The Neighbor Matrix: Generalizing A Graph's Degree Sequence

3 authors, including:

Jonathan Roginski
United States Military Academy West Point

24 PUBLICATIONS   **217** CITATIONS

SEE PROFILE

The Neighbor Matrix:
Generalizing A Graph's Degree Sequence

Jonathan W. Roginski*
Department of Applied Mathematics
Naval Postgraduate School, Monterey, CA
*Corresponding author: jonathan.w.roginski.mil@mail.mil

Ralucca M. Gera
Department of Applied Mathematics
Naval Postgraduate School, Monterey, CA

Erik C. Rye
Department of Applied Mathematics
Naval Postgraduate School, Monterey, CA

Abstract

The newly introduced neighborhood matrix extends the power of adjacency and distance matrices to describe the topology of graphs. The adjacency matrix enumerates which pairs of vertices share an edge and it may be summarized by the degree sequence, a list of the adjacency matrix row sums. The distance matrix shows more information, namely the length of shortest paths between vertex pairs. We introduce and explore the neighborhood matrix, which we have found to be an analog to the distance matrix what the degree sequence is to the adjacency matrix. The neighbor matrix includes the degree sequence as its first column and the sequence of all other distances in the graph up to the graph's diameter, enumerating the number of neighbors each vertex has at every distance present in the graph. We prove this matrix to contain eleven oft-used graph statistics and topological descriptors. We also provide insight into two applications that show potential utility of the neighbor matrix in comparing graphs and identifying topologically significant vertices in a graph.

Key words: adjacency matrix, distance, graph topology, centrality, graph power.

2010 Math Subject Classification: 05C07, 05C12, 05C50

1 Introduction

Matrices such as the adjacency, distance, reciprocal distance, walk, reachability, and Laplacian offer well-studied, compact structures that represent graph information [3, 6–10, 13–15]. Two of these matrix representations that form the foundation of this work are the adjacency matrix and the distance matrix. The adjacency matrix enumerates which pairs of vertices share an edge. The degree sequence is one summary of the adjacency matrix: a list of the adjacency matrix row sums. The distance matrix shows more information, namely the length of a shortest path between vertex pairs. We introduce and explore the neighbor matrix, a new matrix that extends the degree sequence to capture the distribution of each distance in a graph from 1 to k, where k is the graph diameter.

We need only consider Figure 1 to see the topological richness of the neighbor matrix. On the left in the figure is a depiction of the degree sequence of a Barabási-Albert graph on 100 vertices. The right figure depicts the distances accounted for by the neighbor matrix. The $distance - 1$ neighbors are exactly the degree sequence of the graph, depicted in white on the right. The illustration shows $distance - 2$ neighbors in the background of the right picture in medium gray, followed by $distance - 3$ neighbors in dark gray, and $distance - 4$ neighbors in light gray. We depict the small number of $distance - 5$ neighbors in the foreground of the picture, in black.
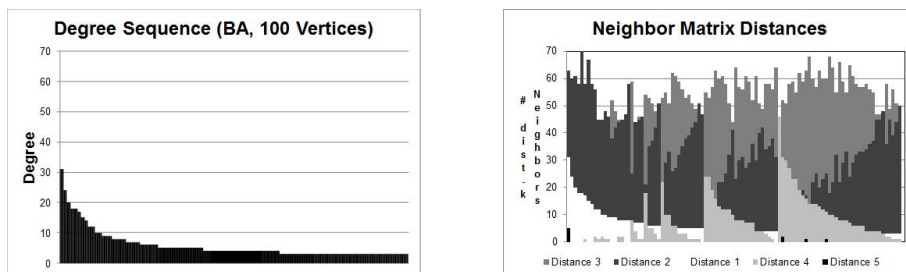

Figure 1: Comparing Topological Information of Neighbor Matrix to Degree Sequence

We use combinatoric techniques to show the neighbor matrix to include many of the statistics and topological characteristics currently used to describe graphs. Finally, we describe the use of the neighbor matrix in two applications that highlight its potential usefulness. We show that where current topological descriptors fail to discriminate between non-isomorphic graphs, the neighbor matrix may enable a graph comparison mechanism. We also present a methodology that leverages the topological information resident in the neighbors matrix to improve upon analysis using currently-defined centrality measures that seek to identify those vertices in the graph which exert the most influence over a graph's structure.

2 Definitions

We first review established definitions of graph theory and complex network terms that are fundamental to the work described in this paper.

This work is limited to analysis of *simple, connected* graphs (no multiple edges nor loops) $G = (V(G), E(G))$, where $V(G)$ is the vertex set and $E(G)$ the edge set. We annotate the vertices of a graph $v_1, v_2, \ldots, v_n$, where n is the number of vertices in the graph. Two vertices are called *adjacent* if there is an edge between them, otherwise they are *nonadjacent*. The edge set $E(G)$ contains each adjacency between vertices. A *cycle* is a sequence of adjacent vertices that begins and ends at the same vertex, but repeats no other vertex. A graph with no cycles is called *acyclic* and all connected, acyclic graphs are called *trees*. A graph H is a *subgraph* of G (denoted $H \subseteq G$) if $V(H) \subseteq V(G)$ and $E(H) \subseteq E(G)$. If $H \subseteq G$ and either $V(H)$ is a proper subset of $V(G)$ or $E(H)$ is a proper subset of $E(G)$, then H is a proper subgraph of G, $H \subset G$.

