Article

Not peer-reviewed version

Collaboration: People, Papers, Average Graphs, Durfee Squares and Metric Dimension

Melissa Holly *

Posted Date: 7 July 2025

doi: 10.20944/preprints202507.0495.v1

Keywords: collaboration network; network evolution; metric dimension; Durfee square; Durfee rank; average graph

doi:10.20944/preprints202507.0495.v1

Article

Collaboration: People, Papers, Average Graphs, Durfee Squares and Metric Dimension

Melissa Holly

Independent researcher, Richmond, VA, USA; mholly1417@gmail.com

Abstract

Utilizing several methods, this note shows that, in any collaboration network analysis, paper exclusion not only creates a loss of information, but can lead to incorrect interpretation of network structure because interpretation of vertex degree in the authors only graph is not well defined. Because the bipartite authors with papers graph is the actual social network, metric dimension is used to show that the relative distance structure of the bipartite graph is often defined by the structure of the papers, not that of the authors. Due to the NP-hard nature of metric dimension, methods that increase computational efficiency for the bipartite authors and papers graph are explored. With a departmental collaboration focus, public data for 245 professors from mathematics, physics and biology departments of three U.S. public universities is analyzed with network structure compared using metric dimension. By discipline, an average graph is defined with average graphs constructed from the collected data for the authors only structure, for the bipartite authors with papers structure and for papers only. Social analysis of the collected data shows that a 27% change in the total number of hubs, along with identifying different professors as hubs, when the authors only graphs are compared to bipartite graphs reiterating the need for paper inclusion in any collaboration study.

Keywords: collaboration network; network evolution; metric dimension; Durfee square; Durfee rank; average graph

MSC: Primary 05C82; Secondary 05C12

---

1. Introduction

Most collaboration network studies examine the authors only structure although it is recognized that some information is lost when papers are excluded. This study shows that excluding papers can lead to misinterpretation of various aspects of the collaboration network structure, possibly making accurate modeling of the network evolution impossible.

Utilizing graphs as sociograms, with a focus on the departmental collaboration of 245 professors, data was collected for three STEM departments in three U.S. public universities where the average department size is similar to Zachary's karate club mentioned in several social network papers [25]. To provide anonymity concerning the selected universities, the concept of an average graph is introduced with data analysis given for both graphs based on the collected data and the constructed average graphs. Used to compare the change in relative distance structure when papers are added to the authors only structure, metric dimension analysis reflects that the authors only structure does not necessarily preserve the relative distance structure of the actual collaboration network. For any bipartite graph G and utilizing G's distance matrix (DM), DM block resolving provides an accurate but more efficient method for finding the metric dimension of G, $\dim(G)$. Denote the bipartite authors and papers graph with G_{AP}, the authors only graph by G_A and the papers only graph as G_P. Let $a \times a$ be the DM block with both rows and columns indexed by author vertices $a \in A$ and $p \times p$ be the square block indexed by paper vertices $p \in P$. Proposition 9.1 shows that DM block resolving can be utilized to find $\dim(G_A)$ and $\dim(G_P)$. Theorem 9.7 states that given $\dim(a \times a) \neq \dim(p \times p)$ only three DM blocks need to

doi:10.20944/preprints202507.0495.v1

be resolved in order to find $\mathsf{dim}(G_{AP})$. Propositions 9.8 and 9.9 provide more efficient methods for determining $\mathsf{dim}(G_{AP})$ given certain conditions such as Durfee rank of a graph comparison. This note appears to be the only social network analysis utilizing metric dimension to compare related social network structures, and the only one utilizing degree diagrams and the related Durfee rank of a graph.

In this study's nine departments, an average of 50% of the faculty performed departmental collaboration between 2019 and 2023. Six of the nine department chairs did departmental collaboration with three of the six acting as collaboration hubs. In comparing the G_A to the bipartite G_{AP}, the G_{AP} represents the actual social network, not the G_A. Although there is an interpretation difference in large vertex degree in a G_A compared to that of a G_{AP}, based on collaboration hubs, a comparison of the G_A to the G_{AP} results in a 27% increase in the number of hubs, plus different authors are identified as the hubs. The large degree interpretation of the G_A is also not well defined. This emphasizes the importance of paper inclusion and gives foundation for Conjecture 6.1 stating that accurately determining social network evolution models requires paper inclusion.

Although the use of graphs in studying networks preceded their paper, Harary and Norman's 1953 article that connected graph theory to social network analysis [13] got serious attention. Price's 1965 article [29] discusses the network structure of scientific papers based on the papers' references. Another important social network study was conducted by Milgram in 1967 [23] where randomly selected individuals were asked to forward a letter to a stock broker in Boston. In 2001 Newman [24] utilized large scientific paper databases to study the collaboration structure of scientific research. Often used to compare networks, metric dimension was published independently by Slater in 1975 [32] and by Harary and Melter in 1976 [14]. Metric dimension was originally shown to be NP-complete in [6] but has been more recently shown to be NP-hard [15]. The metric dimension of complete bipartite graphs is given in [8] while [4] discusses $\mathsf{dim}(G)$ for regular bipartite graphs. No exact method has been found for finding the metric dimension covering all bipartite graphs.

Because this note is written for a variety of possible readers, explanations are kept as simple as possible although basic graph theory knowledge is assumed. Section 2 provides graph theory notation and other possibly unfamiliar concepts important to this note. That section concludes with social network concepts and some results of other social network studies. Section 3 gives the data collection methods for this study. Vertex projection graphs based on the bipartite authors with papers graph are discussed in Section 4, while Section 5 discusses the structural specifics of the bipartite authors and papers graph. The social analysis in Section 6 interprets the data collected on departmental collaboration with respect to department chairs who act as hubs.

Since this study takes a close view of mathematics, physics and biology departments, the data analysis presentation focuses on concealment of which universities were used via average graph construction. The nine average graphs given in Section 7 are accompanied by brief discussions. The use of the degree diagram and the Durfee square with its rank as an analysis tool for bipartite graphs is discussed in Section 8. Section 9 covers the metric dimension analysis of the average graphs and methods for determining the metric dimension for the authors and papers bipartite graph. A focus on future work concludes this paper in Section 10 where the challenges presented by paper inclusion in collaboration studies utilizing large databases is briefly discussed.

2. Background Information

Whether this note's readers are sociologists, statistical physicists or mathematicians, this note assumes basic graph theory familiarity as found in [9]. Utilized in many social analysis studies, a sociogram is a graph where vertices reflect people and/or social groups with edges representing vertex relationships. Graphs constructed from the collected data are referred to here as *data derived graphs*.

The *Pigeonhole Principle* states that given more pigeons than pigeonholes for the pigeons, and assuming that all of the pigeons find a pigeonhole, then at least one pigeonhole contains more than one pigeon.

2.1. Graph Theory Notation

A proper subset A of set X is denoted by $A \subset X$ with set cardinality given by $|X|$ and element x inclusion in X by $x \in X$. Graph G has vertex set $V(G)$ with cardinality $|V(G)|$, or simply n, called G's *order*. Edge set $E(G)$ has cardinality $|E(G)|$, or m, and is G's *size*. A vertex v in a specific graph G is v_G. All G in this note are simple so multiple edges and loops are excluded. Graph G isomorphic to graph H is denoted with $G \cong H$. The *diameter* of a connected G, $\mathrm{diam}(G)$, is the longest distance found over all of G's vertex pairs. *Distance variety* is the set of *possible* distances as determined by connected G's diameter. For instance, suppose G's diameter is 4. If all distances between specific $v \in V(G)$ and all other $v_i \in V(G)$ are determined, this set of distances can only include 0 (from v to itself), 1, 2, 3 and 4 since the diameter is the longest path in G. All of these distances do not necessarily exist for all $v \in V(G)$; instead, these distances are the only *possible* distances for G. For vertices v_1 and v_2, let $d(v_1, v_2)$ denote the distance between v_1 and v_2.

Complete graphs are given by K_n where n is the graph order. A path graph is P_n and a cycle graph is C_n. Distinctly label the two vertex sets of bipartite G so that each set is clearly distinguished from the other, and call this labeling a *distinguishing labeling*. The goal of this labeling is to generate block matrices for the bipartite graph where possible. As an example, during this study, the exact number of authors who did departmental collaboration and the exact number of representative papers was initially unknown. However, for each department, it was a given that the number of authors was < 100. Thus, when department graphs were constructed, the distinguishing labeling gave authors a label < 100 and papers > 100.

The open neighborhood of vertex v is $N(v)$ while the closed neighborhood is $N[v]$. For $v_1, v_2 \in V(G)$, v_1 adjacent to v_2 is $v_1 \sim v_2$. Given v_1 and v_2 and the edge (v_1, v_2) between them, if vertex w is added between v_1 and v_2 creating edges (v_1, w) and (w, v_2), then edge (v_1, v_2) is *subdivided* by w. In comparing graphs G_1 and G_2, if G_2 is G_1 with every edge subdivided, then G_2 is said to be a *subdivided* G_1.

Vertex v with degree of 1, $\deg(v) = 1$, is called a *pendant vertex*. Define *pendant chain* in a bipartite graph as a path subgraph of three or more vertices in G that begins with a pendant vertex and concludes with a vertex incident to at least 3 edges in G with only degree 2 vertices between the pendant vertex and the concluding vertex. The length of the pendant chain is the number of edges between the pendant vertex and the vertex with degree greater than 2. A P_n is considered to have no pendant chains. The maximum degree of a vertex in $V(G)$ is denoted by $\Delta(G)$ and the maximum degree of set X is $\Delta(X)$. The minimum degree in G is $\delta(G)$.

2.2. Distance Matrix and Common Neighbor Matrix

For any G with order n where $v_i, v_j \in V(G)$, the *distance matrix* DM is the symmetric $n \times n$ matrix indexed by $V(G)$. Element $d_{i,j}$ of DM contains the distance between v_i and v_j in G. A DM is only defined for connected graphs or for each connected component of a graph. Each $d_{i,i}$ is zero reflecting the distance of each vertex to itself. In bipartite graphs, distances between elements of the same partite vertex set are all even while distances between elements of the two distinct partite sets are all odd. Given a distinguishing labeling on bipartite G_{AP} with partite sets A and P, G's DM has blocks $a \times a$, $p \times p$, $a \times p$ and $p \times a$ where $a \in A$ and $p \in P$. Blocks $a \times a$ and $p \times p$ contain even distances while the distances in the $a \times p$ and $p \times a$ blocks are odd.

Given G, define the *common neighbor matrix* CNM as a symmetric $n \times n$ matrix with indices defined on $V(G)$ where element $c_{i,j} \in$ CNM at the intersection of $v_i, v_j \in V(G)$ is the number of common neighbors that v_i shares with v_j. A CNM can be defined to include the common neighbors of vertices that are adjacent and/or those that are not adjacent. In this note, it is assumed that any CNM here is defined between nonadjacent vertices. Thus, for nonadjacent $v_i, v_j \in V(G)$, $c_{i,j} = |N(v_i) \cap N(v_j)|$. The CNM based on nonadjacent vertices for any K_n is an all zero matrix. For nonadjacent vertices, if G is bipartite and $c_{i,j}$ is non-zero, then v_i and v_j are in the same partite set. Note that any row (or column) sum can include duplicate vertices since a vertex can be shared between vertices v_i and v_j and also shared between v_i and vertex v_k.

