $d_{i,j} = 1, d_{j,k} = -1, d_{k,s} = 1, d_{s,i} = -1$; all other $d_{u,v} = 0$. The space $\mathbb{C}_S^n$ spanned by all $\{\mathbf{b}_{i,j,k,s}^n\}$, with dimension $\frac{n(n-3)}{2}$, is orthogonal to $\mathbb{CPI}_S^n$. With*

$$\mathcal{A}_{1,2}^n = \{\mathbf{b}_{1,2,j,k}^n\}_{2<j<k\leq n},\ \mathcal{B}_{1,2}^n = \{\mathbf{b}_{1,2,j,3}^n\}_{j=4}^n, \text{ a basis is } \mathcal{A}_{1,2}^n \cup \mathcal{B}_{1,2}^n. \tag{10}$$

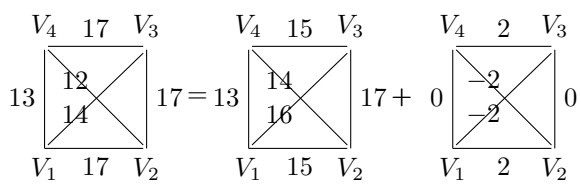
To explain these $\mathbf{b}_{i,j,k,s}^n$ vectors, if the arc lengths of the route around the perimeter of Fig. 6a satisfy the Eq. 8 cpi requirement, then $x_1 - y_1 + x_2 - y_2 = 0$. This equation has the normal vector $(1, -1, 1, -1)$, which in path form, is $\mathbf{b}_{4,1,2,3}^4$. In general the $\mathbf{b}_{i,j,k,s}^n$ path is the four-cycle $V_i \xrightarrow{1} V_j \xrightarrow{-1} V_k \xrightarrow{1} V_s \xrightarrow{-1} V_i$; each vertex has one leg of length 1 and one of length -1. With Fig. 6a, the $\mathbf{b}_{1,2,3,4}^4$ and $\mathbf{b}_{1,4,2,3}^4$ multiples are, respectively, $\frac{1}{4}\{(y_1 + y_2) - (x_1 + x_2)\}$ and $\frac{1}{4}\{(x_1 + x_2) - (z_1 + z_2)\}$. Namely, $\mathbf{b}_{i,j,k,s}^n$ measures how the $\mathcal{G}_S^n$ data deviates from cpi sameness while identifying which data edges of a four-tuple are ridges or valleys.

Proof. That $\mathbf{b}_{i,j,k,s}^n$ is orthogonal to each $\mathbf{B}_t^n$ is immediate. If $t \neq i, j, k, s$, the scalar product is zero. If t is one of these indices, say $t = j$, then one component of $\mathbf{b}_{i,j,k,s}^n$ with vertex V_j is positive and the other is negative, so the scalar product with $\mathbf{B}_t^n$ is zero.

Establishing the linear independence of Eq. 10 follows a switching pattern. Iteratively, it will be shown that all coefficients of $\sum_{2<j<k\leq n} \alpha_{j,k}\mathbf{b}^n_{1,2,j,k} = \mathbf{0}$ must equal zero. For each of the $\binom{n-3}{2}$ top vectors in $\mathcal{A}^n_{1,2}$ (i.e., $j,k \geq 4$), only vector $\mathbf{b}^n_{1,2,j,k}$ has a non-zero $d_{j,k}$, so $\alpha_{j,k} = 0$. For all remaining vectors, either j or k equals 3. Of these, only the top $\mathcal{B}^n_{1,2}$ vector of $\mathbf{b}^n_{1,2,n,3}$ has a non-zero $d_{2,n}$, so $\alpha_{n,3} = 0$. The top remaining $\mathcal{A}^n_{1,2}$ vector is $\mathbf{b}^n_{1,2,3,n}$, where, with the removal of $\mathbf{b}^n_{1,2,n,3}$, only $\mathbf{b}^n_{1,2,3,n}$ has non-zero $d_{3,n}$, so $\alpha_{3,n} = 0$. The obvious induction argument of switching between remaining $\mathcal{A}^n_{1,2}$ and $\mathcal{B}^n_{1,2}$ vectors continues. That is, if s is the upper bound of the remaining j,k values, then only $\mathbf{b}^n_{1,2,s,3} \in \mathcal{B}^n_{1,2}$ has a non-zero $d_{2,s}$ term, so $\alpha_{s,3} = 0$. The top remaining $\mathcal{A}^n_{1,2}$ vector is $\mathbf{b}^n_{1,2,3,s}$; as $\mathbf{b}^n_{1,2,s,3}$ was removed, only $\mathbf{b}^n_{1,2,3,s}$ of the remaining vectors has a non-zero $d_{3,s}$ term, so $\alpha_{3,s} = 0$ and $s-1$ is the largest remaining j,k value. This completes the proof. □

3.1. **Decomposing $\mathbb{G}^n_S$.** Theorem 11 summarizes the above; it is the $\mathbb{G}^n_S$ version of Thm. 2. As with Sect. 2, the decomposition involves $O(n^2)$ computations.

Theorem 11. *For $n \geq 4$, $\mathbb{G}^n_S$ has an n-dimensional linear subspace $\mathbb{CPI}^n_S$ and an orthogonal $\frac{n(n-3)}{2}$ dimensional linear subspace $\mathbb{C}^n_S$. A $\mathcal{G}^n_S \in \mathbb{G}^n_S$ has a unique decomposition $\mathcal{G}^n_S = \mathcal{G}^n_{S,cpi} + \mathcal{G}^n_{S,cyclic}$ where $\mathcal{G}^n_{S,cpi} \in \mathbb{CPI}^n_S$ and $\mathcal{G}^n_{S,cyclic} \in \mathbb{C}^n_S$ are, respectively, the orthogonal projection of $\mathcal{G}^n_S$ to $\mathbb{CPI}^n_S$ and to $\mathbb{C}^n_S$.*



a. $\mathcal{G}^4_S$ **b.** $\mathcal{G}^4_{S,cpi}$ **c.** $\mathcal{G}^4_{S,cyclic}; 2\mathbf{b}_{1,2,4,3}$

Figure 7. Interpreting $\mathcal{G}^n_{S,cyclic}$

Before computing $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cyclic}$, Fig. 7 is used to explain their roles and to relate $\mathcal{G}^n_{S,cpi}$ with $\mathcal{G}^n_{A,cpi}$. According to Thm. 11, $\mathcal{G}^n_{S,cpi}$ is the $\mathbb{CPI}^n_S$ graph that most closely resembles $\mathcal{G}^n_S$; this similarity is apparent when comparing Figs. 7a, b. A defining feature of $\mathcal{G}^n_{A,cpi} \in \mathbb{ST}^n_A$ is that for any set of vertices, the length of all closed paths meeting each vertex once is zero. Similarly, according to Def. 4, for $\mathcal{G}^n_{S,cpi} \in \mathbb{CPI}^n_S$ and for any selected set of vertices, the length of all closed paths that meet each vertex once is the same, but not necessarily zero. The role of the decomposition in $\mathbb{G}^n_A$ and in $\mathbb{G}^n_S$ is to remove these common path length values. Thus the $\mathcal{G}^n_{A,cyclic}$ and $\mathcal{G}^n_{S,cyclic}$ graphs characterize how the data from the original graph ($\mathcal{G}^n_A$ or $\mathcal{G}^n_S$) differs from the cpi sameness to provide valued path length information.