Two graphs, G and H are *isomorphic* if there is a one-to-one correspondence ϕ from $V(G)$ to $V(H)$ such that $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$. A *graph invariant* is a property that has the same value for every pair of isomorphic graphs. An *automorphism* is an isomorphism from a graph G to itself. Since the identity and inverse are both automorphisms and the composition of two automorphisms is itself an automorphism, the set of all automorphisms of a graph G forms a *group* under the operation of composition. We denote the *automorphism group* of G to be $Aut(G)$. Suppose v is a vertex of graph G. The set of all vertices to which v may be mapped by an automorphism of G in an *orbit* of G. The automorphism relating two vertices is an equivalence relation resulting in equivalence classes that are the *orbits* of G. We define the orbit of vertex i in graph G as $o(i_G)$. If G contains a single orbit, then G is *vertex-transitive*.

The *adjacency matrix* A is comprised of entries $a_{ij}, \forall i, j \leq n$, where n is the number of vertices in the graph. An entry $a_{ij} = 1$ represents the adjacency of vertex i and vertex j; $a_{ij} = 0$ otherwise. A shortest path between two nonadjacent vertices is called a *geodesic*. The length of a geodesic between two vertices is the *distance* between the vertices. The *distance matrix* D is comprised of entries $d_{ij}, \forall i, j \leq n$, where n is the number of vertices in the graph and d_{ij} represents distance between vertex i and vertex j. The *average distance in* G is computed as the fraction of all pairwise distances out of all possible distances, $\dfrac{2}{n(n-1)} \sum_{i \neq j} d(v_i, v_j)$. By convention, $d(x, y) = 0$ if x and y are in different components. For a given vertex v in graph G, u is a *k-step (or k-hop) neighbor* of v if $d(u, v) \leq k$ for $1 \leq k \leq n - 1$. Wu and Dai introduced a specification of this measure in [17]: a vertex u is an *exact k-hop neighbor* of v if $d(u, v) = k$ for $1 \leq k \leq n - 1$.

Let v be a vertex of G. The *eccentricity* of the vertex v, $e(v)$, is the maximum distance from v to any other vertex in G: $e(v) = \max\{d(v, w) : w \in V(G)\}$. The *radius of* G is the minimum eccentricity among the vertices of G:

$\mathrm{rad}(G) = \min\{e(v) : v \in V(G)\}$. The *diameter of G* is the maximum eccentricity among the vertices of G: $\mathrm{diam}(G) = \max\{e(v) : v \in V(G)\}$. The *center of G*, $Cen(G)$, is the subgraph induced by those vertices of G having minimum eccentricity: $Cen(G) = G[\{v \in V(G) : e(v) = \mathrm{rad}(G)\}]$. The *periphery of G, $Per(G)$*, is the subgraph induced by those vertices of G having maximum eccentricity: $Per(G) = G[\{v \in V(G) : e(v) = \mathrm{diam}(G)\}]$. The kth *power* of an undirected graph G is the graph $G^k = (V(G), E(G^k))$, where $E(G^k) = \{uv : d_G(u,v) \leq k\}$.

A vertex v_i's *degree* denotes the number of vertices to which v_i is adjacent. The *degree sequence of G* of a graph is an integer sequence $d_1, d_2, \ldots, d_n$ where $n = |V(G)|$ and d_i is the degree of vertex i. We will use the convention that the sequence is non-increasing. The *degree distribution of G* is the probability distribution of the degrees of the nodes in the graph, i.e. what fraction of the vertices have degree k ($\delta(G) \leq k \leq \Delta(G)$), where $\delta(G)$ and $\Delta(G)$ are the *minimum and maximum degree* in G, respectively. The *average degree of G* is $\dfrac{\sum_{v_i \in V(G)} deg(v_i)}{n} = \dfrac{2m}{n}$, where m is the number of edges in the graph. The *density* of a graph is the ratio of possible edges to the edges that are actually present in the graph: $\dfrac{m}{\binom{n}{2}} = \dfrac{2m}{n(n-1)}$.

The *average clustering coefficient* is the ratio of triangles in the graph to the number of connected triples (i.e., connected subgraphs on three vertices):

$$\text{Average Clustering Coefficient} = \frac{\text{number of triangles x 3}}{\text{number of connected triples}}.$$

Pearson's correlation coefficient in graphs is a measure of *assortative mixing*, i.e., the extent vertices with high degree are adjacent to each other. See [16] for a detailed treatment of the topic. The *s-metric* determines the extent to which the graph being examined has a "hub-like" core; see [1, 11].

A variety of "centrality" measures serve to provide insight into which vertices are the most *influential* in a graph [15]. The word influential is emphasized because there is no commonly accepted definition in a graph topological context (and one is not offered here). We will leverage a basic understanding of "importance" in this paper: a vertex with high centrality is more "important" in some sense than a vertex of smaller centrality. The *degree* of vertex i is sometimes referred to as *degree centrality* of vertex i (degree centrality is often normalized through division by $n - 1$). *Closeness centrality* is a measure of the distance from a vertex to all other vertices, calculated as the inverse mean distance. *Betweenness centrality* is the extent to which a given vertex lies on the shortest paths to other vertices.

The reader may refer to [6, 7] for additional graph theory terminology and to [15] for additional complex network terminology.

We now introduce the following definitions as alternative (and more intuitive) terminology to the exact k-hop neighbor. The definitions will facilitate development of the neighbor matrix.