In Section 4, the CNM of the bipartite authors with papers graph is used to construct projection graphs. Based on common neighbors of nonadjacent vertices, a similar matrix is the *κ-th order vertex-adjacency matrix* ${}^{v}\mathrm{A}_{\kappa}$ as given in [17] for $\kappa = 2$. This matrix is a binary matrix with a 1 when v_i and v_j have distance 2 and a zero otherwise. However, the CNM provides the number of common neighbors for nonadjacent vertices, not just that common neighbors exist. Although not within the scope of this note, for two graphs with the same order $n > 6$, consider what a CNM based on nonadjacent vertices of all 1 and zero elements states about G_1 compared to a CNM for G_2 with only 0, 3, 4 and 5 as elements.

2.3. Average Graphs

The motivation for developing an average graph derives from the desire for both anonymity in this study yet using graph/network structure images. An *average graph* is a graph defined on the statistical averages of a parameter set for a collection of graphs. This concept has meaning when the graphs in the collection of graphs are specifically related. Although the departments within each university in this study are related to each other by institution, the average graphs in Section 7 represent the averages of the data derived graphs for each discipline as this aligns with this study's objective.

The three data derived physics graphs (authors only, authors with papers and papers only graphs) differ significantly from those for mathematics and biology; but the three data derived physics graphs have similar structure to each other. The same can be said for the three data derived mathematics graphs and the three biology graphs.

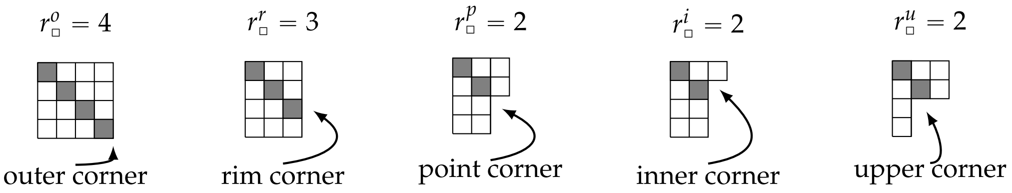
In constructing this study's average graphs, average graph order and size are used, along with average diameter, average degree and, as discussed later in this note, average metric dimension. Because many averages generate decimal values, truncation value to rounded value ranges are used as targets in average graph construction. If an integer value results, then that value is used. Average degree is determined as the sum of the three graphs' degree sequences divided by the sum of the graph orders $\pm$ one standard deviation ($\pm\sigma$) with the target range based on the truncation value to the rounded value range as with the other averages. For bipartite authors with papers graphs, the average number of papers and the average number of authors is calculated. The average number of components is determined along with the average number for each component type such as the average number of P_3 components, etc. Regarding the larger components that are distinct, an average large component(s) is found that is average with respect to order, size, degree, diameter and metric dimension.

2.4. Durfee Square

The concept of Young diagrams, and their included Durfee squares, are used in [20] and [21] with respect to the degree sequence of a graph. In [21] and [31], these two concepts are utilized in connection to threshold graphs. The Durfee square in [1] is used for h-index enhancement, while in [30], the Durfee square is utilized in measurement of scholarly impact.

A *Young diagram* (also called *Young tableau* and related to *Ferrer's diagram*) gives a visual image of a non-negative non-increasing integer partition. Imagine the number 4 as four horizontal squares reflecting 4+0. Then 3+1 can be visualized as  with a maximal partition integer as the top row and each subsequent row below the top row representing a partition integer less than or equal to its predecessor row partition integer.

The *Durfee square* of a Young diagram is the largest square of squares anchored by the upper left square of the diagram. The *Durfee rank*, $r_{\square}$, is the number of squares along one edge of the Durfee square, which is also the number of squares along the Durfee square diagonal ($r_{\square}$ is also called the *Frobenius rank* [10] or *partition trace* [20]). In [2], a Young diagram corner ⌜ is referred to as an *inner* corner of the Young diagram while a ⌟ corner is called an *outer* corner. Figure 1 displays five types of corners associated with $r_{\square}$, each of which relays different information regarding the large degree structure of a graph. Let $R_{r_{\square}}$ indicate the lowest (from the top) degree diagram row that includes $r_{\square}$, and let $\ell(R_{r_\square})$ denote the length of this row. The row above $R_{r_\square}$ is row $R_{r_\square-1}$ and the row below $R_{r_\square}$ is row $R_{r_\square+1}$. The table at the bottom of Figure 1 reflects the relationship of each $r_\square$ corner type with respect to the degree diagram rows $R_{r_\square-1}$, $R_{r_\square}$ and $R_{r_\square+1}$.



$r_\square$ corner type	$\ell(R_{r_\square-1})$ above	$\ell(R_{r_\square})$	$\ell(R_{r_\square+1})$ below
Outer: $r^o_\square$	$= r_\square$	$= r_\square$	$< r_\square$
Rim: $r^r_\square$	$= r_\square$	$= r_\square$	$= r_\square$
Point: $r^p_\square$	$> r_\square$	$> r_\square$	$= r_\square - 1$
Inner: $r^i_\square$	$> r_\square$	$= r_\square$	$= r_\square$
Upper: $r^u_\square$	$> r_\square$	$> r_\square$	$< r_\square$

Figure 1. Five types of $r_\square$ corners and a table with respect to degree diagram row length.

Any graph with at least one edge, has a degree sequence that is integral. An integer sequence from which a graph can be constructed is called *graphic*. In the same manner that 4 can be partitioned as either 2+2 or 3+1, a degree sequence can be viewed as a partition of twice a graph's size or $2m$. A *degree diagram* is a Young diagram of a graph's' degree sequence, or a degree sequence of a vertex subset that might not be graphic. Certain requirements exist for an integer sequence to be graphic (see [9,21] for general information). There exist two easy indicators that a degree sequence is graphic. First, the number of odd degree vertices must be even. Also, for a degree sequence X, $\Delta(X) < |X|$ is required.

As explained in the next subsection, the larger degrees in G impact G's metric dimension. For G, $r_\square(G)$ reflects that there exist at least $r_\square(G)$ number of vertices that have degree at least $r_\square(G)$; thus giving a minimum value for the large degree structure in G. The identification of the $r_\square$ corner conveys additional large degree information as shown by the table in Figure 1.

Isolated vertices K_1 are excluded from this note; and they are excluded from the degree diagram concept. Given a vertex subset X of a connected G, $r_\square(X)$ is the Durfee rank of X's degree sequence. Although the degree diagram of a vertex subset X can have a single row, any degree diagram of G must have at least two rows.

Consider a bipartite G_{XY} with partite sets X and Y. The degree sum of either vertex set is $m = |E(G)|$; thus, the degree sequences of X and Y are each partitions of m. The degree sequence of either set is most likely not graphic, but placing each sequence in a degree diagram provides significant information regarding G. If the partite degree sequences are placed in degree diagrams, then the visual partition image of one diagram is simply a rearrangement of the squares in the other partition image. Discussed later in this note, Figures 10 and 11 each display a bipartite G_{AP} with the degree diagrams for the two vertex sets A and P.

Assume a connected G. If $r_\square(G) = 1$ then the second row of the given degree diagram must contain a single square; and the first row must contain one or more squares. If the first row contains a single square, then G must be bipartite K_2 which is a star graph, $K_{1,n-1}$, with $n = 2$. If the first row contains more than one square, then G is again a star graph. Thus, if and only if $G \cong K_{1,n-1}$, then $r_\square(G) = 1$. This proves Proposition 2.1.

Proposition 2.1. *A connected G has $r_\square(G) = 1$ if and only if G is a star graph, $K_{1,n-1}$, with $n \geqslant 2$.* $\square$

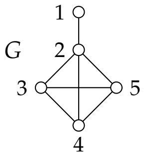
Although complete characterization of simple G based on $r_{\square}(G)$ is beyond the scope of this note, if $r_{\square}(G) = 2$, then $\Delta(G) \geqslant 2$ for two or more vertices. Graphs with $r_{\square}(G) = 2$ include C_n and, for $n \geqslant 4$, P_n. There is more to explore in the Durfee square and Durefee rank concepts than what is contained in this note.

2.5. Metric Dimension

Introduced independently by Slater in [32] (1975) and Harary and Melter in [14] (1976), the metric dimension of a graph has found numerous uses related to the comparison of network structures. If two graphs have the same metric dimension then, based on relative distance, the two graphs have a similar structure.

Suppose G is a simple graph with $v_i, v_j \in V(G)$, and let $d(v_i, v_j)$ indicate the distance between vertices v_i and v_j. Imagine ordered subset W of vertices in G such that every vertex in G has a unique combination of distances to the members of W. Then W is called a *resolving set* of G. The fact that the elements in W are considered to be ordered is critical. The cardinality of a minimum resolving set is the *metric dimension* of G, $\dim(G)$. Set W is the *metric generator* of G, and the elements of a minimum cardinality W is a *metric basis* of G [8]. There can be more than one resolving set W with minimum cardinality.

As an example, consider the graph given in Figure 2 along with two W sets, W_1 and W_2. Each vertex v has a unique *distance vector*, $r(v|W_i)$ (also known as the *metric representation* of v or *metric code* of v), that contains the distances from v to each of the members in that particular W. W is a resolving set if and only if all distance vectors for all v in G are unique. Any vertex v in W has a unique $r(v|W)$ as a zero is in v's distance vector at v's position in W; so both W_1 and W_2 in the figure are resolving sets of G. If vertex 2 is added to each W, W is still a resolving set but it is not minimal. However, if any vertex is removed from either W_1 or W_2 then the distance vectors for the vertices not in the altered W are no longer unique, indicating that both W_1 and W_2 are minimal resolving sets for G; and G has $\dim(G) = |W_i| = 3$ where $i \in \{1, 2\}$. Additional information on the metric dimension of a graph can be found in [8].



$W_1 = \{1,3,5\}$
$r_1(2|W_1) = (1,1,1)$
$r_1(4|W_1) = (2,1,1)$

$W_2 = \{1,3,4\}$
$r_2(2|W_2) = (1,1,1)$
$r_2(5|W_2) = (2,1,1)$

$$\begin{array}{c|ccccc|}
 & 1 & 2 & 3 & 4 & 5 \\
\hline
1 & 0 & 1 & 2 & 2 & 2 \\
2 & 1 & 0 & 1 & 1 & 1 \\
3 & 2 & 1 & 0 & 1 & 1 \\
4 & 2 & 1 & 1 & 0 & 1 \\
5 & 2 & 1 & 1 & 1 & 0
\end{array}$$

Distance matrix DM for G

Figure 2. Graph G and two resolving sets for G plus the distance matrix DM for G.