In Fig. 7b, the sums of its vertical edges, horizontal edges, and diagonals all equal 30. Thus, all Fig. 7b Hamiltonian paths have the length 60. Three of the six Fig. 7a Hamiltonian circuits are $V_1 \xrightarrow{17} V_2 \xrightarrow{17} V_3 \xrightarrow{17} V_4 \xrightarrow{13} V_1$ with length 64, $V_1 \xrightarrow{17} V_2 \xrightarrow{12} V_4 \xrightarrow{17} V_3 \xrightarrow{14} V_1$ with length 60, and $V_1 \xrightarrow{14} V_3 \xrightarrow{17} V_2 \xrightarrow{12} V_4 \xrightarrow{13} V_1$ with length 56; the other three are reversals. The average length of these paths is 60, which agrees with its Fig. 7b value. This comparison accurately suggests that for any set of vertices used to define closed paths, what happens in $\mathcal{G}^n_{S,cpi}$ is the average of what happens in $\mathcal{G}^n_S$. For instance, the length of a Hamiltonian path for Fig. 7a is the sum of its lengths in Fig.7b and Fig. 7c. A "subtraction" argument, similar to that used with Fig. 2, is that the portion of a path entry contributing to the Fig. 7b average Hamiltonian length is subtracted from the actual leg value. What remains determines how the path length differs from the average, so it is

used in the computation; the average length of a Hamiltonian path is replaced at the end. Thus, as developed below, path lengths in $\mathcal{G}^n_{S,cyclic}$ (e.g., Fig. 7c) measure differences from the average. As $V_1 \xrightarrow{-2} V_3 \xrightarrow{0} V_2 \xrightarrow{-2} V_4 \xrightarrow{0} V_1$ in Fig. 7c has the shortest length of -4, this defines the shortest Fig. 7a path that has length -4 from the average of 60, or 56.

3.2. **Computing $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cyclic}$.** Computing the $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cyclic}$ for a $\mathcal{G}^n_S$ follows the lead of Sect. 2.2. This is because $\mathcal{G}^n_{S,cpi}$ is the orthogonal projection of $\mathcal{G}^n_S$ to $\mathbb{CPI}^n_S$, and a basis for $\mathbb{CPI}^n_S$ is known (Thm. 9). Entries for $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cyclic}$ are based on the following.

Definition 5. *For V_j of $\mathcal{G}^n_S \in \mathbb{G}^n_S$, let $\mathcal{S}_S(V_j)$ be the sum of the arc lengths attached to vertex V_j, $j = 1, \ldots, n$. Let $T(\mathcal{G}^n_S) = \frac{1}{n-1}\sum_{j=1}^n \mathcal{S}_S(V_j)$.*

Because $\frac{1}{n-1}S(V_j)$ is the average length of an arc with V_j as a vertex, it follows that $T(\mathcal{G}^n_S)$ is the *average $\mathcal{G}^n_S$ Hamiltonian path length.* As $\mathcal{G}^n_{S,cyclic}$ consists of $\mathbf{b}^n_{i,j,k,s}$ cycles, each $\mathbf{b}^n_{i,j,k,s}$ arc entering a vertex has a leaving arc with the same weight but opposite sign, so $S_S(V_j) = 0$. This equation requires the $S_S(V_j)$ values for $\mathcal{G}^n_S$ and $\mathcal{G}^n_{S,cpi}$ to agree. Because $T(\mathcal{G}^n_S)$ sums these values, the average Hamiltonian path lengths in $\mathcal{G}^n_S$ and in $\mathcal{G}^n_{S,cpi}$ agree (as suggested with Fig. 7), or

$$T(\mathcal{G}^n_S) = T(\mathcal{G}^n_{S,cpi}) = 2\sum_{j=1}^n \omega_j. \tag{11}$$

Agreement between $S_S(V_j)$ values in $\mathcal{G}_S$ and $\mathcal{G}^n_{S,cpi}$ provides equations for the unknowns $\{\omega_j\}^n_{j=1}$. Illustrating with Fig. 8a, as $S_S(V_1) = 81$ for $\mathcal{G}^5_S$, the same value holds for $\mathcal{G}^5_{S,cpi}$, which means that $\sum_{j=2}^5(\omega_1 + \omega_j) = 81$. In general, the unknown $\{\omega_j\}^n_{j=1}$ satisfy

$$S_S(V_j) = \sum_{k\neq j}(\omega_j + \omega_k) = (n-1)\omega_j + \sum_{k\neq j}\omega_k = (n-2)\omega_j + \sum_{k=1}^n \omega_k = (n-2)\omega_j + \frac{1}{2}T(\mathcal{G}^n_{S,cpi}), . \tag{12}$$

Using $T(\mathcal{G}^n_S) = \mathcal{T}(\mathcal{G}^n_{S,cpi})$ (Eq. 11), the values of the $\mathcal{G}^n_{S,cpi}$ weights are

$$\omega_j = \frac{1}{n-2}[S_S(V_j) - \frac{1}{2}T(\mathcal{G}^n_S)],\ j = 1, 2, \ldots, n. \tag{13}$$

These ω_j weights, which define $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cyclic}$, lead to a result concerning path lengths.

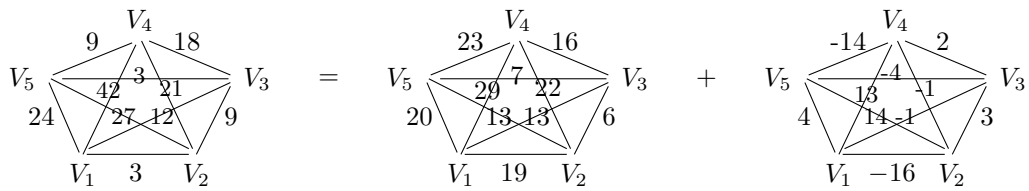
Theorem 12. *For $\mathcal{G}^n_S \in \mathbb{G}^n_S$, Eq. 13 defines the weights of its $\mathcal{G}^n_{S,cpi}$ component. Let $\mathcal{G}^n_{S,cyclic} = \mathcal{G}^n_S - \mathcal{G}^n_{S,cpi}$. The $\mathcal{G}^n_S$ length of a Hamiltonian circuit equals $T(\mathcal{G}^n_S)$ plus its $\mathcal{G}^n_{S,cyclic}$ path length.*

Proof: A $\mathcal{G}^n_S$ path length is the sum of its $\mathcal{G}^n_{S,cpi}$ and $\mathcal{G}^n_{S,cpi}$ lengths. All $\mathcal{G}^n_{S,cpi}$ Hamiltonian paths have length $T(\mathcal{G}^n_{S,cpi}) = T(\mathcal{G}^n_S)$, so Thm. 12 follows. □

According to Thm. 12, all essential closed path properties of $\mathcal{G}^n_S$ are based on the structure of $\mathcal{G}^n_{S,cyclic}$ and its four-cycle symmetries. Thus, general properties characterizing $\mathcal{G}^n_{S,cyclic}$ are useful.