Definition 1. *Let G be a graph with $u, v \in V(G)$. Vertex u is a distance-k neighbor of v if and only if $d(u,v) = k$, where $k = diam(G) (1 \leq k \leq n-1)$.*

Definition 2. *Let G be a graph, with $v \in V(G)$. The distance-k neighborhood of v is*

$$N^{dist-k}(v) = \{u \in V(G) : d(u,v) = k\}.$$

3 The Neighbor Matrix of Graph G: $X^{dist-k}(G)$

We introduce a matrix that is a multi-dimensional analog of the degree sequence. As the degree sequence summarizes the adjacencies in the adjacency matrix, the neighbor matrix summarizes the distributions of each vertex-vertex geodesic length found in the distance matrix.

The neighbor matrix is a multi-dimensional analog of the degree sequence. As the degree sequence summarizes the adjacencies in the adjacency matrix, the neighbor matrix summarizes the distributions of each vertex-vertex geodesic length found in the distance matrix. See Definition 3 and Figure 2.

Definition 3. *Let G be a graph with $V(G) = \{v_1, v_2, \ldots, v_n\}$. The neighbor matrix,*

$$X^{dist-k}(G) = [x_{ij}], (1 \leq i \leq n, 1 \leq j \leq k),$$

where $n = |V(G)|$, $k = diam(G)$ $(1 \leq k \leq n-1)$ and $x_{ij} = |N^{dist-j}(i)|$, $i \in V(G); 1 \leq j \leq k$. We sort the rows of X^{dist-k} in a reverse lexicographic manner by organizing the rows of $X^{dist-k}(G)$ in non-increasing order of the entries in the first column. If there is a tie in some column $j < k$, we sort affected rows in non-increasing order of column $j + 1$.

The entries x_{ij} represent the count of dist-j neighbors of vertex i. This is the same number defined by Bloom, et al, and later by Buckley and Harary as the d_{ij}^{th} entry of the *distance degree sequence* [4, 5]. In addition, we define $X_i^{dist-k}(G)$ to be row i of the neighbor matrix associated with graph G, the row associated with $v_i \in V(G)$.

Figure 2 depicts a graph G and its corresponding neighbor matrix $X^{dist-3}(G)$.

Lemma 1. *There is a unique neighbor matrix $X^{dist-k}(G)$ for each graph G.*

Proof: Suppose otherwise that a given graph can have two different neighbor matrices. Let $K^1 = X_1^{dist-k}(G)$ and $K^2 = X_2^{dist-k}(G)$ such that $K^1 \neq K^2$. Then, there is an entry $k_{ij}^1 \neq k_{ij}^2$. By construction and ordering of the neighbor matrix, vertex i in G simultaneously has two different numbers of neighbors at distance j, which is not possible. ∎

Given Lemma 1, the following observation is immediate:

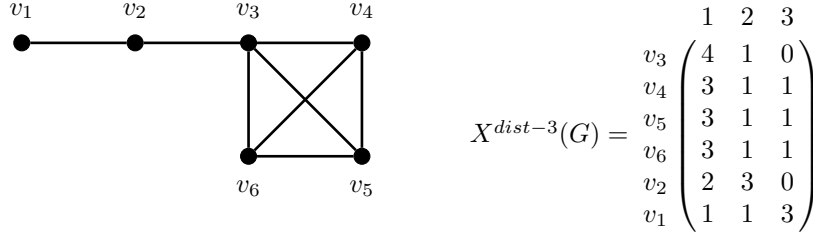**Proposition 1.** *Let G and H be two graphs. If $G \cong H$, then $X^{dist-k}(G) = X^{dist-k}(H)$.*

$$X^{dist-3}(G) = \begin{array}{c} \\ v_3 \\ v_4 \\ v_5 \\ v_6 \\ v_2 \\ v_1 \end{array} \begin{array}{ccc} 1 & 2 & 3 \\ \left(\begin{array}{ccc} 4 & 1 & 0 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \\ 3 & 1 & 1 \\ 2 & 3 & 0 \\ 1 & 1 & 3 \end{array}\right) \end{array}$$

Figure 2: Graph G with associated neighbor matrix

Proof: By definition of isomorphism, there is a one-to-one correspondence ϕ from $V(G)$ to $V(H)$ such that $uv \in E(G)$ if and only if $\phi(u)\phi(v) \in E(H)$. Therefore, the distance in G from vertex i to vertex j is the same as the distance in H from $\phi(i)$ to $\phi(j)$, $\quad \forall i,j \in G$. It follows directly by construction of the neighbor matrix that $X^{dist-k}(G) = X^{dist-k}(H)$. ■

Remark 1. *The converse of Theorem 1 is false. That is,*

$$X^{dist-k}(G) = X^{dist-k}(H) \;\not\Longrightarrow\; G \cong H.$$

Proof: We present an example depicting two infinite classes of non-isomorphic graphs with the same neighbor matrix below.

We define graphs G and H in accordance with Table 1. Vertices 1 to n define a K_n. In Figure 3, $n = 8$.

	G	H
vertex $n+1$ adjacent to	$n, n-1, 3, 4$	$n, 2, 3, 4$
vertex $n+2$ adjacent to	$1, 2, 3, 4$	$1, 2, 3, 4$
vertex $n+3$ adjacent to	$n, n-1, n-2, n-3$	$n, n-1, n-2, n-3$
vertex $n+4$ adjacent to	$1, 2, n-2, n-3$	$1, n-1, n-2, n-3$

Table 1: Non-isomorphic Graphs G and H with Identical Neighbor Matrices

Figure 3 is an illustration of the graphs constructed in Table 1. The edges drawn in heavier pitch are those edges which are different between graphs G and H.