2.5.1. Known Metric Dimensions

The metric dimension of some graph families has been determined. All P_n have metric dimension of 1, $\dim(K_n) = n - 1$ for all K_n and $\dim(C_n) = 2$. Bipartite star graphs have $\dim(K_{1,n-1}) = n - 2$. There exist additional graph families with known $\dim(G)$ not mentioned here, and finding the metric dimension remains an active area of research. From Proposition 2.1, if $r_{\square}(G) = 1$, then $G \cong K_{1,n-1}$ and $\dim(G) = n - 2$.

2.5.2. Distance Matrix and Metric Dimension

All G have $\dim(G)$. From a technical linear algebra standpoint, matrix columns represent vector space bases while rows map to the field. With respect to any DM, because a DM is symmetric, the metric dimension of G can be found utilizing either the columns or the rows of G's DM [3]. Rows are utilized in this note because this seems more natural than using columns. The manner in which the DM is used to find the metric dimension of a graph is best explained with an example. Figure 2 contains the distance matrix DM for the displayed G. The goal is to find a minimum number of rows that provide a set of unique ordered combinations of distances in the selected rows' column combinations. Any specific v_i indexing a DM row is assumed to be in a W set and the ordered column combinations of a collection of v_i are the distance vectors for the $v_i \in W$ to $v_j \in V(G)$. Alternatively, if columns are used, the vertices v_j indexing the columns are in W and the ordered row combinations are the distance vectors for the v_j to v_i. Note that any column combination in row v_i that contains a zero is unique as this combination indicates that $v_i \in W$.

In Figure 2, first notice that the DM row for pendant vertex 1 has a unique number of 1s and 2s compared to the other rows. A unique set of distances is often true for pendant vertices. Second, notice that the row for vertex 2 contains all 1s except for the single zero; so getting unique column combinations with vertex 2's row is difficult. Let $W = \{1,2\}$ and consider the DM rows for vertices 1 and 2. For this W, the column combinations for vertices 3, 4 and 5 are $r(3|W) = r(4|W) = r(5|W) = (2,1)$ so W is not a resolving set. Now select the DM rows of 1 and 3 placing 1 and 3 in W. Compared to rows 1 and 2, rows 1 and 3 have one less repeated column combination. However, still $r(4|W) = r(5|W) = (2,1)$. So let either vertex 4 or 5 be in W making $|W| = 3$. Rows 1, 3 and 5 provide the set of unique column combinations that are the unique distance vectors $\{(0,2,2),(1,1,1),(2,0,1),(2,1,1),(2,1,0)\}$, so $W = \{1,3,5\}$ is a resolving set. Comparing all combinations of rows, all minimal resolving sets contain three elements so $\dim(G) = 3$. Note that the closed neighborhoods of vertices 3, 4 and 5 in Figure 2 are the same.

2.5.3. Diameter and Metric Dimension

Since the diameter is the longest possible distance between any two vertices in G, the diameter reflects the maximum distance variety over the vertices of G. Consider G with diameter of 2 so $n > 2$. Then, excluding the zero in each DM row and utilizing $\mathrm{diam}(G)$, the only possible distances contained in G's DM are 1 and 2. As the order of G increases, so does the length of the rows in DM. Excluding zero, let n be the number of distances in G's distance variety and let r be the number of DM rows; so r is also the number of zeros in the r rows. There exist a maximum of $n^r + r$ possible unique ordered combinations. In the example with $\mathrm{diam}(G) = 2$, for distances 1 and 2 and for 2 rows, there exist 6 maximum unique column combinations. Thus, two rows cannot give unique column combinations required by $\dim(G)$ if the rows are longer than 6.

Given a fixed graph order, as diameter decreases, the metric dimension tends to increase. For simple connected G, if G's diameter is $n-1$ then G is P_n and $\dim(G) = 1$. If G's diameter is 1, then $\dim(G) = \dim(K_n) = n-1$.

2.5.4. Degree and Metric Dimension

Vertex degree in G also plays a significant role in $\dim(G)$ because as general degree increases for a fixed order G, diameter tends to decrease due to more vertices becoming adjacent to each other. As diameter decreases, the variety of distances in G's DM tends to decrease since the number of 1s increases in some rows. As the distance variety decreases, then the number of DM rows required for unique column combinations tends to increase. In other words, for a fixed graph order, a general increase in vertex degree also tends to increase metric dimension.

2.5.5. Twin Vertices and Metric Dimension

Given distinct vertices v_1 and v_2 in G, if either $N(v_1) = N(v_2)$ or $N[v_1] = N[v_2]$, then v_1 and v_2 are *twin vertices* ([19] [37]); and distances from v_1 and v_2 to the other vertices in G are the same. Therefore, either v_1 or v_2 must be in a minimal W. A set of twin pairs can have more than two vertices, all of which have the same set of distances. As noted above, in Figure 2, $N[3] = N[4] = N[5] = \{2,3,4,5\}$ indicating that this vertex set is a set of three twin pairs. This forces two of the three twins to be in a minimum W. In other words, given x number of twin vertices in the same twin set, then $x-1$ of the vertices must be in a minimal W for G. Any K_n has a twin set of cardinality n making $\dim(K_n) = n-1$ as is known.

2.6. Social Network Background

Considered to be social networks, collaboration networks have been one of the most active areas of research for the past couple of decades. In this note, the graphs formed by authors without papers, by authors with papers and by papers without authors are discussed since the papers, and their connected research activities, are the social groups. For this study, the assumption is made that departments have a physical existence where professors see each other on a regular basis, giving them the opportunity to discuss their research.

A *research group* is a collection of collaborating authors within the same organization while a *research network* includes collaborators from more than one organization [18]. Given these definitions, this note is focused on the departmental research groups that may exist in a research network. Collaboration can lead to a larger number of publications, career advancement and increased access to funds [35].

Milgram's impactful 1960s study [23] involved 160 random individuals in Nebraska who were requested to forward a letter to one of Milgram's Boston friends. A requirement to the forwarding was that the letter be sent only to people who the sender knew on a first-name basis. Even though Milgram's study was on a small scale, Milgram's requirement of first-name basis has been used to justify using collaboration networks as representative social networks as opposed to film actors in films [24] because it is assumed that coauthors tend to know each other on a first-name basis. As mentioned in [24], some papers have a very large laboratories as authors, so a first-name basis seems unlikely in those cases. The social network in any department is undoubtedly one where first names are known among its faculty.

Consider the different environments found in the three disciplines covered by this study. Both physics and biology can have complex physical laboratories while the mathematician's laboratory is typically paper and pen, or marker and board, or one or more computers. This difference results in a larger total number of collaborators for biology and physics compared to mathematics papers as discussed in [24]. Although a paper may have many authors, the only authors considered in this study are those from the same university department. Any given author team may produce a number of papers; but in this note, only a single paper that represents a distinct author collaboration structure is considered.

3. Data Collection Methods and Approach

This section discusses the data collection methods used in this study with a focus on the use of representative papers. The collected data's purpose is to generate collaboration network graphs on which metric dimension is used to compare the structures.

3.1. Data Collection Methods

Five years of public information (2019 to 2023) as found on Google Scholar, ResearchGate and Web of Science is used for the professors in the mathematics, biology and physics departments of three U.S. public universities. Duplicate papers are excluded. Utilizing the same logic as given in [24], preprints are included in this study when they do not duplicate published papers.

Selection of the three United States public universities is based on the following common characteristics as determined directly from each university's web site.

1. Total student enrollment is between 25,000 and 30,000 with a primary campus that includes a medical school and hospital. Primary campus is defined as containing at least 70% of the student population.
2. The basic structure of all three discipline departments is fundamentally the same; so each department has an applied faculty who work on medically related mathematics along with general research areas for that discipline.

Collaboration focuses exclusively on tenure track faculty in the same department. Professors who perform only research (have no teaching responsibilities) are eliminated because not all of the departments have these positions. Any professor officially listed on the web as being in more than one department is removed. In all nine of the departments, some faculty collaborate with both medical school personnel as well as members of their departments. In this case, the focus is exclusively on the collaborating authors within the studied department. Department inclusion of faculty during the study's five year period is determined from various public sources including institutional reference on published papers.

It is assumed that the web sites are accurate and current. This assumption is applied to both the university web sites as well as those for the nine departments. The assumption is made that professors are accurately listed in the various research areas.

3.2. Data Approach-Representative Papers

As our objective is to analyze the collaboration structure and not the total amount of collaboration activity, only representative papers are utilized. In other words, if two department authors collaborate on 20 papers within the five year period, only one representative paper is recorded for the collaboration. However, if a third author in the department is periodically added to the collaborating team, then a second representative paper is documented. Thus each representative paper has a unique department collaboration authorship.

3.3. Collaboration Group Size

The collaboration group in this note is the number of professors in the same department who are authors on a representative paper, not the total number of authors on an actual paper. In almost all instances, the actual number of authors is greater than the number who are from the same department. Exceptions to the last statement are typically found in the three math departments where the mathematicians from the same department are the only authors on the published paper.

4. People, Papers and Graphs

Let G_A be a graph that contains only authors where edges connect the authors collaborating together on research papers. Denote a graph based only on representative papers as G_P where an edge between two papers reflects that the papers have at least one author in common. In collaboration network studies, because the social groups are the papers, the actual social situation is the bipartite graph, G_{AP}, constructed with author set A and related paper set P.

4.1. Projection Graphs

Graph G_A is the graph discussed in most collaboration network analysis studies. This graph is a projection graph derived by projecting the author vertices $a \in A$ onto the papers $p \in P$ in the bipartite G_{AP} [28]. Graph G_P is the projection graph of the set of p onto the set of a in G_{AP}; so G_P depicts the structure of the research groups identified by the representative papers in this note. Thus $V(G_A)$ is $A \subset V(G_{AP})$ and $V(G_P)$ is $P \subset V(G_{AP})$. Figure 3 depicts G_A, G_{AP} and G_P of a few large components based on this study's collected data. Note, in some of the depicted graphs, but not all, G_{AP} is generated by subdividing every edge of G_A. Also note that for two of the graph trios, the G_A are isomorphic K_3 while their related G_{AP} are quite different, as are the two G_P.

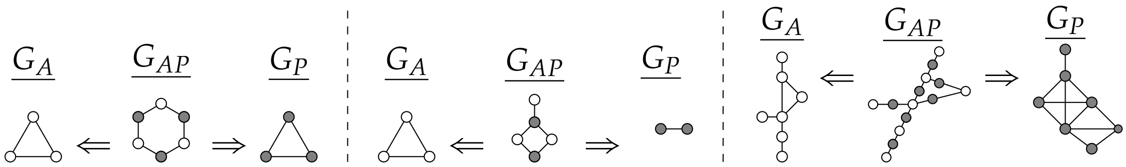


Figure 3. Examples of largest components from data derived G_A, G_{AP} and G_P.