Corollary 6. *If $\mathcal{G}^n_S$ has the property that $S_S(V_j) = 0$, $j = 1, \ldots, n$, then $\mathcal{G}^n_S \in \mathbb{C}^n_S$.*

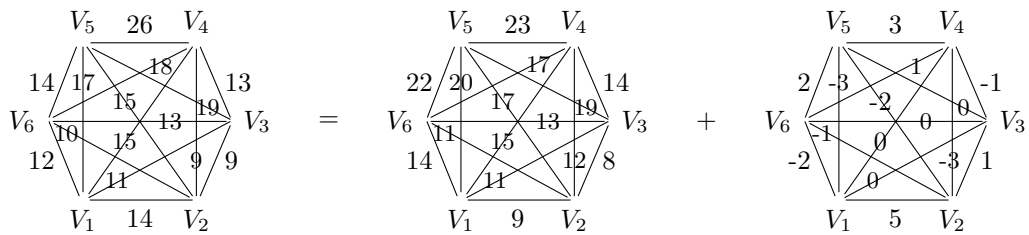
Proof: This condition requires $T(\mathcal{G}^n_S) = 0$ (Def. 5) and $\omega_j = 0$, $j = 1, \ldots, n$ (Eq. 13). As $\mathcal{G}^n_{S,cpi} = 0$, it follows that $\mathcal{G}^n_S = \mathcal{G}^n_{S,cyclic}$. □



a. Original; $\mathcal{G}_S^5$ **b.** $\mathcal{G}_{S,cpi}^5$ **c.** $\mathcal{G}_{S,cyclic}^5$

Figure 8. Decomposition of a $\mathcal{G}_S^5 \in \mathbb{G}_S^5$

To illustrate Thm. 12, the Fig. 8a computations from $\mathcal{G}_S^5$ are $S_S(V_1) = 81, S_S(V_2) = 60, S_S(V_3) = 42, S_S(V_4) = 90, S_S(V_5) = 63$, so $T(\mathcal{G}_S^5) = 84$. This means that (Eq. 13) $\omega_1 = \frac{1}{3}[81 - 42] = 13$, $\omega_2 = 6$, $\omega_3 = 0$, $\omega_4 = 16$, $\omega_5 = 7$, from which $\mathcal{G}_{S,cpi}^5$ and $\mathcal{G}_{S,cyclic}^5$ of Figs. 8b, c follow.



a. $\mathcal{G}_S^6$ **b.** $\mathcal{G}_{S,cpi}^6$ **c.** $\mathcal{G}_{S,cyclic}^6$

Figure 9. A $\mathcal{G}_S^6$

The Fig. 9 six-alternative example is similarly obtained. The $\mathcal{G}_{S,cpi}^6$ weights are $\omega_1 = 6, \omega_2 = 3, \omega_3 = 5, \omega_4 = 9, \omega_5 = 14, \omega_6 = 8$. As required by Def. 4, for any rectangle in $\mathcal{G}_{S,cpi}^6$ (Fig. 9b), the sums of its horizontal edges, its vertical edges, and its diagonals are the same. For any five vertices, the lengths of any $\mathcal{G}_{S,cpi}^6$ closed curves meeting all five vertices once are the same. All $\mathcal{G}_{S,cpi}^6$ Hamiltonian paths have the same length. Similar to Cor. 5, $\mathcal{G}_{S,cpi}^n$ can dominate the $\mathcal{G}_S^n$ format.

Turning to $\mathcal{G}_{S,cyclic}^n$, negative arc values normally are avoided with symmetric costs because cycling can generate an arbitrarily small path length. This problem is sidestepped here because such cycling increases the value that replaces $T(\mathcal{G}_{S,cpi}^n)$ in Thm. 12; e.g., if each vertex is met twice, then the value is $2T(\mathcal{G}_{S,cpi}^n) = 2(2\sum_{j=1}^n \omega_j)$. As $\mathcal{G}_{S,cyclic}^n$ entries indicate "differences from average," following arcs with negative lengths is following "below average cost" arcs; a concept that does not exist for $\mathcal{G}_S^n$.

An importance of the reduction is that the $\mathcal{G}_{S,cyclic}^n$ arc lengths have a distinct meaning for path lengths, so even the GA can be successful where it would fail with $\mathcal{G}_S^n$. With Fig. 8c, the GA identifies shortest $\mathcal{G}_{S,cyclic}^n$ Hamiltonian circuit of $V_1 \xrightarrow{-16} V_2 \xrightarrow{-1} V_4 \xrightarrow{-14} V_5 \xrightarrow{-4} V_3 \xrightarrow{-1} V_1$, which uses all five negative cost arcs, has length -36. Its $\mathcal{G}_S^5$ length (Thm. 12) is $T(\mathcal{G}_S^5) - 36 = 84 - 36 = 48$. But GA is thrown off the track with $\mathcal{G}_S^5$ (Fig. 8a) because of the $\mathcal{G}_{S,cpi}^5$ terms.

Similarly, the GA identifies the shortest Fig. 9c Hamiltonian path $V_1 \xrightarrow{-3} V_5 \xrightarrow{-2} V_2 \xrightarrow{-3} V_4 \xrightarrow{-1} V_3 \xrightarrow{0} V_6 \xrightarrow{-2} V_1$ of -11. Using the ω_j values for Fig. 9b, $T(\mathcal{G}_{S,cpi}^6) = 90$, so the shortest Hamiltonian path in Fig. 9a is 11 below this average, or $90 - 11 = 79$. Again, the GA fails for $\mathcal{G}_S^6$ because the $\mathcal{G}_{S,cpi}^6$ entries divert it.

The construction leads to an easily computed lower bound for Hamiltonian path lengths.

Corollary 7. *For $\mathcal{G}_S^n$, let the adjustment $\mathcal{A}(\mathcal{G}_{S,cyclic}^n)$ be the sum of the n smallest arc lengths in $\mathcal{G}_{S,cyclic}^n$. All $\mathcal{G}_S^n$ Hamiltonian path lengths are bounded below by $T(\mathcal{G}_S^n) + \mathcal{A}(\mathcal{G}_{S,cyclic}^n)$. The shortest Hamiltonian graph is bounded above by $T(\mathcal{G}_S^n)$,*

The last statement follows because $T(\mathcal{G}_S^n)$ is the average length of a Hamiltonian path. Thus some Hamiltonian path length is smaller than $T(\mathcal{G}_S^n)$ and $\mathcal{A}(\mathcal{G}_{S,cyclic}^n) < 0$ iff $\mathcal{G}_{S,cyclic}^n \neq 0$. For Fig. 8, $\mathcal{A}(\mathcal{G}_{S,cyclic}^5) = -36$, so the lower bound is $84 - 36 = 48$, which equals the length of the shortest Hamiltonian path. With Fig. 9, $\mathcal{A}(\mathcal{G}_{S,cyclic}^5) = -12$ for the lower bound of $90 - 12 = 78$, but the shortest Hamiltonian path has the larger length of 79. The reason is that the -1 length of $\widehat{V_2V_6}$ can not be used. By using the four-cycle geometry, sharper estimates can be derived.