Given G and H, $X^{dist-k}(G) = X^{dist-k}(H)$. In graph G, each degree-4 vertex has a maximum of two dist-1 neighbors shared with any other degree-4 vertex. In graph H, each degree-4 vertex has one vertex with which it shares three dist-1 neighbors. Therefore $G \not\cong H$. ■

It is clear that the implication stated in Proposition 1 holds for trees. Buckley and Harary provide a counterexample for Remark 1 which may be extended to show identical neighbor matrices do not imply isomorphism for trees [5]. As algorithms exist to construct neighbor matrices, the contrapositive of Theorem 1 is a method to verify the non-isomorphism of two graphs.
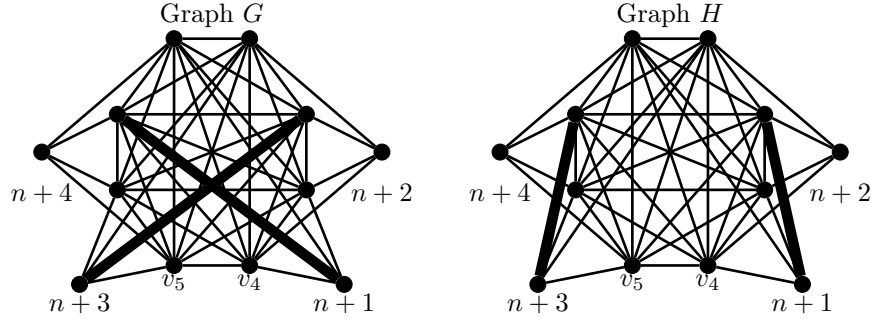
Figure 3: Graphs G and H

The beauty of the newly introduced neighbor matrix extends from how naturally it captures the topology of an arbitrary graph, coupled with the simplicity of calculating matrix entries. A neighbor matrix of a graph captures each vertex's "view" of the graph through shortest paths, thus reaching every vertex of the graph, implicitly accounting for edges, cycles and other subgraphs of G (see Figure 4, two views of the Petersen graph).
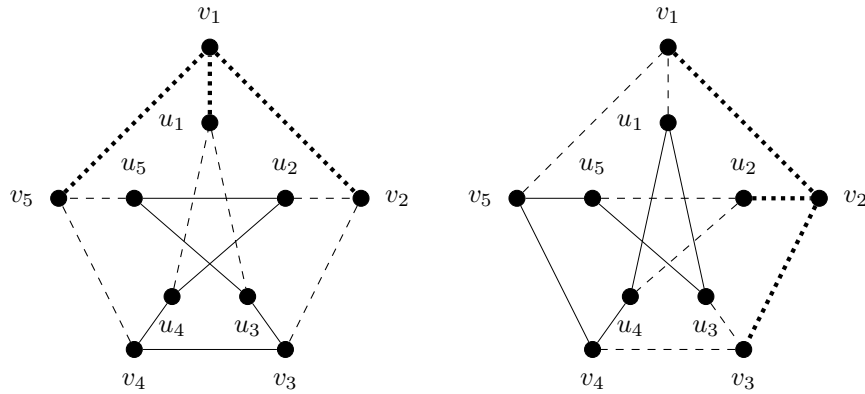

Figure 4: Two views of distance-1 and distance-2 neighbors (from v_1 and v_2) in the Petersen graph

From the topological information encoded in the neighbor matrix, we may extract many graph parameters, as shown in Theorem 1.

Theorem 1. *For a simple graph G the following topological information can be observed from the neighbor matrix $X^{dist-k}(G) = [x_{ij}], (1 \leq i \leq n, 1 \leq j \leq k)$:*

(a) G is connected if and only if $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} = n(n-1)$.

7

- *The number of components in G, $\kappa(G)$, is $\sum_{d=1}^{n-1} \frac{|\mathcal{S}_d|}{d+1}$, where*

$$\mathcal{S}_d = \{v_i : \sum_{j=1}^{k} v_{ij} = d\}.$$

(b) *The radius of the graph is $Rad(G) = \max\{j : x_{ij} \neq 0, \forall i, 1 \leq i \leq n\}$.*

(c) *The center of the graph is $Cen(G) = G[\{v_i : x_{i\,\mathrm{rad}(G)} \neq 0, x_{ij} = 0, \forall j > rad(G)\}]$.*

(d) *The closeness centrality of vertex i in a connected graph is*

$$CC_i = \frac{n-1}{\sum_{j=1}^{k} j \cdot x_{ij}}.$$

(e) *The average distance between vertices in a connected graph G is*

$$\frac{1}{n(n-1)} \sum_{i=1}^{n} \sum_{j=1}^{k} (j \cdot x_{ij}) = \frac{1}{n(n-1)} \sum_{j=1}^{k} \sum_{i=1}^{n} (j \cdot x_{ij}).$$

(f) *Given $\mathrm{diam}(G) = k$ (the number of columns in $X^{dist-k}(G)$), $(1 \leq k \leq n-1)$, the graph periphery is $Per(G) = G[\{v_i : x_{ik} \neq 0 \quad \forall i, \quad 1 \leq i \leq n\}]$.*