4.2. *Construction of Projection Graphs from CNM*

Let a and p be vertices in $V(G_{AP})$, $a_{G_A} \in V(G_A)$ and $p_{G_P} \in V(G_P)$. The open neighborhood of a_{G_A} (or p_{G_P}) is the set union of a's neighbors' neighborhoods less a in G_{AP} (and the same for p) so duplicate vertices are eliminated. For a, let $p_i \in N(a)$ where $N(p_i) - a$ is a set of a_j less a (for p, let $a_j \in N(p)$ and $N(a_j) - p$ is the set of p_i less p.) Thus, $\deg(a_{G_A}) = \bigcup N(p_i) - a$ where $p_i \in N(a)$ and similarly for p_{G_P}. Hence, $N(a_{G_A})$ is based on $N(a)$, the degree of each $p_i \in N(a) - a$ less the number of neighbors shared among the set of p_i.

Given any G_{AP}, its CNM based on vertex nonadjacency can be utilized to construct G_A and G_P as follows. For bipartite G_{AP} with partite vertex sets A and P, define a graph G_A on vertex set A where vertices $a_i, a_j \in A$ and $a_i \sim a_j$ in G_A if there is a non-zero value in G_{AP}'s CNM at $c_{i,j} \in$ CNM. A similar graph G_P can be defined for vertices $p_i, p_j \in P$ in G_{AP}. The count of the non-zero entries in any row of G_{AP}'s CNM then gives the degree of vertex a_{G_A} indexing the CNM row and similarly for any $p_{G_P} \in G_P$. The order of G_A is $|A| \subset V(G_{AP})$ and $|G_P| = |P| \subset V(G_{AP})$. G_A and G_P are the two projection graphs of G_{AP} derived from G_{AP}'s CNM.

Thus, due to the bipartite nature of G_{AP}, if a distinguishing labeling is given to G_{AP} that clearly separates set $a \in A$ from the members of the set $p \in P$ (such as papers labeled $\geqslant 100$ and authors < 100), and the indices of CNM are in numeric order, then the CNM is a block matrix with $a \times a$, $a \times p$, $p \times a$ and $p \times p$ blocks. Based on nonadjacent vertices, the $a \times p$ and $p \times a$ blocks in the CNM for bipartite G_{AP} are all zero blocks. Note, the DM of a G_{AP} can also be used where the G_A (and G_P) is constructed based on vertex pairs in the $a \times a$ (also $p \times p$) block with distance 2.

5. Structural Specifics of the Authors with Papers Graph

The use of representative papers makes the structure of the G_{AP} very specific. However, consider any social group network such as actors and films, or women and their participation in Southern US social groups [5], or the collaboration structure found in current large research paper databases [24] with the concept of representative groups. Thus, the defined structure of the G_{AP} in this note applies to many similar social situations. Note, only even C_n with $n \geqslant 6$ and odd P_n with $n \geqslant 3$ are G_{AP}.

Projection graphs G_A and G_P can be either bipartite or non-bipartite. In the 18 data derived projection graphs, 22% of the G_A largest components and 44% of the G_P largest components are bipartite.

5.1. *Pendants, Degrees and Neighborhoods*

The papers in this study require at least two faculty members for them to be considered in a G_{AP}. This requirement and the use of representative papers restricts the possibilities for the structure of these graphs. Below is a list of the specific structural aspects of a G_{AP} large component and its related G_A and G_P. Proofs of the following statements are left to the reader.

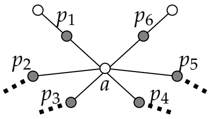
1. PENDANT VERTICES AND PENDANT CHAINS:
    (a) In any G_{AP}, only authors can be a pendant vertex.
    (b) Any pendant author in a G_{AP} is also pendant in its G_A.
    (c) Any pendant chain in a G_{AP} includes at least one author and one paper.
    (d) Any pendant paper in a G_P is not pendant in the related G_{AP}.
    (e) Any pendant vertex in a G_P must be in a pendant chain with minimum length of 3 in the related G_{AP}.
2. DEGREE AND DURFEE RANK: Since only authors can be pendant vertices in a G_{AP}, the minimum possible degree for authors is 1 while that of papers is 2.
    (a) For the set of paper vertices in a G_{AP}, $2 \leqslant r_{\square}(P) \leqslant \Delta(G_{AP})$.
    (b) For author vertices in a G_{AP}, $1 \leqslant r_{\square}(A) \leqslant \Delta(G_{AP})$.
    (c) $r_{\square}(G_{AP})$ can be greater than either $r_{\square}(A)$ or $r_{\square}(P)$.
    (d) In a G_{AP}, vertex $a \in A$ always projects to a $p \in P$ that has at least one neighbor in the G_{AP}.

   (e) Vertex p can project to an a vertex that has no other neighbor (i.e. a is pendant).

3. NEIGHBORHOODS: Each paper is a representative paper resulting in unique neighborhoods for all papers in any G_{AP}.
   (a) No paper vertex can be a twin of another paper vertex in a G_{AP}.
   (b) Author vertices can be in more than one twin pair.
   (c) All bipartite G_{AP} are planar due to the distinct neighborhoods of all p vertices.

5.2. Degree Projection

Consider the situation depicted in Figure 4 where author a has collaborated with other authors on six representative papers shown as gray vertices. Then each paper has vertex a as a common neighbor with the other five papers; so the p vertices are nonadjacent common neighbors of each other. This gives each of the papers in the related G_P *at least* a degree of 5, and places the six paper vertices in a K_6 subgraph in G_P. In other words, a high degree in one of the vertex sets of G_{AP} is projected onto the vertices of the other set in the latter set's projection graph. The degrees for papers p_2, p_3, p_4 and p_5 increase past 5 depending on the degree of the other author vertices to which each paper is adjacent in G_{AP}.



When G_P is created, the degree 6 of the central vertex a gets projected onto paper vertices. Thus, vertices p_1, p_2, p_3, p_4, p_5 and p_6 all obtain *at least* degree 5, and are in a K_6 subgraph in G_P.

Figure 4. Example with central degree 6 author and adjacent gray papers.

Proposition 5.1. *Assume connected G_{AP} has partite sets A of authors and set P of representative papers so $\Delta(G_{AP}) = \max\{\Delta(A), \Delta(P)\}$. In the projection of P vertices to the vertices in A, $\Delta(A)$ generates a $K_{\Delta(A)}$ subgraph in G_P; and by projection of A onto P, $\Delta(P)$ generates a $K_{\Delta(P)}$ subgraph in G_A.*

Proof. Given a G_{AP} as described, if $\Delta(A) = x$, there exists vertex $a \in A$ that is adjacent to x number of $p \in P$. Call the set of x number of p vertices P_x. Since the members of P_x share a as a neighbor, they have each other as common neighbors of a. So each pair of vertices in P_x has a nonzero value at their intersection element in G_{AP}'s CNM. Hence, there is a K_x subgraph in the G_P. The same reasoning applies to $\Delta(P) = y$ producing a K_y subgraph in G_A. □

If G_{AP}'s size m is even, the degree sequence of either A or P (most often P) can be a collection of all 2s. When this occurs, projection results in a degree sequence that is isomorphic to that of the vertex set in G_{AP}. This is due to degree projection creating a set of K_2 in the projection graph. As an example, consider C_6 with three $a \in A$ and three $p \in P$. The degree sequence for A is isomorphic to that of P and both sequences are $[2, 2, 2]$ so $\Delta(A) = \Delta(P) = 2$. These sequences generate G_A and G_P isomorphic to K_3. This is not a conflict to maximum degree generating a complete subgraph based on the maximum degree because three K_2 subgraphs are generated in both G_A and G_P; and K_3 is also C_3. In fact, if G_{AP} is an even C_n with $n \geqslant 6$ (required for distinct $N(p)$), then G_A and G_P are isomorphic cycle graphs each with order $\frac{n}{2}$ due to the all 2s degree sequences of both partite sets. If G_{AP} is an odd P_n with $n \geqslant 5$ then $|A| = |P| + 1$, and G_A and G_P are both path graphs due to $\Delta(A) = \Delta(P) = 2$.

There is a compounding effect that can occur in the projection graphs. Figure 5 displays two G_{AP} with gray $p \in P$ and white $a \in A$, where each G_{AP} is a C_8 with two chords. In both cases for that specific G_{AP}, the A and P degree sequences are identical and the projection graphs are isomorphic. For both G_{AP}, $\Delta(A) = \Delta(P) = 3$ yet $G_{AP}1$ on the left has a G_A (and isomorphic G_P) that is K_4 while $G_{AP}2$ on the right has a G_A (and also G_P) that is C_4 with a chord. Although metric dimension, $\dim(G)$, is discussed later in greater detail, $\dim(G_{AP}1) = 3$ while $\dim(G_{AP}2) = 2$. Although this may seem like a contradiction to Proposition 5.1, it is not. In both instances, the projection of the maximum degree 3 vertex produces K_3 subgraphs in the projection graph. The difference is due to the disparity in the structure of the neighborhoods due to the different locations of the degree 3 vertices. In $G_{AP}1$, author vertex b with degree 2 is adjacent to two degree 3 papers so b is adjacent to a, c and d in G_A. On the other hand, $G_{AP}2$ contains no degree 2 vertex adjacent to two degree 3 vertices. Thus, the four degree 2 vertices have only two neighbors in their respective projection graphs. Due to the importance of Proposition 5.1, next is a large component example derived from the collected data.

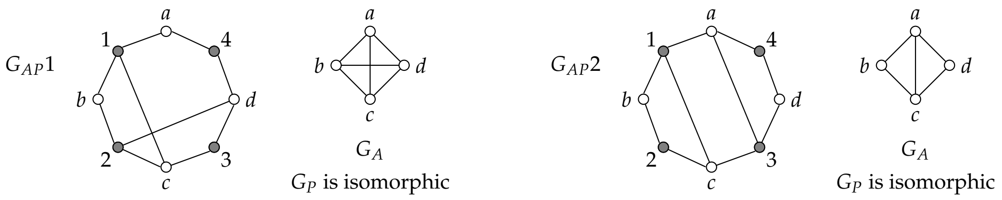


Figure 5. Two G_{AP} with the G_A and $G_P \cong G_A$.

For partite sets $a_i \in A$ and $p_j \in P$, it is a fact that $\sum_{i=1}^{|A|} \deg(a_i) = \sum_{j=1}^{|P|} \deg(p_j) = m$ where m is the size of the related G_{AP}. As with $G_{AP}1$ in Figure 5, the data derived large component of a G_{AP} in Figure 6 appears to contradict Proposition 5.1. However, due to the degree sum of A equaling the degree sum of P, the K_4 in G_A is still much smaller than the K_6 found in G_P due to $\Delta(A) = 6$ but $\Delta(P) = 3$. This particular G_P contains two vertices with degree 8 and two vertices with degree 7 as shown later in Figure 11.