Closely related to Cor. 7 is an approach to find the shortest Hamiltonian tour by ranking $\mathcal{G}_{S,cyclic}^n$ arcs according to their lengths where "smaller is better." If marking first n shortest arcs does not define a Hamiltonian tour, iteratively add arcs from this list until the marked legs do define such a closed path. (All of the shortest Hamiltonian circuits in this section were verified in this simple manner. This approach can be improved by using properties of the four-cycles.)

a. From Fig. 8c **b.** From Fig. 9c.

Figure 10. Finding paths

To illustrate with Fig. 10a (from Fig.8c), just marking the five legs with negative costs already defines the shortest Hamiltonian graph. In Fig. 10b, the six legs with the smallest (all negative) values do not define a Hamiltonian circuit, so add an additional leg with the next smallest cost (here zero). The $V_3 \xrightarrow{0} V_6$ arc completes the $V_1 \xrightarrow{-3} V_5 \xrightarrow{-2} V_2 \xrightarrow{-3} V_4 \xrightarrow{-1} V_3 \xrightarrow{0} V_6 \xrightarrow{-2} V_1$ circuit, which, by construction, is the shortest.

This approach applies to other types of closed paths. Suppose the goal in Fig. 9 is to find find the shortest closed path that passes once through each of the four vertices $\{V_1, V_2, V_4, V_5\}$. The six $\mathcal{G}_{S,cyclic}^6$ arc lengths of these vertices are $\{-3, -3, -2, 0, 3, 5\}$ where marking the first four on $\mathcal{G}_{S,cyclic}^6$ already defines the minimal closed path $V_2 \xrightarrow{-3} V_4 \xrightarrow{0} V_1 \xrightarrow{-3} V_5 \xrightarrow{-2} V_2$ of length -8. This set's T value is $2(\omega_1 + \omega_2 + \omega_4 + \omega_5)$, which, in $\mathcal{G}_{S,cpi}^6$, is the sum of the rectangle's vertical and horizontal legs or 64. So the length of this shortest $\mathcal{G}_S^6$ closed path over these vertices is $64 - 8 = 56$.

3.3. **Four cycle structure.** A complication in determining which four-cycles define a given $\mathcal{G}_{S,cyclic}^n$ is that some of these four-cycles must overlap on certain edges. To handle this complexity, the switching, iterative approach used in the proof of Thm. 10 is used.

Theorem 13. *To express a $\mathcal{G}_{S,cycle}^n \in \mathbb{C}_S^n$ in terms of the basis in Eq. 10, for $4 \leq j < k$, the multiple of $\mathbf{b}_{1,2,j,k}^n \in \mathcal{A}_{1,2}^n$ is $d_{j,k}$ from the $V_j \xrightarrow{d_{j,k}} V_k$ arc in $\mathcal{G}_{S,cycle}^n$. (If the arc is not in the graph, its value is zero.) After determining the multiple of a basis vector, remove the associated four-cycle from the graph. In what remains, the multiple of the top $\mathbf{b}_{1,2,n,3}^n \in \mathcal{B}_{1,2}^n$ is the negative of the $d_{2,n}$ value in*

the of $V_2 \xrightarrow{d_{2,n}} V_n$ *arc in the reduced graph.*[3] *After removing this four-cycle, the top remaining* $\mathcal{A}^n_{1,2}$ *vector is* $\mathbf{b}^n_{1,2,3,n}$*; its coefficient is the length in* $V_3 \xrightarrow{d_{3,n}} V_n$ *in the reduced graph, which leaves* $n-1$ *as the largest remaining index in the reduced graph. In general, if the largest remaining index is* s*, the multiple of the top remaining* $\mathcal{B}^n_{1,2}$ *vector,* $\mathbf{b}^n_{1,2,s,3}$*, is the negative of* $d_{2,s}$ *from the reduced graph's* $V_2 \xrightarrow{d_{2,s}} V_s$ *arc. The top of the remaining* $\mathcal{A}^n_{1,2}$ *vectors is* $\mathbf{b}^n_{1,2,3,s}$*; its multiple is the* $d_{3,s}$ *value of the* $V_3 \xrightarrow{d_{3,s}} V_s$ *arc in the reduced graph.*

Proof, The proof is essentially that of Thm. 10; removing basis vectors in the specified manner leaves, at each stage, a single $d_{u,v}$ value of a certain type. Because $\mathbb{C}^n_S$ is the sum of these four-cycles, the existence of this $d_{u,v} \neq 0$ requires the associated $\mathbf{b}^n_{1,2,k,s}$ to be in the decomposition; the form of this four-cycle requires $d_{u,v}$ to be the vector's multiple. A difference is that if $d_{u,v}$ identifies a vector from $\mathcal{B}^n_{1,2}$, the multiple is the negative of $d_{u,v}$, as required by the form of the associated four-cycle. If the vector is from $\mathcal{A}^n_{1,2}$, then $d_{u,v}$ is the multiple. □

Using this approach, the four cycles of $\mathcal{G}^5_{S,cyclic}$ in Fig. 8c are $-14\mathbf{b}^5_{1,2,\mathbf{4},5}$, $-\mathbf{b}^5_{1,\mathbf{2},\mathbf{5},3}$, $10\mathbf{b}^5_{1,2,\mathbf{3},\mathbf{5}}$, $-13\mathbf{b}^5_{1,\mathbf{2},\mathbf{3},4}$ and $15\mathbf{b}^5_{1,2,4,3}$.

3.4. **Extensions.** With minor modifications, all other results developed in Sect. 2 for the asymmetric $\mathbb{G}^n_A$ transfer to the symmetric $\mathbb{G}^n_S$. For instance, to analyze connected and closed path properties that involve a subset of vertices, carry out the above with that subset. Other samples follow.

Theorem 14. *Consider the class of paths starting at* V_j *and ending at* V_k *that pass through vertices* $\{V_i\}_{i\in\mathcal{D}}$ *where, for each* $i \in \mathcal{D}$*, the path passes through* V_i κ_i *times. The length of such a path in* $\mathcal{G}^n_S$ *is its path length in* $\mathcal{G}^n_{S,cyclic}$ *plus* $(\omega_j + \omega_k) + 2\sum_{i\in\mathcal{D}} \kappa_i\omega_i$*.*

As an example, consider all paths in Fig. 9 that start at V_1, end in V_5 and pass through each of V_2, V_3, V_4 twice. According to the weights of $\mathcal{G}^6_{S,cpi}$, the length of any of these paths in $\mathcal{G}^6_S$ is its length in $\mathcal{G}^6_{S,cyclic}$ plus its $\mathcal{G}^6_{S,cpi}$ length of $(6+4)+4(3+5+4)$.