(g) *The first column is a representation of the degree sequence, in which the degree centrality of vertex i is the first entry in row i, x_{i1}. The number of edges in the graph is the half-sum of the entries in the first column:*

$$m = \frac{1}{2} \cdot \sum_{i=1}^{n} x_{i1}.$$

- *The density of a graph is the ratio of possible edges to the edges that are actually present in the graph:*

$$\frac{m}{\binom{n}{2}} = \frac{2m}{n(n-1)}.$$

(h) *For each column $j \in X^{dist-k}(G)$, $\sum_{i=1}^{n} x_{ij} = 2\left(|E(G^j)| - |E(G^{j-1})|\right)$.*

(i) *For each s $\left(1 \leq s \leq \mathrm{diam}(G)\right)$, the number of edges in the power graph G^s is given by*

$$\sum_{j=1}^{s} \sum_{i=1}^{n} x_{ij} = 2|E(G^s)|.$$

Proof:

(a) Suppose G is connected. Fix a vertex v, and note that $\forall u \in V(G - v)$, a geodesic exists such that $d(u, v) = j$ $(1 \le j \le k)$. There are $n - 1$ choices for u in $G - v$. By construction of $X^{dist-k}(G)$, the length of a geodesic between all vertex pairs $u, v \in V(G)$ is counted. So, for each of the n rows of the matrix there are $n - 1$ distance entries, totalling $n(n - 1)$. For the converse, assume, to the contrary, that G is not connected and $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} = n(n - 1)$. Then, there is a component G' containing q vertices $(1 \le q \le n - 1)$; we choose $v \notin V(G')$. There are at most $n - 1 - q$ vertices adjacent to v. By construction of $X^{dist-k}(G)$ the entries of the row corresponding to vertex v sum to $n - 1 - q$, and also there are at most $n - 2$ geodesics that are counted in the other rows of $X^{dist-k}(G)$. Therefore $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} \le (n - 1)(n - 2) + (n - 1 - q) < n(n - 1)$, which contradicts the initial statement that $\sum_{i=1}^{n} \sum_{j=1}^{k} x_{ij} = n(n - 1)$. ∎

- We proved in (a) that a connected graph (i.e., graph with 1 component) has n neighbor matrix row sums of $n - 1$. It follows that each connected component of order d will have d neighbor matrix row sums of $d - 1$. Therefore, given $\mathcal{S}_d = \{v_i : \sum_{j=1}^{k} v_{ij} = d\}$, $\forall d : 1 \le d \le n - 1$, the number of connected components in G of order d is
$$\kappa(G)_d = \frac{|\mathcal{S}_d|}{d + 1},$$
and, $\kappa(G) = \sum_{d=1}^{n-1} \kappa(G)_d.$ ∎

By construction of X^{dist-k}, each value x_{ij} is the number of geodesics originating at vertex i of length j $(1 \le j \le k)$ with different terminal vertices. We weight the entries of the k-matrix row by multiplying each value x_{ij} by its associated distance, j. As seen in (a), the number of possible geodesics in G in a simple, connected graph is $n(n - 1)$.

(d) After summing over row i, division by the number of geodesics provides the weighted average which is the average distance in G from vertex i to all other vertices. As stated in Section 2, the closeness centrality is the inverse of this average distance. ∎

(e) After summing over rows and columns, division by the number of geodesics provides the weighted average which is the average distance between vertices in G. ∎

9

(h) The left hand side of the equation double counts the sum of distance j neighbors ($1 \leq j \leq \text{diam}(G)$) from each vertex in G. The right hand side does the same by calculating the number of vertex pairs of distance at most j apart, and then it removes the number of vertex pairs of distance at most $j - 1$. ∎

(i) The left hand side of the equation double counts the sum of distance 1, 2, ..., j neighbors ($1 \leq j \leq \text{diam}(G)$) from each vertex in G. The right hand side does the same by calculating the number of vertex pairs of distance at most j apart. ∎

4   The Neighbor Matrix and Graph Orbits

As stated in Section 2, vertices in the same orbit of a graph form an equivalence class. These equivalence classes are related to neighbor matrix rows, as stated in Proposition 2.

Proposition 2. *Given vertices v_i and v_j in graph G we have*

$$o_G(i) = o_G(j) \implies X_i^{dist-k}(G) = X_j^{dist-k}(G).$$

Proof: Suppose otherwise, that v_i and v_j are in the same orbit and $X_i^{dist-k}(G)$ $\neq X_j^{dist-k}(G)$. Then, there is at least one distance value f for which v_i has more (or less) $distance - f$ neighbors than v_j. Therefore, there is no automorphism that maps v_i to v_j in $Aut(G)$. This is a contradiction, as v_i and v_j are in the same orbit, such a permutation of vertices must indeed exist in $Aut(G)$. ∎

Remark 2. *The converse of Proposition 2 is false. That is,*

$$X_i^{dist-k}(G) = X_j^{dist-k}(G) \; \not\!\!\implies o_G(i) = o_G(j).$$

Proof: The graph G in Figure 5 shows an example of a graph with a cycle. Note that vertices v_i and v_j have identical neighbor matrix rows. There are two different paths from vertex v_i to vertex v_k, a distance-2 neighbor. All paths from vertex v_j to distance-2 neighbors are unique; therefore there is no automorphism in $Aut(G1)$ that maps v_i to v_j and the two vertices are in different orbits.