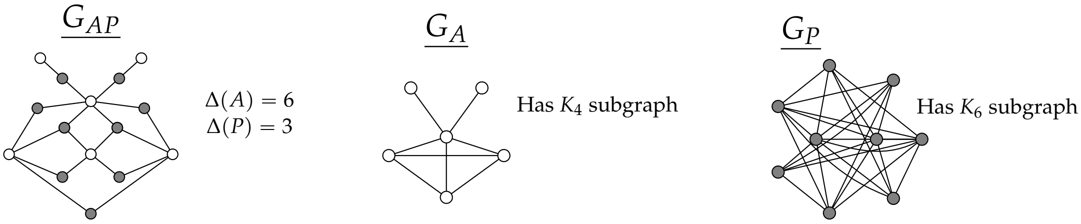


Figure 6. A G_{AP} with related G_A and G_P.

Remark 5.2. *Define* $\epsilon(G) = \frac{|E|}{|V|}$ *as in [12] and let* $\delta(G)$ *be the minimum degree of* G*. Excluding isolated vertices, Proposition 1.2.2 in [12] shows that large degree vertices are not scattered in vertices with smaller degrees. In other words, in any graph there exists a subgraph* H*, where* $H = G$ *may be true, such that* $\delta(H) > \epsilon(H) \geqslant \epsilon(G)$*. The proof of the proposition in [12] contains the following process. For* G *with at least one edge, construct an induced subgraph sequence* $G = H_0 \supseteq H_1 \supseteq \cdots$ *such that any* $v_i \in V(H_i)$ *where* $\deg(v_i) \leqslant \epsilon(H_i)$ *is deleted and* $H_{i+1} = H_i - v_i$*. The process stops when there are no more* $v_i \in V(H_i)$ *that can be deleted. This results in* $\epsilon(H_{i+1}) \geqslant \epsilon(H_i)$ *for all* i*, and an induced subgraph that contains the vertices with the larger degrees in* G*.*

Because invariant $\epsilon(G) = \frac{|E|}{|V|}$ *reflects the proportion of graph size to graph order,* $\epsilon(G) < 1$ *indicates a tree or forest. If* $\epsilon(G) = 1$*, then graph size equals graph order, one example of which is* C_n*. For* K_n*,* $\epsilon(K_n) = \frac{x-1}{2}$*; so for* $K_2 \cong P_2$*,* $\epsilon(K_2) = \frac{1}{2} < 1$ *and for* $K_3 \cong C_3$*,* $\epsilon(K_3) = 1$*.*

The situation that produces the closest complete subgraphs in the projection graphs is when at least one of the partite sets of G_{AP} has a degree sequence that consists of all 2s. Because vertex $a \in A \subset V(G_{AP})$ can have degree 1 and the minimum degree of $p \in P \subset V(G_{AP})$ is 2, if the degree sequence of A is all 2s then the degree sequence of P is also all 2s reflecting that G_{AP} is an even cycle graph with $n \geqslant 6$. In this case, the only complete projection graph is when G_{AP} is C_6 and both projection graphs are $K_3 \cong C_3$. For even C_n with $n \geqslant 6$, the projection graphs are isomorphic cycles composed of K_2 subgraphs. If G_{AP} is a subdivided G_A, then the degree sequence of P in G_{AP} is all 2s and $\Delta(A) \geqslant \Delta(P)$. For this situation, the degree sequence of G_A is isomorphic to the degree sequence of A in G_{AP} due to the all-2s P degree sequence.

Based on Proposition 5.1, if $\Delta(A) = x$ and $\Delta(P) = y$ where $x > y$, can the maximal induced complete subgraph in G_A have greater order than the maximal induced complete subgraph in G_P? The answer is "yes" for a specific case that follows where $\Delta(A) > \Delta(P)$ but the maximal complete subgraph in G_A exceeds the order of the complete subgraph in G_P.

Suppose G_{AP} is a subdivided $G_A \cong K_{|A|}$ where $\Delta(A) = x$. Because it poses a "small graph" exception concerning the number of K_x subgraphs in G_P, first let $x = 3$ so $|A| = 4$. Then G_{AP} is the subdivided K_4 and $G_A \cong K_4$, so G_A contains four K_3 subgraphs. G_P has $|P| = 6$, each $p \in V(G_P)$ has regular degree 4 and by Proposition 5.1, there exist eight K_3 subgraphs but these subgraphs do not form a K_4 or K_5. Instead there exist three sets of twin pairs in G_P. In this case, $3 = r_{\square}(G_A) < r_{\square}(G_P) = 4$ and $\mathsf{dim}(G_A) = \mathsf{dim}(G_P)$. Now more generally, let $\Delta(A) = x > 3$. Then in G_{AP}, $|A| = x+1$, all $a_i \in A$ have degree x and each pair in set A shares a single paper. Thus, $|P| = \frac{(x+1)(x)}{2} = y$ and G_P has $\Delta(G_P) = 2(x-1)$ for all p and has $\frac{x(x+1)(x-1)}{2}$ number of edges. There exist $x+1$ number of K_x subgraphs in $G_A \cong K_{x+1}$. Based on Proposition 5.1, the y number of $p_j \in V(G_P)$ are in K_x subgraphs but each p has degree $2(x-1)$ that is greater than degree $x-1$. To compare these particular G_A and G_P, consider ϵ given in Remark 5.2. For the G_A, $\epsilon(G_A) = \frac{x}{2}$ while $\epsilon(G_P) = x-1$; and $\frac{x}{2} < x-1$ for all $x > 3$. This shows that proportionally, there are more vertices with larger degrees in G_P than in G_A. However, the placement of the edges in G_{AP} only allows a p pair to have at most $x-1$ common neighbors. Thus, the largest complete subgraph in G_P is K_x which is smaller than $K_{x+1} \cong G_A$.

Also note that for the two regular degrees, $x < 2(x-1)$ so $x = r_{\square}(G_A) < r_{\square}(G_P) = 2(x-1)$. A graph's degree structure affects its metric dimension; so compared to G_A's DM, either equal or more rows in G_P's DM are required to produce unique column combinations. Hence, $\mathsf{dim}(G_A) \leqslant \mathsf{dim}(G_P)$.

Corollary 5.3. *Suppose connected G_{AP} has partite sets A of authors and set P of representative papers and G_{AP} is not a subdivided K_n. If $\Delta(A) > \Delta(P)$ then a maximal complete subgraph in G_A has smaller order than a maximal complete subgraph in G_P. If $\Delta(A) < \Delta(P)$ then a maximal complete subgraph in G_P has smaller order than a maximal complete subgraph in G_A.*

Proof. Suppose $\Delta(A) > \Delta(P)$ where $\Delta(A) = x$ and $\Delta(P) = y$. Assume that the vertex elimination process described in Remark 5.2 has been done on connected G_{AP} generating a graph G^*_{AP} containing only the larger degree vertices in G_{AP}. Let a_Δ have degree x and p_Δ have degree y. To create a maximal situation, let $a_\Delta \sim p_\Delta$ in G^*_{AP} and G_{AP}. Denote the set of p_Δ neighbors of a_Δ as $p_\Delta i$ where $1 \leqslant i \leqslant x$, and a_Δ neighbors of p_Δ as $a_\Delta j$ with $1 \leqslant j \leqslant y$. To get maximum degrees of a_Δ and p_Δ in their respective projection graphs, if it were possible that the neighbors of a_Δ and p_Δ shared no neighbors, then maximum $\deg_{G_A}(a_\Delta) = \left(\sum_{i=1}^{\deg(a_\Delta)} \deg(p_{\Delta i})\right) - \deg(a_\Delta) = xy - x$ and similarly for p_Δ making maximum $\deg_{G_P}(p_\Delta) = yx - y$. Since $x > y$, $\deg_{G_A}(a_\Delta) < \deg_{G_P}(p_\Delta)$ for all $a_\Delta \in G_A$ and all $p_\Delta \in G_P$. Similar logic shows $\deg_{G_P}(p_\Delta) > \deg_{G_A}(a_\Delta)$.

Still assuming that all a_Δ are adjacent to all p_Δ in G^*_{AP} (and G_{AP}) where $x > y$, now let the neighbors of a_Δ and p_Δ share neighbors, let $N(a_\Delta)$ be a set of $p_\Delta i$ where $N(p_\Delta i)$ is a set of a_Δ so $|N(a_\Delta)| > |N(p_\Delta)|$. Thus the probability of shared neighbors for the smaller distinct neighborhoods of $p_\Delta i$ is greater than the probability of shared neighbors in the larger possibly non-distinct $a_\Delta i$ neighborhoods resulting in $\deg(a_\Delta)$ in G_A being less than $xy - x$ that is already less than $xy - y$ for the p_Δ. Thus for $a_\Delta \in G_A$, $|N(a_\Delta)| \in G_A$ is less than $|N(p_\Delta)| \in G_P$. Because all $p_\Delta i$ are neighbors of a_Δ, they form a $K_{\Delta(A)}$ subgraph in G_P, and the same for the $a_\Delta i$ in G_A. Hence, the maximal complete subgraph in G_A is then smaller than the maximal complete sungraph in G_P.

When $x < y$, the larger neighborhoods are distinct. Thus, there exists a greater probability of shared neighbors in the smaller $N(a_\Delta i)$ because the vertices can have non-distinct neighborhoods (twins are permitted) and $N(p_\Delta)$ have more $a_\Delta i$ in this case. This gives the p_Δ in G_p a degree smaller than $xy - y$ that is smaller than $xy - x$ in this case.

First assume $\Delta(A) > \Delta(P)$ in G_{AP} that is not a subdivided K_n. For the sake of contradiction let the a_Δ in the associated G_A be in a larger maximal K_n subgraph than the p_Δ in the maximal K_n of the related G_P. This implies that each p_Δ has more unique (non-shared) common neighbors compared to the a_Δ in G_{AP}'s CNM. However, the probability of the last statement referring to G_{AP}'s CNM is zero due to the $N(a_\Delta i)$ having a greater chance of shared neighbors in G_{AP}. Now assume that $\Delta(A) < \Delta(P)$ in G_{AP} and that the a_Δ in the associated G_A are in a smaller maximal K_n subgraph than the p_Δ in the maximal K_n of the related G_P. Using similar logic as when $\Delta(A) > \Delta(P)$ again reveals a zero probability of this situation. $\square$

6. Social Aspects of the Data Collected

In this section, the data collected from the three university web sites is examined first, followed by discussing each of the three disciplines regarding the department chairs as research focal points.

Assortative mixing occurs when members of social groups associate with each other based on specific characteristics. In this study, all professors have specific areas of focus within their general research area as given on the department web site. As expected, with the exception of two papers, professors collaborate with other professors in the department who share their specific research area. The two exceptions are education papers where department professors who do not have education as their research area, coauthor with the education researcher in their department. Although not given in this note, dendrograms based on distances successfully identified clear communities in the G_A and G_{AP} centered on research areas.