Graphs with incomplete symmetric costs are handled the same way as in Sect. 2. That is, complete the graph by adding arcs of any desired length to obtain $\tilde{\mathcal{G}}^n_S$. Then compute $\tilde{\mathcal{G}}^n_{S,cpi}$ and $\tilde{\mathcal{G}}^n_{S,cyclic}$ For incomplete graphs, ∞ often is assigned to inadmissible arcs; do so only with $\tilde{\mathcal{G}}^n_{S,cyclic}$.

Theorem 15. *For an incomplete symmetric graph* $\mathcal{G}^n_S$*, let* $\tilde{\mathcal{G}}^n_S$ *include the missing* $\mathcal{G}^n_S$ *arcs where each has an arbitrary selected length. Compute* $T(\tilde{\mathcal{G}}^n_S)$ *and* $\tilde{\mathcal{G}}^n_{S,cyclic}$*. The length of a* $\mathcal{G}^n_S$ *Hamiltonian path is* $T(\tilde{\mathcal{G}}^n_S)$ *plus its* $\tilde{\mathcal{G}}^n_{S,cyclic}$ *length.*

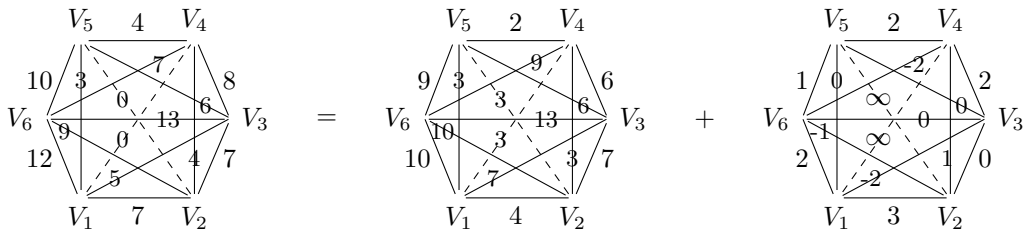
Computations can be simplified by adding arcs of zero length so that the $S_S(V_j)$ values for $\mathcal{G}^n_S$ and $\tilde{\mathcal{G}}^n_S$ agree, and $T(\mathcal{G}^n_S) = T(\tilde{\mathcal{G}}^n_S)$.

Proof: A Hamiltonian path length in $\mathcal{G}^n_S$ is the same in $\tilde{\mathcal{G}}^n_S$, which equals $T(\tilde{\mathcal{G}}^n_S)$ plus its length in $\tilde{\mathcal{G}}^n{}_{S,cyclic}$. The result follows. □

As $\widehat{V_1V_4}$ and $\widehat{V_2V_5}$ are not admitted in Fig. 11, include them in Fig. 11a with zero lengths (the two dashed arcs in Fig. 11a). The $S_S(V_j)$ values of $\tilde{\mathcal{G}}^n_S$ are $S_S(V_1) = 27, S_S(V_2) = 27, S_S(V_3) = 39, S_S(V_4) = 23, S_S(V_5) = 23, S_S(V_6) = 51$. Thus $T(\tilde{\mathcal{G}}^n_S) = 38$, $\omega_1 = \frac{1}{4}[27-19] = 2, \omega_2 = 2, \omega_3 = 5, \omega_4 = 1, \omega_5 = 1, \omega_6 = 8$, and Figs. 10b, c follow. The two inadmissible Fig. 11c arcs (both

[3]The associated arc for $\mathbf{b}^n_{1,2,n,3}$ is $V_1 \xrightarrow{1} V_2 \xrightarrow{-1} V_n \xrightarrow{1} V_3 \xrightarrow{-1} V_1$, so for $V_2 \xrightarrow{d_{2,n}} V_n$ to hold, the coefficient for $\mathbf{b}^n_{1,2,n,3}$ must be the negative of $d_{2,n}$.

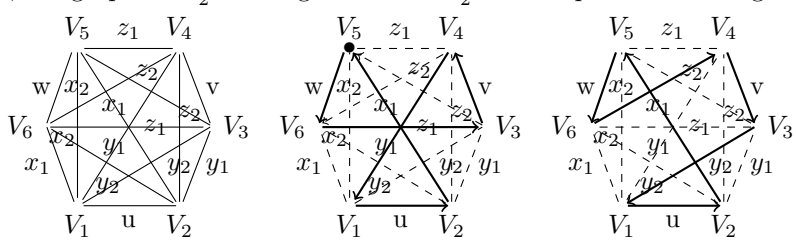
with length of -3) could be dropped or, as in Fig. 11c, replaced with ∞. The shortest $\mathcal{G}^6_{S,cyclic}$ Hamiltonian path $V_1 \xrightarrow{-2} V_3 \xrightarrow{0} V_2 \xrightarrow{-1} V_6 \xrightarrow{-2} V_4 \xrightarrow{2} V_5 \xrightarrow{0} V_1$ of length -3, which includes all allowed arcs with negative costs, can be found in the above described manner. In $\mathcal{G}^6_S$ this path has the "below average" length of $T(\mathcal{G}^6_S) - 3 = 35$.



a. From $\mathcal{G}^6_S$ to $\widetilde{\mathcal{G}^6_S}$ **b.** $\widetilde{\mathcal{G}^6_{S,cpi}}$ **c.** $\widetilde{\mathcal{G}^6_{S,cyclic}}$

Figure 11. An incomplete $\mathcal{G}^6_S$

While the method associated with Fig. 10 is more efficient than the GA, it is worth using the structure of the decomposition to explain certain GA traits. Parallel to the $\mathcal{G}^n_{A,cyclic}$ concern whether a vertex can be a source, a $\mathcal{G}^n_{S,cyclic}$ worry is whether all of a vertex's legs are positive. But $S_S(V_j) = 0$, so this cannot happen. Thus, if all $\mathcal{G}^n_{S,cyclic}$ vertices have an arc with non-zero length, then the number of negative length arcs is bounded below by $\frac{n}{2}$ and above by $\frac{n(n-2)}{2}$. Dropping $\mathcal{G}^n_{S,cpi}$ eliminates one GA difficulty, but another is caused by the number of options. This can be seen with Fig. 12a, which is given by $\{-x_1\mathbf{b}^6_{1,2,5,6} - x_2\mathbf{b}^6_{1,2,6,5}\}$, $\{-y_1\mathbf{b}^6_{1,2,3,4}, -y_2\mathbf{b}^6_{1,2,4,3}\}$, and $\{-z_1\mathbf{b}^6_{3,4,5,6}, -z_2\mathbf{b}^6_{3,4,6,5}\}$, where each bracket defines a rectangle. Thus $u = -(x_1 + x_2) - (y_1 + y_2)$, $v = -(y_1 + y_2) - (z_1 + z_2)$, and $w = -(x_1 + x_2) - (z_1 + z_2)$. Should the x's, y's, and z's have positive values, the graph has $\frac{n}{2} = 3$ negative and $\frac{n(n-2)}{2} = 12$ positive arc lengths.