Consider graph T in Figure 5. Note that vertices v_y and v_z have identical neighbor matrix rows. The subgraphs rooted at vertices v_y and v_z, respectively each have two distance-2 neighbors. In the subgraph rooted at v_y, the paths to the two distance-2 neighbors share a common vertex. In the subgraph rooted at v_z, the paths to the two distance-2 neighbors do not share a common vertex. Therefore, there is no automorphism in $Aut(T)$ that maps v_y to v_z and the two vertices are in different orbits. ∎

An arbitrary graph (G) A tree (T)

$$X_i^{dist-6}(G1) = X_j^{dist-6}(G1) =$$
$$\begin{bmatrix} 3 & 3 & 1 & 0 & 0 \end{bmatrix}$$

$$X_y^{dist-6}(G2) = X_z^{dist-6}(G2) =$$
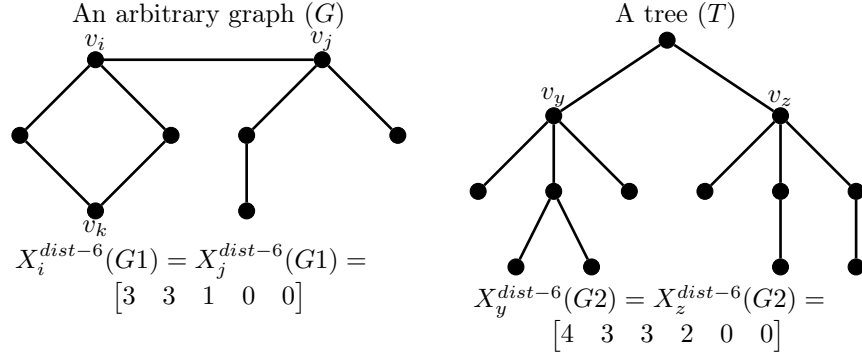$$\begin{bmatrix} 4 & 3 & 3 & 2 & 0 & 0 \end{bmatrix}$$

Figure 5: Graphs $G1$ and $G2$, Counterexamples to Proposition 2

5   Finding the Neighbor Matrix

The neighbor matrix may be fully determined using powers of adjacency matrices, or any of a variety of computer algorithms that determine all-pairs shortest paths. Propositions 3 and 4 present two algebraic methods of obtaining a neighbor matrix representation of G. For these propositions we use X_A^{dist-k} to refer to the neighbor matrix obtained from an adjacency matrix A of G, and $\vec{1}$ to represent the $n \times 1$ vector whose entries are 1.

Proposition 3. *Each neighbor matrix can be obtained using adjacency matrices of powers of the original graph G, as shown in equation 1. Entries of -1 in $A(G^i) - A(G^{i-1})$ are replaced by 0 before multiplication by $\vec{1}$:*

$$X_A^{dist-k} = \begin{bmatrix} A(G)\vec{1} & \left(A(G^2) - A(G)\right)\vec{1} & \cdots & \left(A(G^k) - A(G^{k-1})\right)\vec{1} \end{bmatrix}. \quad (1)$$

Proof: Notice that from the definition of G^k, the adjacency matrix of G^k, namely $A(G^k)$, includes entries of 1 for all pairs of vertices of distance k or less in the original graph G. Therefore, when we subtract $A(G^{k-1})$ from $A(G^k)$ and replace the -1 entries by 0, we are left with a $(0, 1)$-matrix with 1 entries representing pairs of vertices exactly distance k apart in G. Multiplication by $\vec{1}$ creates a vector consisting of the number of vertices at distance k from each vertex in G.   ∎

We let $\mathcal{A}^j$ be the Boolean matrix obtained from $\mathcal{A}^j = (A^j)^b - \mathcal{A}^{j-1}$, where $j \geq 2$, by replacing the value of -1 and diagonal entries with zero; $\mathcal{A}^1 = A$, the adjacency matrix.

Proposition 4. *Each unsorted neighbor matrix, $X^{dist-k(u)}$ can be obtained using the boolean matrices $\mathcal{A}^j (1 \leq j \leq k)$, as shown in equation 2*

$$X_{\mathcal{A}}^{dist-k(u)}(G) = \begin{bmatrix} \mathcal{A}^1\vec{1} & \mathcal{A}^2\vec{1} & \mathcal{A}^3\vec{1} & \dots & \mathcal{A}^k\vec{1} \end{bmatrix}. \qquad (2)$$

Proof: Notice that multiplying the adjacency matrix $\mathcal{A}^1$ of G by the column vector $\vec{1}$ we obtain the first column of the matrix $X^{dist-k}(G)$, the degree sequence. The second column of $X^{dist-k}(G)$ is $\mathcal{A}^2\vec{1} - \mathcal{A}\vec{1}$, as it counts all the vertices of distance 2 or less, and then it subtracts the vertices of distance 1, i.e. the neighbors of each fixed vertex in $V(G)$, as shown in Theorem 1. Similarly, $\mathcal{A}^k\vec{1} - \mathcal{A}^{k-1}\vec{1}$ counts the number of vertices k-hops away, and it subtracts the number of vertices $k-1$ hops away. $\blacksquare$

After a reverse lexicographic sort, $X^{dist-k(u)}$ becomes the neighbor matrix X^{dist-k}.

We next show that the neighbor-matrices X^{dist-k} obtained from different adjacency matrices of a graph G are related through the same permutation matrices as the adjacency matrices themselves.