6.1. Analysis of University Data

Various characteristics of the three universities are collected from the institutions' web sites with averages (means) and standard deviations (σ) presented here. Focus is exclusively on each university's primary campus. The average student to teacher ratio is 15:1 with $\sigma = 1.5$. The average total student population is 27,136 ($\sigma = 2,023$). Of the total student population, there is an average of 20,340 undergraduates ($\sigma = 1,064$) representing 75% of the student body, and 6,795 graduate students ($\sigma = 2,126$) for 25%. There is an average of 22,857 full-time students ($\sigma = 1,219$) or 84%. The average in-state student population is 19,818 students ($\sigma = 7,639$) or 73%.

By discipline, Table 1 displays the mean department size and the mean number and percent of faculty performing departmental collaboration. Overall, six of the nine department chairs (67%) collaborate within their department.

Table 1. Faculty collaboration information for the nine departments.

Discipline	Mean # faculty	Mean # and (%) collaborate
Math	29	14 (48%)
Physics	26	13 (50%)
Biology	27	14 (52%)

6.2. Analysis of Discipline Data: Hubs Analysis

Define a *hub* as any author vertex whose degree is greater than the average department degree plus one standard deviation. An analysis of hubs is done for both the nine G_A and the nine G_{AP} (with paper degrees excluded) with a focus on department chairs. Professors in a hub position may exert greater influence as far as research in a department is concerned [36].

in any G_{AP}, author vertices with larger degree indicate professors involved with a greater number of representative papers that reflect distinct research groups in the department. In any G_A, the interpretation of vertices with the greater degree is not well defined as there exist two possible meanings. In a G_A, a high degree a can reflect professors associated with either representative papers that have a larger number of departmental faculty authors or a large number of representative papers that may have a few authors.

doi:10.20944/preprints202507.0495.v1

The data in Table 2 gives the hubs analysis for the nine data derived G_A and the nine data derived G_{AP}. Overall, 33% of the department chairs are hubs in their departments. Any G_{AP} depicts the actual social network, not its G_A projection graph. In this analysis, the value defining the hubs is calculated separately for the nine G_A and the nine G_{AP}. The G_{AP} value is based on the author degree only (paper degree is excluded). Due to degree projection, in this analysis, 33% of the 18 graphs examined display a different set of hubs between the G_A and the related G_{AP}.

Table 2. G_A and G_{AP} hub analysis based on collected data from the nine departments.

Discipline	G_A Total # hubs	G_A chair as hub	G_{AP} Total # hub	G_{AP} chair as hub
Math	9	1	11	2
Physics	5	1	9	0
Biology	8	1	8	1

Utilizing the G_{AP} to assess the number of hubs produced an increase in total number of hubs from 22 (see Table 2) to 28 (27% increase). Although the total number of chairs acting as hubs remained the same, the number shifted from 1 to zero for physics and from 1 to 2 for mathematics. In comparing the two table sections, notice that there is no difference for the Biology row. An examination of the three data derived biology G_A to their related G_{AP} reveals that each G_{AP} is a subdivided version of the G_A in all three cases.

Focusing exclusively on the structure of G_A can give misleading interpretations as shown by the different hub analysis results between the G_A and the related G_{AP} and the fact that the interpretation of the large degree structure in a G_A is not well defined. These differences can impact an analysis of a network's evolution over time.

Conjecture 6.1. *Given an author and paper collaboration network structure, analysis of projection graphs G_A and G_P, plus bipartite G_{AP}, is required in the prediction of future network links or edges and general network evolution.*

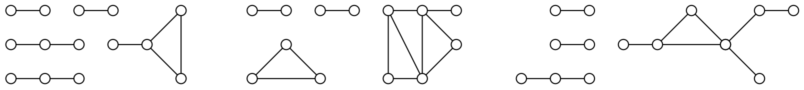
7. The Nine Average Graphs

Derived from the 27 data derived graphs, the average graphs are covered in this section. For each discipline, the average G_A, average G_{AP} and average G_P are given, and are only briefly discussed. Focus in the discussion is on the large component, L_G, of each average graph. Only the metric dimension for the large component L_G of each average graph is displayed.

An important requirement in the transition between the average G_A and the average G_{AP}, and between the average G_P and the average G_{AP}, is that projecting the a in the average G_{AP} onto the p in the average G_{AP} must produce the average G_A determined from calculating the average parameters; and similarly projecting the $p \in V(G_{AP})$ onto the $a \in V(G_{AP})$ must generate the calculated average G_P.

7.1. Authors Only Average Graphs and Analysis

As expected, in Figure 7, the number of components in all of the average graphs is similar since overall, 50% of the professors do departmental collaboration with little difference between average discipline department size as displayed in Table 1. Compared to the results in [24], the percent of vertices in the large components here is much less.



L_{G_A} $\dim(G_A) = 2$ L_{G_A} $\dim(G_A) = 2$ L_{G_A} $\dim(G_A) = 2$

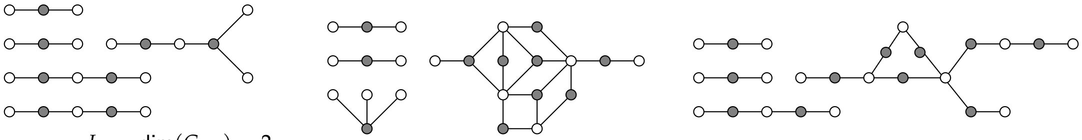
Average math G_A Average physics G_A Average biology G_A

Figure 7. The average G_A graphs for the three disciplines.

In comparing the large components of the three average G_A in Figure 7, notice that this component is much smaller for mathematics. The greater connectivity of the physics G_A is explained by the fact that the physics G_{AP}'s large component in Figure 8 has 5 K_3 subgraphs while the other two G_{AP} contain a single K_3. Although $\Delta(G_A)$ is close for the three disciplines' large components, the physics G_A has $r_{\square}(G_A) = 3$ while math and biology have $r_{\square}(G_A) = 2$. As shown in the figure, for the L_{G_A}, $\dim(G_A) = 2$ for all three disciplines.

7.2. Authors with Papers Average Graphs and Analysis

Figure 8 displays the three average G_{AP} for this study. Consider the L_G for the average physics G_{AP}. Notice that one author in the physics L_G has degree 5 while two other authors have degree 4. The relatively high degrees for the physics author vertices indicate that these professors have significant variety in their research group construction since representative papers are utilized. As depicted and previously mentioned, the average physics G_{AP} structure in Figure 8 reflects the three data derived G_{AP}. In other words, all of the physics departments in this study display a high amount of variety in their departmental collaboration research groups. The three G_{AP} have distinct $\dim(G_{AP})$ for their $L_{G_{AP}}$.



$L_{G_{AP}}$ $\dim(G_{AP}) = 2$ $L_{G_{AP}}$ $\dim(G_{AP}) = 4$ $L_{G_{AP}}$ $\dim(G_{AP}) = 3$

Average math G_{AP} Average physics G_{AP} Average biology G_{AP}

Figure 8. The average G_{AP} graphs for the three disciplines.

The total number of papers produced during the study's time frame by all considered professors was determined. On average, the physics professors produced 3.5 times as many papers as the other two research areas. In two of the physics departments, the number of representative papers outnumbered the authors by 150% due to the variety in the research group construction. Could the high level of productivity be due to the departmental collaboration style of the physics professors? Greater variety in research group construction allows for a broader range of skills and knowledge in collaborative research.

7.3. Papers Only Average Graphs and Analysis

Figure 9 clearly reflects the departmental collaboration style difference between physics and the other two research areas. Notice that the L_{G_P} has $\dim(G_P) = 1$ for math, $\dim(G_P) = 4$ for physics and $\dim(G_P) = 3$ for biology.

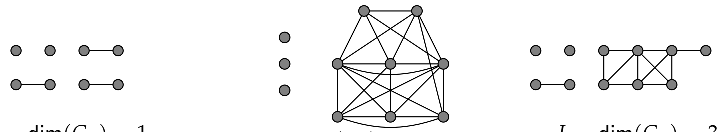


Figure 9. The average G_P graphs for the three disciplines.

8. The Degree Diagram and the Durfee Rank

Figure 10 displays a connected G_{AP} along with the degree diagrams for G_{AP}'s vertex sets A and P. Although $\Delta(A) = \Delta(P) = 3$, set A has more vertices with degree 3 than P; so $r^o_\square(A) = 3$ and $r^i_\square(P) = 2$. One impact of the difference in the two $r_\square$ is that $\dim(G_A) = 2$ and $\dim(G_P) = 3 = \dim(G_{AP})$. Can G_P more accurately reflect the actual collaboration in a department instead of the related G_A? Notice that $|A| < |P|$ in this case.

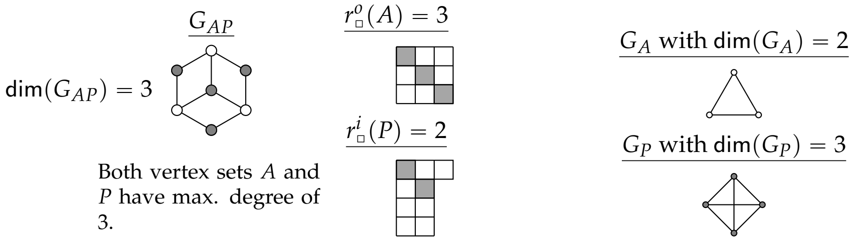


Figure 10. G_{AP} with maximum degree 3 in both A and P vertex sets.

Consider the large component $L_{G_{AP}}$ from a data derived G_{AP} depicted in Figure 11 (also displayed in Figure 6). On the top far left is the degree diagram of the G_{AP} degree sequence. Set $V(G_{AP})$ has $\Delta(G_{AP}) = 6$ and $r^o_\square(G_{AP}) = 4$. On the right side of G_{AP} are the degree diagrams for sets A and P in G_{AP}. Although the degree sequence for A is not graphic, each degree diagram clearly displays the degree relationships within each of the vertex sets compared to the other set. In comparing the two degree diagrams, $\Delta(A) = 6$ and $r^o_\square(A) = 4$ while $\Delta(P) = 3$ and $r^p_\square(P) = 2$. Note that the degree diagram of G_{AP} is merely the union of the degree diagrams for sets A and P with $\Delta(G_{AP}) = \Delta(A)$ and $r^o_\square(G_{AP}) = r^o_\square(A)$ in this case.