a. A choice of $\mathcal{G}^6_{S,cyclic}$ **b.** Selecting z_1 **c.** Selecting z_2

Figure 12. Potential failures of the Greedy Algorithm

It follows from the material following Cor. 7 that a shortest Hamiltonian path must include the three arcs with negative lengths (of u, v, and w). Thus the four-cycle symmetry requires all Hamiltonian paths of Fig. 12a with these negative length arcs to have one of only four sizes where $u + v + w$ is supplemented by $x_1 + y_1 + z_1, x_1 + y_2 + z_2, x_2 + y_1 + z_1$ or $x_2 + y_2 + z_2$. Starting the GA at V_1 and assuming that x_1 is smaller than x_2, y_1, y_2, the first three moves (Figs. 12b, c) are $V_1 \xrightarrow{u} V_2 \xrightarrow{x_1} V_5 \xrightarrow{w} V_6$. The next move is to the $\widehat{V_3V_4}$ arc. Whichever way $\widehat{V_3V_4}$ is entered determines the last arc.

Should $z_1 < z_2$, then Fig. 12b represents the fourth GA step $V_6 \xrightarrow{z_1} V_3$. But should y_1 be much larger than y_2, the Fig. 12c route would be shorter. As the arrangement of the $\mathcal{G}^n_{Scyclic}$ four-cycles can affect the success of an algorithm, the algebra of these four-cycles needs to be better understood.

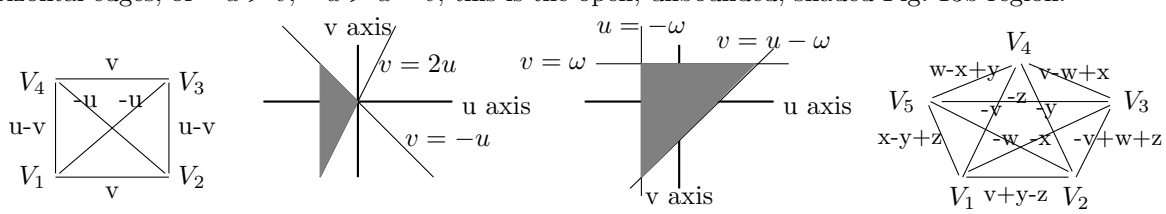
For small values of n, and theoretically for all values, the basis for $\mathbb{C}^n_S$ exhibits all possible $\mathcal{G}^n_{S,cyclic}$ choices. To illustrate an application, if arc costs represent Euclidean distances in the planar problem and the triangle inequality is satisfied, then minimal Hamiltonian paths cannot have a self intersection. It is reasonable to wonder where else does the triangle inequality ensure this behavior. Here, the structure of $\mathcal{G}^n_{S,cpi}$ plays a role.

Theorem 16. *For $n \geq 3$, $\mathcal{G}^n_{S,cpi}$ satisfies the triangle inequality iff all weights ω_j are non-negative.*

According to Eq. 13, $\omega_k < 0$ represents where the associated $S_S(V_k)$ is bounded above by $\frac{1}{2}T(\mathcal{G}^n_S)$. That is, the average of arc lengths attached to V_k is much smaller than average over the graph.

Proof: The $\mathcal{G}^n_{S,cpi}$ arc length for $\widehat{V_iV_j}$ is $\omega_i + \omega_j$. For a triplet, the length of the two arcs $\widehat{V_iV_j}$ and $\widehat{V_jV_k}$ is $\omega_i + 2\omega_j + \omega_k$, which differs from the $\widehat{V_iV_k}$ arc length by $2\omega_j$. Thus, the triangle inequality is satisfied iff $\omega_j \geq 0$. This must hold for all legs of all triplets, so $\omega_j \geq 0$ for all j. □

Turning to $n = 4$, Fig. 13a represents all possible $\mathcal{G}^4_{S,cyclic}$ structures (with $u\mathbf{b}^4_{1,4,2,3} + v\mathbf{b}^4_{1,2,3,4}$), so it characterizes all closed path properties and their lengths for $\mathcal{G}^4_S \in \mathbb{G}^4_s$. Assuming the rectangular Fig. 13a faithfully represents the geometry of a considered concern (e.g., using actual rather than Euclidean costs), the issue is to understand which (u, v) values (that is, which $\mathcal{G}^4_{S,cyclic}$) require a shortest Hamiltonian path to avoid the crossing diagonals. The answer follows from Fig. 13a as it requires the length of the two diagonals to be greater than that of the two vertical and the two horizontal edges, or $-u > v$, $-u > u - v$; this is the open, unbounded, shaded Fig. 13b region.



a. All $\mathcal{G}^4_{S,cyclic}$ **b.** No diagonal crossings **c.** Triangle inequality **d.** All $\mathcal{G}^5_{S,cyclic}$ choices

Figure 13. Finding properties of $\mathcal{G}^n_{S,cyclic}$

To compare this wedge with what happens should $\mathcal{G}^4_S$ satisfy the triangle inequality, it follows from Thm. 16 that in a $\{V_i, V_j, V_k\}$ triangle, the sum of the $\widehat{V_iV_k}$ and $\widehat{V_kV_j}$ leg lengths in $\mathcal{G}^4_{S,cyclic}$ plus $2\omega_k$ must be an upper bound for $\widehat{V_iV_j}$'s $\mathcal{G}^4_{S,cyclic}$ leg length. Applying this to the three triangles where vertex V_k is off the triangle's compared edge leads to

$$\omega_k \geq -u, \quad \omega_k \geq v, \quad \omega_k \geq u - v. \tag{14}$$

Theorem 17. *For $\mathcal{G}^4_S$, if any $\omega_j < 0$, then $\mathcal{G}^4_S$ does not satisfy the triangle inequality. Let $\omega = \min(\omega_1, \omega_2, \omega_3, \omega_4)$. The region where $\mathcal{G}^4_S$ satisfies the triangle inequality is defined by substituting ω for ω_k in Eq. 14; it is depicted by the closed shaded triangle in Fig. 13c.*

Proof: If $\omega_k < 0$. then Eq. 14 cannot be satisfied. The remainder follows from the above. □

If $\omega = 0$, the triangle inequality is satisfied only for $u = v = 0$, which is $\mathcal{G}^4_{S,cyclic} = 0$ so $\mathcal{G}^4_S = \mathcal{G}^4_{S,cpi}$. Both the triangle inequality and the non-crossing of the diagonals in the shortest Hamiltonian path hold in the intersection of the shaded portions of Figs. 13b, c; this is the Fig. 13b shaded triangle limited on the left by $u \geq -\omega$. What remains are regions (i.e., choices of $\mathcal{G}^4_{S,cyclic}$) where the triangle inequality is satisfied but the shortest Hamiltonian path includes the diagonals,

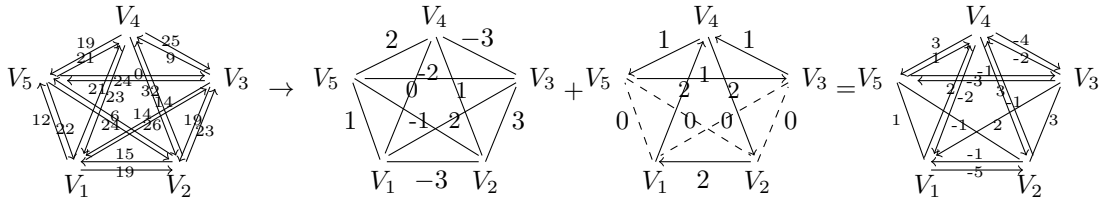
and a sizable region (the shaded Fig. 13b region for $u < -\omega$) where the diagonals are not in the shortest Hamiltonian circuit and the triangle inequality is not satisfied.