Proposition 5. *Given A and B, two different adjacency matrices of graph G (so $B = P \cdot A \cdot P^T$, for some permutation matrix P),*

$$X_B^{dist-k}(G) = P \cdot X_A^{dist-k}(G) \cdot P^T.$$

Proof: Let A and B be two different adjacency matrices of G, so $B = P \cdot A \cdot P^T$ for some permutation matrix P. Since P is an orthogonal matrix, we have that $B^i = (PAP^T)^i = PA^iP^T$, for $1 \leq i \leq k$. Therefore $B^j - B^{j-1} = P(B^j - A^{j-1})P^T$ for each $j(1 \leq j \leq k)$, which implies that $X_B^{dist-k}(G) = P \cdot X_A^{dist-k}(G) \cdot P^T$. $\blacksquare$

6 Applications

There are many potential applications for the neighbor matrix, as it is a compact structure and rich in topological information. Two such applications are graph comparison and the identification of vertices with high topological significance, relative to other vertices in the graph.

6.1 Graph Comparison

The neighbor matrix enables the analysis of two graphs and insight into their topological *similarity*, which we consider to be a relaxation of isomorphism. We consider two graphs to be *similar* if they have the same distance distributions, i.e., their neighbor matrices are the same. For example, using this definition, graphs G and H in Figure 3 are *similar*.

It has previously been established that using individual characteristics provides an incomplete picture of a graph's structure [1, 11]. This is because information is missing that is needed to fully characterize the graph. Several measures follow that illustrate the problem of different structures with the same (or close) comparison criteria:

- Degree Sequence: a cycle C_{3k} and the disconnected graph obtained from the union of k copies of the C_3 cycle (or, kC_3) each have constant, identical degree sequences

- Average Distance: a wheel on 32 vertices $W_{1,31}$ and the 16-vertex star graph $K_{1,15}$ have the average path length of 1.875.

- Average Clustering Coefficient (close, positive): a wheel on 21 vertices $W_{1,20}$ has the average clustering coefficient of 0.63888, while the $(7,7)$ Barbell graph (two copies of K_7 connected by a path with 7 additional vertices) has average clustering coefficient of .63945.

- Average Clustering Coefficient (same): a graph of small diameter (star) and a large diameter (path) each have a clustering coefficient of 0.

- Pearson Coefficient: a wheel on $n+1$ vertices ($n \geq 5$) $W_{1,n}$ and the $(3,5)$ Barbell graph (two copies of K_3 connected by a path with 5 more vertices) have the Pearson correlation coefficient of $-.33333$.

- s-metric: for each $r-$regular graph, the normalized s-metric is 1 because there is no variability in the degree sequence in the background set of graphs.

Though variability of graphs with the same or similar graph invariant values has been observed before, it has not been adequately addressed. A widespread desire remains in the community of decision makers operating in a connected world to compare graph and network structures in a scientifically-viable manner. This, coupled with the well-documented lack of a comprehensive, intuitive metric useful in making this comparison served as a motivation for development of the neighbor matrix to describe graphs.

Consider Table 3, which includes several classes of graphs that are defined in Table 2. The entries in bold are the graph invariants previously discussed that have identical values when calculated from different graphs. Note that eight of the graphs represented in Table 3 have the same average clustering coefficient: 0; but neighbor matrix dimensions are different. This phenomenon persists through the other examples shown. A similar result comes from looking at the dimension of neighbor matrices associated with graphs manifesting the same average path length, Pearson correlation coefficent, and normalized s-metric. So, simple inspection seems to indicate utility in the neighbor matrix as a topological discriminator where other invariants fail, even before more detailed analysis.

In linear algebra, we use norms to determine the "size" of a matrix. As we see in Table 3, the Frobenius norm is unique for each graph present, even those for which multiple measures are the same or even undefined. This uniqueness provides the foundation for future work that uses the norm to calculate a graph comparison metric that is potentially more topologically insightful than measures currently in use.

Graph	Description	Graph	Description
$B_{7,7}$	2 x K_7 connected by P_7	$W_{1,20}$	wheel on 21 vertices
$K_{1,15}$	star on 16 vertices	P_{16}	path on 16 vertices
$W_{1,31}$	wheel on 32 vertices	$B_{10,12}$	2 x K_{10} connected by P_{12}
K_{32}	complete graph on 32 vertices	$K_{16,16}$	32-vertex complete bipartite
CL_{32}	circular ladder on 32 vertices	C_{32}	cycle on 32 vertices
H_{32}	hypercube on 32 vertices	$L_{21,11}$	K_{21} connected to P_{11}
P_{32}	path on 32 vertices	$K_{1,31}$	star on 32 vertices

Table 2: Graphs Analyzed

Graph	AVG Distance	AVG Cluster Coeff	Pearson Correl Coeff	s-metric (Norm'd)	dimension (X^{dist-k})	X^{dist-k} Frobenius Norm
$B_{7,7}$	5.0	**0.640**	0.719	0.981	21 x 10	43.36
$W_{1,20}$	1.81	**0.640**	**-0.333**	0.323	21 x 2	79.75
$K_{1,15}$	**1.875**	**0**	-1	0.133	16 x 2	56.39
P_{16}	5.330	**0**	-0.077	0.981	15 x 14	17.55
$W_{1,31}$	**1.875**	0.648	**-0.333**	0.207	32 x 2	159.8
$B_{10,12}$	7.323	0.613	0.866	0.990	32 x 15	77.05
K_{32}	1.0	1.0	*	**1**	32 x 1	175.4
$K_{16,16}$	1.484	**0**	*	**1**	32 x 2	124.1
CL_{32}	4.645	**0**	*	**1**	32 x 9	60.66
C_{32}	8.258	**0**	*	**1**	32 x 16	44.18
H_{32}	2.581	**0**	*	**1**	32 x 5	89.62
$L_{21,11}$	4.105	0.653	0.942	0.998	32 x 12	116.0
P_{32}	11.0	**0**	-0.033	0.992	32 x 31	38.37
$K_{1,31}$	1.94	**0**	-1	0.064	32 x 2	170.0

Table 3: Graph Comparison Metrics

*The Pearson correlation coefficient is undefined in regular graphs because the denominator of the calculation is a variance of zero.