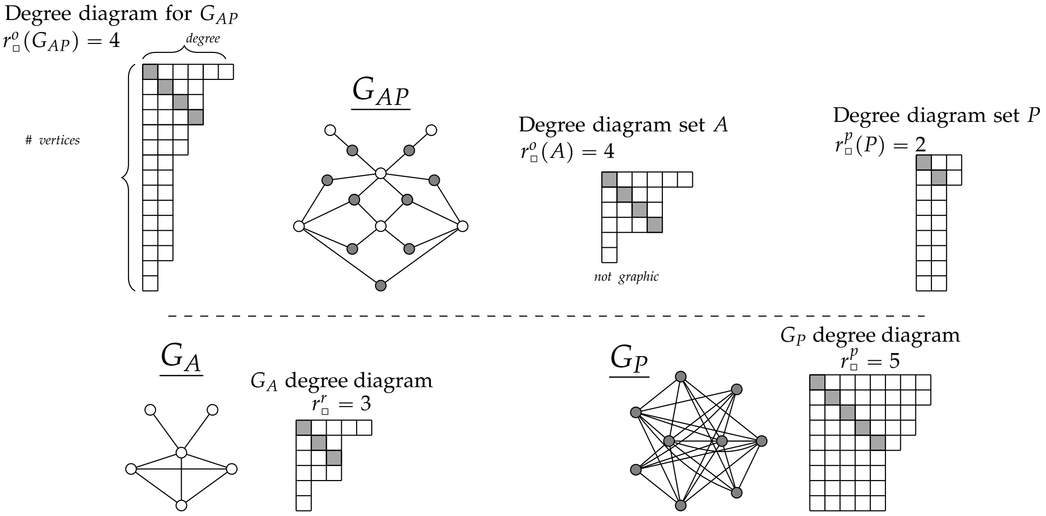


Figure 11. TOP: Data derived G_{AP} and the degree diagrams for G_{AP}, vertex set A and set P. BOTTOM: The related G_A and G_P with degree diagrams.

doi:10.20944/preprints202507.0495.v1

Now examine G_A and G_P in the lower portion of Figure 11. Here $\Delta(G_A) = 5$ and $r^r_\square(G_A) = 3$ reflecting a decrease in these figures compared to those for set A in G_{AP}. For G_P, $\Delta(G_P) = 8$ compared to $\Delta(P) = 3$ and $r^p_\square(G_P) = 5$ versus $r^p_\square(P) = 2$. This significant change is due to degree projection in G_{AP} where $\Delta(A) = 6$ generates a K_6 subgraph in G_P. Due to $|A| < |P|$ in G_{AP} and the Pigeonhole Principle, all of the high degree vertices are authors as shown by $r^o_\square(G_{AP}) = r^o_\square(A) = 4$ and $\Delta(P) = 3$. This results in all of the papers that are coauthored by the high degree authors gaining a combination of the high degrees in G_P.

Because metric dimension is used to compare networks, and based on the change in the hubs analysis between some G_A and their related G_{AP}, the next section examines the change in metric dimension given a G_{AP} and its related G_A and G_P.

9. The Metric Dimension of the Authors with Papers Graph

This section examines changes in the relative distance structure of G_A, G_P and G_{AP}, provides methods for finding $\dim(G_{AP})$. The metric dimension of G with multiple components is the sum of the metric dimension of each component in G. Although the average graphs all have multiple components, this section is focused exclusively on the changes in any single large component treated here as a connected graph. First the change between $\dim(G_{AP})$ to $\dim(G_A)$ and to $\dim(G_P)$ for the average graphs is examined.

9.1. Changes Between the Three Data Derived Average Graphs

For each discipline, the metric dimension of that discipline's three average graphs' L_G are compared. Unless otherwise stated, reference to G_A is referring to the large component of G_A; and the same is true for G_{AP} and G_P. The comparison begins with the mathematics discipline followed by physics and concludes with biology.

9.1.1. Metric Dimension in Average Mathematics Graphs:

With respect to the three average mathematics graphs, $\dim(G_A) = 2$, $\dim(G_{AP}) = 2$ and $\dim(G_P) = 1$. In this case, $\Delta(G_{AP}) = \Delta(P) = 3$ generating a K_3 in G_A. Thus the relative distance structures of G_A and G_{AP} are similar, and the projection generating G_P is not structure preserving with respect to the relative distance structure of G_{AP} even though $d_{G_P}(p_i, p_j) = \frac{1}{2} d_{G_{AP}}(p_i, p_j)$ where $p_i, p_j \in P$.

9.1.2. Metric Dimension in Average Physics Graphs:

In examining the three average physics graphs, $\dim(G_A) = 2$, $\dim(G_{AP}) = 4$ and $\dim(G_P) = 4$ with $\Delta(G_{AP}) = \Delta(A) = 5$ producing a K_5 subgraph in G_P where $\Delta(G_P) = 7$. In this case the distance structures of G_A and G_{AP} are different while the distance structures of G_P and G_{AP} are similar. Here the projection generating G_A does not preserve the relative distance structure of G_{AP}. This again reflects the importance of paper inclusion.

9.1.3. Metric Dimension in Average Biology Graphs:

The metric dimensions of the three average biology graphs are $\dim(G_A) = 2$, $\dim(G_{AP}) = 3$ and $\dim(G_P) = 3$. For biology, $\Delta(G_{AP}) = \Delta(A) = 4$ generating a K_4 in G_P with $\Delta(G_P) = 5$. Similar to physics, the relative distance structure of G_{AP} is that of G_P, not G_A.

9.2. Double Distance and Diameter

As mentioned, $d_{G_{AP}}(a_i, a_j)$, is double $d_{G_A}(a_i, a_j)$; and similarly for $d_{G_{AP}}(p_i, p_j)$ and $d_{G_P}(p_i, p_j)$. The double distance fact does not apply to the diameter of G_A, G_P and G_{AP}. In all cases, $\text{diam}(G_{AP}) > \text{diam}(G_A)$ and $\text{diam}(G_{AP}) > \text{diam}(G_P)$ due to the double distance relationship. In some cases, but not all, when the diameter in G_{AP} is an author to author path, then $\text{diam}(G_{AP}) = 2 \cdot \text{diam}(G_A)$. In all cases, $\frac{\text{diam}(G_{AP})}{2} > \text{diam}(G_P)$ and $\text{diam}(G_A) \geqslant \text{diam}(G_P)$ since no paper can be a pendant vertex in G_{AP}. Recall that, by affecting the possible distance variety in a graph's DM, diameter impacts metric dimension of a graph. When the diameter is a to a, or p to p, it is even; and when the diameter is a to p, it is odd.

9.3. Using the DM for Any Graph'S Metric Dimension

For any G with a labeling that generates distinct DM blocks, the blocks can be utilized to determine $\dim(G)$ resulting in a significant increase in computational efficiency. Finding a resolving set is relatively simple. However finding a *minimal* resolving set is NP-hard [15] as all possibilities must be explored. Regarding a graph's DM, all rows must be compared to all other rows in order to find a minimal number of rows that resolve G with unique column combinations. Thus, being able to find $\dim(G)$ using matrix blocks gives significant computational efficiency. Theorem 9.7 states that given $\dim(a \times a) \neq \dim(p \times p)$ only three of the four DM blocks need to be resolved in order to find $\dim(G_{AP})$. Focus in this note is now on methods for finding $\dim(G_{AP})$.

9.4. DM Block Resolving

Throughout the rest of this note, it is assumed that any G_{AP} has a distinguishing labeling that generates blocks in its DM. As reflected in the DM blocks for any G_{AP}, because G_{AP} is bipartite, distances between elements in the same vertex set are all even, while those between the two partite sets are odd.

Definition 1. DM block resolving *is the process of using the portions of DM rows (or columns) contained in the blocks to find the minimum number of rows in each DM block that give unique column (or row) combinations.*

The phrase "block resolving" is used when it is clear that the blocks are those in a DM. When using block resolving either rows, or columns, can be utilized but it is critical to use the same method for all blocks that are resolved. In this note, rows are used for block resolving and it is assumed that the reader understands that the choice of columns also exists.

All column combinations that contain a zero are unique with the zero implying that the row index vertex is in a W set for the block. The minimum number of rows that gives unique column combinations for that block is denoted by $\dim(a \times a)$, $\dim(a \times p)$, $\dim(p \times a)$ and $\dim(p \times p)$.

An *even block* is a block that contains only even entries; so the $a \times a$ and $p \times p$ blocks are even. Analogously, the $a \times p$ and $p \times a$ blocks are *odd blocks*. General terms referring to the minimum number of rows that resolve a block are $\dim(even)$ and $\dim(odd)$. The term $\dim(block)$ refers to minimally block resolving of a general block, either even or odd.

When using DM block resolving to find $\dim(G_{AP})$, the focus is on the even blocks as these blocks have row and column indices from a particular partite set, and as shown later, these block are also related to the projection graphs. The odd blocks are sometimes referred to as a *block extension* because literally, these blocks extend the rows of the even blocks by giving the relative distance relationships to the vertices in the other partite set.

Prior to giving four example G_{AP} selected for the variety of their block resolving results, the characteristics of the even and odd blocks are discussed.

9.5. DM Block Characteristics

Reference to a *block row* refers only to the portion of the DM row contained in the specific block being discussed.

9.5.1. Even Blocks:

Even blocks are always square and symmetric across their diagonal. The dimension of an even block is the partite set cardinality whose vertices index the block. The entries are all even with a single zero in each row, but the entries in the $a \times a$ block compared to the $p \times p$ block are not necessarily identical sets of even numbers.

Even blocks contain only even diameters. Because any graph has only one diameter measure, an even diameter plus the zeros, provide an even block with greater resolving efficiency by providing greater distance variety compared to its extension block. An even diameter can be in either the $a \times a$ block only, in the $p \times p$ block only or in both even blocks.

Recall that when row resolving, the zero indicates that that row's index vertex is in an ordered W set, so any column combination with a zero is automatically unique. Thus the existence of the single zero automatically reduces the number of column combinations that need to be checked for uniqueness.

9.5.2. Odd Blocks

Odd blocks are only square when $|A| = |P|$. The entries in the odd blocks are all odd with 1s indicating the degree of the vertex that is the row's index. Thus, $\Delta(A)$ and $\Delta(P)$ impact the distance variety in the odd blocks while $r_{\square}(A)$ and $r_{\square}(P)$, including their $r_{\square}$ type, indicate the number of rows that might have a larger number of 1s. The odd blocks' symmetry is to each other, across the diagonal of the DM, resulting in the two odd blocks having the same sets of odd integers. In other words, the rows of the $a \times p$ block are the columns of the $p \times a$ block and vice versa.

An odd diameter is in both odd blocks and increases distance variety. As extensions of the even blocks, the odd blocks can restrict the use of the related even block vertices in a minimal W set for the G_{AP}.

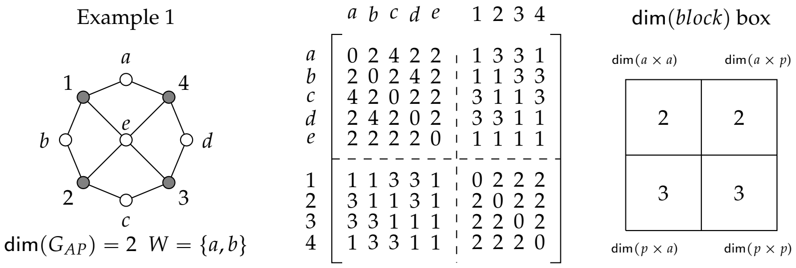
9.6. DM Block Resolving Examples

When block resolving, $\dim(a \times a)$ and $\dim(p \times p)$ always need to be determined. Based on the results of these two metric dimensions, when $\dim(a \times a) \neq \dim(p \times p)$ resolving only one more block is required. Following are four simple example G_{AP} with brief explanations of their block resolving.