Results for $\mathcal{G}^5_{S,cyclic}$ follow in a similar manner. The basis for Fig. 13d, which captures all $\mathcal{G}^5_{S,cyclic}$ behaviors, is $\{v\mathbf{b}^5_{1,2,3,4} + w\mathbf{b}^5_{2,3,4,5} + x\mathbf{b}^5_{3,4,5,1} + y\mathbf{b}^5_{4,5,1,2} + z\mathbf{b}^5_{5,1,2,3}\}$.

4. Graphs with general asymmetric costs

Other systems can be similarly reduced. Graphs where all closed paths have a fixed length identify components of $\mathcal{G}^n$ entries that frustrate a closed path analysis. The subspace's normal bundle measures deviations from neutrality, so it is critical when determining closed path properties.

None of this is necessary for the standard space of graphs with asymmetric costs. The reason is that, for each pair, the arc lengths $V_j \xrightarrow{x} V_k$ and $V_k \xrightarrow{y} V_j$ can be represented as an {average cost, excess cost} pair; e.g., $\{a = \frac{x+y}{2}, V_j \xrightarrow{x-a} V_k\}$. By applying the Sects. 2 and 3 approaches to each component, the above results about incomplete graphs, path lengths, etc., transfer.



a. A general $\mathcal{G}^5$ **b.** $\mathcal{G}^5_{S,cyclic}$. **c.** $\mathcal{G}^5_{A.cyclic}$ **d.** Reduced $\mathcal{G}^5$

Figure 14. Decomposing an asymmetric $\mathcal{G}^5$

To illustrate this program with Fig. 14, by representing the Fig. 14a costs as {average cost, excess cost} pairs, the original graph becomes $\mathcal{G}^5 = \mathcal{G}^5_S + \mathcal{G}^5_A \in \mathbb{G}^5_S \times \mathbb{G}^5_A$, where $\widehat{V_jV_k}$'s length in $\mathcal{G}^5_S$ is the average cost of its arcs, and $\mathcal{G}^5_A$ represents how costs differ from the average. Thus, with $V_1 \xrightarrow{26} V_3$ and $V_3 \xrightarrow{14} V_1$ from Fig. 14a, $\widehat{V_1V_3}$'s length in $\mathcal{G}^5_S$ is 20 and $\mathcal{G}^5_A$ has $V_1 \xrightarrow{6} V_3$.

The analysis of $\mathcal{G}^5_S + \mathcal{G}^5_A$ follows as above: find each graph's cpi and cyclic components. Removing $\mathcal{G}^5_{S,cpi}$ and $\mathcal{G}^5_{A,cpi}$ leaves Figs. 14b, c. A $\mathcal{G}^5$ Hamiltonian path length (with $\mathcal{G}^5_{S,cpi}$ weights $\omega_1 = 10$, $\omega_2 = 10$, $\omega_3 = 8$, $\omega_4 = 12$, $\omega_5 = 6$) equals $T(\mathcal{G}^5_{S,cpi}) = 92$ plus the sum of its $\mathcal{G}^5_{S,cyclic}$ and $\mathcal{G}^5_{A,cyclic}$ lengths. Expressing $\mathcal{G}^5_{S,cyclic} + \mathcal{G}^5_{A,cyclic}$ in a standard Fig. 14d form, its shortest Hamiltonian path of $V_1 \xrightarrow{-5} V_2 \xrightarrow{-1} V_4 \xrightarrow{-4} V_3 \xrightarrow{-3} V_5 \xrightarrow{1} V_1$ follows. Its $\mathcal{G}^5$ path length is $T(\mathcal{G}^5_{S,cpi}) - 12 = 80$.

5. Summary

Components of a graph's entries that hamper finding closed path properties are identified. Eliminating them determines the graph's essence—a reduced graph with smaller degrees of freedom where all closed graph properties are expressed in terms of inherent symmetry structures.

6. Proofs

Results not proved above or in [4], are proved here.

Proof of Theorem 1: For a triplet $\{V_i, V_j, V_k\}$ in a cpi graph, the closed path $V_i \xrightarrow{x} V_j \xrightarrow{y} V_k \xrightarrow{z} V_i$ has path length zero, so $x + y = -z$. Thus, $V_i \xrightarrow{x} V_j \xrightarrow{y} V_k = V_i \xrightarrow{-z=x+y} V_k$ satisfies Eq. 3. As all triplets satisfy Eq. 3, a cpi graph is strongly transitive.

A triplet $\{V_i, V_j, V_s\}$ in a strongly transitive graph satisfies $V_i \xrightarrow{x} V_j \xrightarrow{y} V_s = V_i \xrightarrow{z} V_s$ where $z = x + y$. Applying a fourth alternative V_t to this relationship yields

$$(V_i \xrightarrow{x} V_j \xrightarrow{y} V_s) \xrightarrow{u} V_t = (V_i \xrightarrow{z} V_s) \xrightarrow{u} V_t = V_i \xrightarrow{w} V_t,$$

where $w = z + u = x + y + u$. With the obvious induction argument, it follows that any path from V_i to V_k has the same length as the direct path from V_i to V_k. (This proves Cor. 1.) A closed path has $V_k = V_i$, so its length is that of V_i to V_i, or zero. Hence, a strongly transitive graph is cpi.