6.2 Application 2: Identification of Topologically Influential Vertices

Analysis using the topological information stored in the newly-developed neighbor matrix can rank the vertices in a graph based upon the structural importance of that vertex. If the removal of vertex v_i results in a relatively large change to the distance distributions in a graph, we may assert v_i is structurally important to the graph. Relatively small change to the distances in a graph upon vertex removal indicates less structural importance. There are other measures of topological importance currently in use—the centralities described in Section 2; however, these measures do not directly address how vertices impact the distance between vertices in a graph.

This information can be leveraged towards recommendations for network attack and defense. In the attack, the attacker would choose to interdict the vertex or vertices that cause the greatest structural change. On the defense, the defender would choose to protect or "harden" those nodes that cause greatest structural change in the network.

7 Conclusion

This work introduced and explored the neighbor matrix as an algebraic structure that contains significant graph descriptive and topological information. We proved this topological richness by proving the presence of 11 graph invariants in the neighbor matrix and relating the neighbor matrix to graph orbits. Though it does not inform isomorphism in arbitrary graphs, the neighbor matrix does provide a technique to verify the non-ismorphism of two graphs. The neighbor matrix has the potential to enable greater understanding of "graph space," as it is simultaneously more compact and richer in information than current structures used in graph exploration. For example, the degree sequence has long been used as a foundational element for the exploration of families of graphs. Since the neighbor matrix extends the degree sequence through all the distances that comprise the graph, it promises to enhance our current capability to model, analyze, and understand graphs in all distance dimensions. Further exploration is warranted into how we may use the neighbor matrix to further refine models that use the degree sequence as an input, such as the configuration model of graph generation. We proved the neighbor matrix to contain eleven of the most commonly used invariants used to describe graphs. Future work will leverage this topological richness toward generating insights into graph comparison and further, toward informing the identification of vertices whose influence on graph topology is significant. The neighbor matrix is a mathematically-manipulable structure that may move us toward a single characterization of certain classes of graphs, such as is found in the idea of a "graph signature" [2, 12]. Finally, as the neighbor matrix presents the sequences of all distances which manifest in a graph, it may provide a tool to verify the unigraphicality of certain classes of graphs.

8 Acknowledgements

References

[1] Alderson, D. L. and Li, L. (2007). Diversity Of Graphs With Highly Variable Connectivity. *Phys. Rev. E (3)*, 75(4):046102.

[2] Asadpour, M., Sproewitz, A., Billard, A., Dillenbourg, P., and Ijspeert, A. J. (2006). Graph Signature for Self-Reconfiguration Planning. *IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 863–869.

[3] Axelrod, R. (2015). *Structure Of Decision: The Cognitive Maps Of Political Elites*. Princeton University Press.

[4] Bloom, G. S., Kennedy, J. W., and Quintas, L. V. (1983). Some Problems Concerning Distance And Path Degree Sequences. In *Graph Theory*, pages 179–190. Springer.

[5] Buckley, F. and Harary, F. (1990). *Distance In Graphs*. Addison-Wesley Longman.

[6] Chartrand, G. (2006). *Introduction To Graph Theory*. Tate McGraw-Hill Education.

[7] Harary, F. (1969). *Graph Theory*. Addison-Wesley.

[8] Ivanciuc, O., Balaban, T.-S., and Balaban, A. T. (1993). Design of Topological Indices. Part 4. Reciprocal Distance Matrix, Related Local Vertex Invariants And Topological Indices. *J. Math. Chem.*, 12(1):309–318.

[9] Lele, S. (1991). Some Comments on Coordinate-Free and Scale-Invariant Methods in Morphometrics. *American Journal of Physical Anthropology*, 85(1):407–417.

[10] Lele, S. (1993). Euclidean Distance Matrix Analysis (EDMA): Estimation of Mean Form and Mean Form Difference. *Mathematical Geology*, 25(5):573–602.

[11] Li, L., Alderson, D., Doyle, J. C., and Willinger, W. (2005). Towards a Theory of Scale-Free Graphs: Definition, Properties, and Implications. *Internet Math.*, 2(4):431–523.

[12] MacDougall, J. A., Eggleton, R. B., and Adams, P. (2006). Graphs With A Given Degree Sequence. *Congr. Numer.*, 179:15–31.

[13] Merris, R. (1994). Laplacian Matrices Of Graphs: A Survey. *Linear Algebra Appl.*, 197:143–176.

[14] Mihalić, Z., Veljan, D., Amić, D., Nikolić, S., Plavšić, D., and Trinajstić, N. (1992). The Distance Matrix In Chemistry. *J. Math. Chem.*, 11(1):223–258.

[15] Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.

[16] Newman, M. E. (2002). Assortative Mixing In Networks. *Phys. Rev. Lett.*, 89(20):208701.

[17] Wu, J. and Dai, F. (2005). Efficient Broadcasting with Guaranteed Coverage in Mobile Ad Hoc Networks. *IEEE Transactions On Mobile Computing*, 4(3):1.