To prove that the set of strongly transitive graphs forms a linear subspace, notice that a multiple μ of a strongly transitive graph in $\mathbb{G}_A^n$ changes all path lengths by this multiple; thus the new graph's arcs remain strongly transitive. Therefore the multiple defines another $\mathbb{G}_A^n$ strongly transitive graph. (If $\mu < 0$, then positive cost arcs in the original graph become negative cost arcs in the new graph.) Similarly, for two strongly transitive $\mathbb{G}_A^n$ graphs and any $\{V_i, V_j, V_k\}$ triplet, the first graph satisfies $V_i \xrightarrow{x} V_j \xrightarrow{y} V_k = V_i \xrightarrow{x+y} V_k$ while the second satisfies $V_i \xrightarrow{\tilde{x}} V_j \xrightarrow{\tilde{y}} V_k = V_i \xrightarrow{\tilde{x}+\tilde{y}} V_k$. Combining these graphs leads to $V_i \xrightarrow{x+\tilde{x}} V_j \xrightarrow{y+\tilde{y}} V_k = V_i \xrightarrow{(x+y)+(\tilde{x}+\tilde{y})} V_k$, which satisfies Eq. 3. Thus the set of strongly transitive graphs in $\mathbb{G}_A^n$, $\mathbb{ST}_A^n$, is a linear subspace. □

Proof of Cor. 2: The Eq. 4 basis of $\{V_1 \xrightarrow{1} V_j \xrightarrow{1} V_k \xrightarrow{1} V_1\}_{1<j<k\leq n}$ satisfies Cor. 2 because in this set, only the three cycle $V_1 \xrightarrow{1} V_s \xrightarrow{1} V_k \xrightarrow{1} V_1$ has a $\widehat{V_s V_k}$ arc. For the independence of the arcs, if $\mathbf{c}_{j,k}$ represents the only $\mathcal{CB}_A^n$ cycle with a $\widehat{V_j V_k}$ arc, it must be shown that $\sum x_{j,k}\mathbf{c}_{j,k} = \mathbf{0}$ iff all $x_{j,k} = 0$. But as $\mathbf{c}_{j,k}$ is the only vector with a non-zero j, k component, $x_{j,k} = 0$.

That theses cycles are in $\mathbb{ST}_A^n$'s normal bundle is proved in [4]. As this set consists of $\binom{n-1}{2}$ linearly independent elements that are orthogonal to $\mathbb{ST}_A^n$, it is a basis for the normal bundle. □

Proof of Thm. 3: Equation 5 is an immediate consequence of the representation of $\mathbb{G}_A^n$ into the orthogonal subspaces $\mathbb{ST}_A^n$ and $\mathbb{C}_A^n$. The last comment is proved above.□

Proof of Cor. 5: That the Def. 3 relationship is an equivalence relationship (reflexive, symmetricaA, transitive) follows immediately from the equality of the cyclic components. The difference between any two graphs in $\mathbb{G}_A^n$ is the difference between their cpi and cyclic components. As their cyclic components agree, the difference is the difference between cpi components. Because $\mathbb{ST}_A^n$ is a linear subspace, this difference also is in $\mathbb{ST}_A^n$. □

Proof of Thm. 6: The fact that $\mathcal{G}_{A,cpi}^n$ normally has a sink and source for positive and negative directions follows from the fact that all triplets are transitive, so there is a maximum and a minimum term. That its arc lengths are non-zero means that this top and bottom alternative are unique. For positive directions, the top alternative is a source, the bottom alternative is a sink.

The assertion that $\mathcal{G}_{A,cyclic}^n$ cannot have a source or a sink follows from the fact that $S_A(V_j) = 0$ for each vertex. (This statement follows from the fact that each three-cycle attache to a vertex has one leg pointing in and one leg, of same magnitude, pointing out.) Thus, each $\mathcal{G}_{A,cyclic}^n$ vertex with non-zero arcs has at least one positive direction pointing in and at least one pointing out.

What remains is to show that the longest Hamiltonian path in $\mathcal{G}_{A,cyclic}^n$, $n = 4, 5$, has all positive directions. It already has been shown that $\mathcal{G}_{A,cyclic}^n$ does not have a sink or source. The next possible problem is three-cycle with, say, positive cost arcs, e.g., $V_2 \to V_3 \to V_4 \to V_2$. To avoid having all positive arcs in the longest paths, the cycle must be attracting. But for $n = 4$, that would require all positive cost directions to point away from V_1, making V_1 a source, which it cannot be. For $n = 5$, all positive cost arrows from V_1 and V_5 point to the cycle. The positive cost arrow between V_1 and V_5 points away from one of these vertices, making it a source, which is a contradiction. An

attracting four cycle would make the remaining vertex a source. Everything extends in the same manner for sinks and for negative cost directions. □

Proof of Thm. 8: While the linear algebra proof used for $n = 4$ extends, an iterative argument provides insight. One set of closed paths involves the vertices $\{V_1, V_2, V_3, V_4\}$; these cpi graphs are based on weights ω_j attached to V_j, $j = 1, \ldots, 4$. Increasing the graph size to involve vertices $\{V_1, V_2, V_3, V_4, V_5\}$ requires analyzing all closed graphs in $\{V_1, V_2, V_3, V_5\}$. Again, the solution has weights ω'_k, $k = 1, 2, 3, 5$ assigned to the appropriate vertices. To be consistent with arc lengths in the first 4-tuple, it must be that $\omega'_k = \omega_k$ for $k = 1, 2, 3$. Continuing in this iterative manner extends the proof to all vertices. That the length of a closed path passing through the vertices $\{V_j\}_{j\in\mathcal{D}}$ once is $2\sum_{j\in\mathcal{D}} \omega_j$ is an immediate computation. As a path enters and leaves vertex V_j, the length is increased by $2\omega_j$. □

Proof of Thm. 9: Set $\{\mathbf{B}^n_j\}^{n-1}_{j=1}$ is independent because only $\mathbf{B}^n_j$ has a non-zero $d_{j,n}$ coordinate. If $\{\mathbf{B}^n_j\}^n_{j=1}$ is not independent, there is a summation $\sum_{j=1}^{n-1} x_j \mathbf{B}^n_j = \mathbf{B}^n_n$. In the sum, each $x_j = 1$ to capture $\mathbf{B}^n_n$'s $d_{j,n} = 1$ component. But then $d_{1,2} = 2$ (from $\mathbf{B}^n_1 + \mathbf{B}^n_2$), rather than the required zero of $\mathbf{B}^n_n$, so the linear subspace spanned by $\{\mathbf{B}^n_j\}^n_{j=1}$ is n-dimensional. This space captures the structure of $\mathcal{G}^n_{S,cpi}$ graphs because the $d_{i,j}$ component of $\sum_{s=1}^n \omega_s \mathbf{B}^n_s$ is the required $\omega_i + \omega_j$. □

Proof of Thm. 14: A path's length in $\mathcal{G}^n_S$ is the sum of its lengths in $\mathcal{G}^n_{S,cpi}$ and in $\mathcal{G}^n_{S,cyclic}$. □

References

[1] Arrow, K., (1963) *Social choice and individual values.* (New York, NY: Wiley).

[2] Cook, W. (2012) *In Pursuit of the Traveling Salesman*, (Princeton University Press, Princeton).

[3] Jessie, D., and D. G. Saari (2019) *Coordinate Systems for Games: Simplifying the "me" and "we" interactions."* (Springer, New York).

[4] Saari, D. G. (2021) Seeking consistency with paired comparisons: A systems approach, *Theory and Decisions*, 91 (3) 377-402.

Institute for Mathematical Behavioral Sciences, University of California, Irvine; Irvine CA 92697-5100

Email address: dsaari@uci.edu