The adjacency matrix for a K_n, $A(K_n)$ has a distinct recognizable structure of all 1s except for the all zero diagonal. The matrix resulting from $2 \cdot A(K_n)$ is $A(K_n)$ with the 1s replaced by 2s. When the $2 \cdot A(K_n)$ structure is found as a subblock in a DM even block, it is called a K_n *double subblock*. The importance of these subblocks relates back to Proposition 5.1, and the double distance relationship between G_{AP} and its projection graphs.

9.6.1. Example 1:

Figure 12 displays a G_{AP} with its DM and $\dim(block)$ box that displays the metric dimension of each block.



$$\begin{array}{c|ccccc|cccc}
 & a & b & c & d & e & 1 & 2 & 3 & 4 \\
\hline
a & 0 & 2 & 4 & 2 & 2 & 1 & 3 & 3 & 1 \\
b & 2 & 0 & 2 & 4 & 2 & 1 & 1 & 3 & 3 \\
c & 4 & 2 & 0 & 2 & 2 & 3 & 1 & 1 & 3 \\
d & 2 & 4 & 2 & 0 & 2 & 3 & 3 & 1 & 1 \\
e & 2 & 2 & 2 & 2 & 0 & 1 & 1 & 1 & 1 \\
\hline
1 & 1 & 1 & 3 & 3 & 1 & 0 & 2 & 2 & 2 \\
2 & 3 & 1 & 1 & 3 & 1 & 2 & 0 & 2 & 2 \\
3 & 3 & 3 & 1 & 1 & 1 & 2 & 2 & 0 & 2 \\
4 & 1 & 3 & 3 & 1 & 1 & 2 & 2 & 2 & 0
\end{array}$$

Figure 12. Example 1 for DM block resolving.

In examining the DM in Figure 12, the entire rows for a and b have unique column combinations for all vertices in the graph as reflected by $\dim(a \times a) = 2 = \dim(a \times p)$. There is a K_4 double subblock in $p \times p$ making $\dim(p \times p) = 3$. The row length in $p \times a$ is 5 but resolving two odd digits in two rows gives only $2^2 = 4$ unique column combinations; so the row length of 5 requires 3 rows for $\dim(block)$.

9.6.2. Example 2:

Figure 13 displays another G_{AP}. Both even blocks contain a K_n double subblock and $\dim(odd)$ does not agree with $\dim(even)$ for which the odd block is an extension.

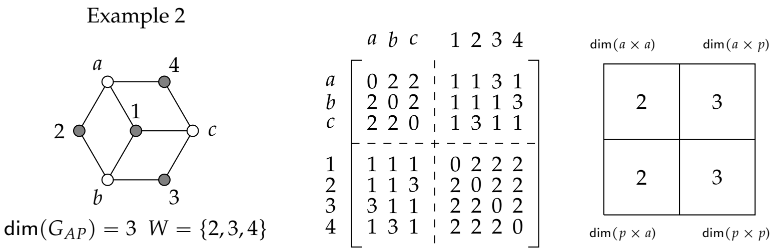


Figure 13. Example 2 for DM block resolving.

In this case, utilizing $\dim(a \times a) = 2$ is not possible for a minimal W in G_{AP} because the number of 1s in the extension block generates $\dim(a \times p) = 3$. Thus, a minimal W is either all three a; or three p, the choice of which depends on the rows that resolve $p \times a$. Note that because $\dim(a \times a) \neq \dim(p \times p)$ once these values are known, checking whether the smaller $\dim(even)$ provides $\dim(G_{AP})$ can be done by resolving only the extension block for the block with the smallest $\dim(even)$.

9.6.3. Example 3:

Example 3 shown in Figure 14 displays a G_{AP} where $\dim(a \times a) = \dim(p \times p)$ but $\dim(odd)$ is greater than $\dim(even)$.

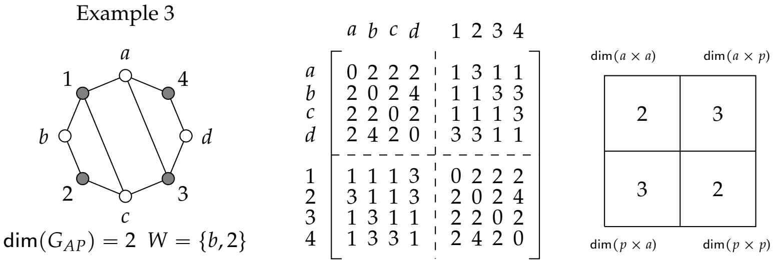


Figure 14. Example 3 for DM block resolving.

In this case, a minimal W must be constructed with both a and p vertices; so for this example, $W = \{b, 2\}$ where both DM rows contain $\text{diam}(G_{AP}) = 4$. Utilizing three a or three p as dictated by $\dim(odd)$ resolves G_{AP} but not minimally. Thus, when $\dim(a \times a) = \dim(p \times p)$ all four blocks should be resolved.

9.6.4. Example 4:

As shown in Figure 15, the G_{AP} in Example 4 generates $\dim(a \times a) \neq \dim(p \times p)$ and $\dim(a \times p) \neq \dim(p \times a)$.

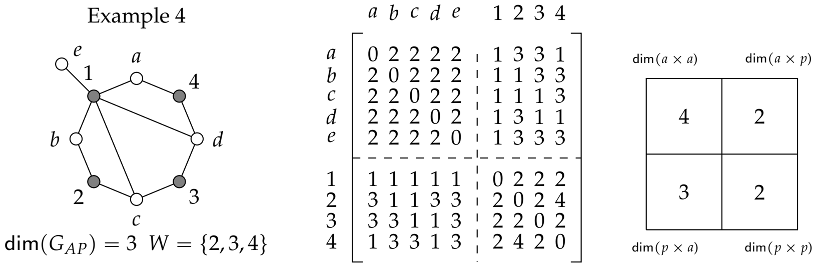


Figure 15. Example 4 for DM block resolving.

This example provides an additional situation where $\dim(a \times a) \neq \dim(p \times p)$ and resolving the extension block for the block with the smaller $\dim(even)$ provides $\dim(G_{AP})$ by restricting $\dim(a \times a) = 2$. In other words, two a vertices do not resolve G_{AP} but three a give $\dim(G_{AP})$. Note that in this case

$\mathrm{dim}(G_{AP}) \neq \mathrm{dim}(even)$ and that $r^p_\square(A) = 2$, $r^i_\square(P) = 2$ but $r^o_\square(G_{AP}) = 3$. As mentioned, there is more to explore in the $r_\square$ concept than what given in this note.

9.7. Relation of the DM Blocks to the Projection Graphs

The following proposition relates the DM blocks to the two projection graphs.

Proposition 9.1. *For the distance matrix (DM) of G_{AP} with authors set A and representative papers set P where G_{AP} has a distinguishing labeling,* $\mathrm{dim}(a \times a) = \mathrm{dim}(G_A)$ *and* $\mathrm{dim}(p \times p) = \mathrm{dim}(G_P)$*.*

Proof. Because G_A is constructed by projecting the vertices in A onto the vertices of P in the related G_{AP}, for each $a_i, a_j \in V(G_A) \subset V(G_{AP})$, $\deg_{G_A}(a_i)$ is the number of 2s in row a_i of the $a \times a$ block of G_{AP}'s DM. In other words, if there exists a 2 in the $a \times a$ block at the intersection of row a_i and column a_j, then $a_i \sim a_j$ in G_A because $d_{G_{AP}}(a_i, a_j) = 2 \cdot d_{G_A}(a_i, a_j)$ indicating that a_i shares a neighbor p with a_j in G_{AP}. Thus, if all entries in the $a \times a$ block are multiplied by $\frac{1}{2}$ the resultant block is isomorphic to G_A's DM. It follows then that a combination of rows that minimally resolve $a \times a$ also minimally resolve G_A. Hence, $\mathrm{dim}(a \times a) = \mathrm{dim}(G_A)$. Given $p_i, p_j \in V(G_P) \subset V(G_{AP})$, the same reasoning applies resulting in $\mathrm{dim}(p \times p) = \mathrm{dim}(G_P)$. ☐

Suppose that for a G_{AP} and its related G_A and G_P, $\mathrm{dim}(G_A) = x$ and $\mathrm{dim}(G_P) = y$ where x and y may, or may not, be equal. Due to the double distance relationship between G_{AP} and its projection graphs, then x number of a minimally resolve all $a \in A$; but x number of a do not necessarily minimally resolve the vertices in A to those in P becuase the double distance relationship does not apply. Thus, x number of vertices do not necessarily minimally resolve G_{AP}. Given that $\mathrm{dim}(G_P) = y$, then y number of p vertices minimally resolve the vertices in P, but not necessarily minimally resolve set A nor minimally resolve G_{AP}.

Proposition 9.2. *Any minimal resolving set of G_A minimally resolves the vertices $a_i, a_j \in A \subset V(G_{AP})$ but not necessarily the vertices in $P \subset V(G_{AP})$; so not necessarily G_{AP}. Likewise any minimal resolving set of G_P minimally resolves vertices $p_i, p_j \in P \subset V(G_{AP})$ but not necessarily the vertices in $A \subset V(G_{AP})$; so not necessarily G_{AP}.* ☐

9.8. Complete Graph Double Subblocks

As stated previously, if a 2 is at the DM intersection of v_i and v_j then these vertices are adjacent in the related projection graph. If the same v_i and v_j are found in a K_n double subgraph of an even DM block, then v_i and v_j are found in a K_n subgraph in the related projection graph, thus impacting the metric dimension of the projection graph. Proof of the following proposition is given by the gray boxes of the DM in Figure 16.

Proposition 9.3. *Given the distance matrix (DM) for a G_{AP} with authors set A where $\Delta(A) = x$ and representative papers set P with $\Delta(P) = y$, if G_{AP} is given a distinguishing labeling that is consecutive around its even cycle subgraphs, then a K_y double subblock is possible in the $a \times a$ block and a K_x double subblock is possible in the $p \times p$ block.* ☐

9.9. Existence of Twin Pairs

The existence of twin pairs can affect the possible results for the DM blocks. Figure 16 shows that it is possible to have $\mathrm{dim}(p \times a) = 0$ reflecting the impact of twins on block resolving.

Proposition 9.4. *Given a G_{AP} with authors set A and representative papers set P and utilizing row block resolving of G_{AP}'s DM,* $\mathrm{dim}(p \times a) = 0$ *if and only if G_{AP} has at least one twin pair.*

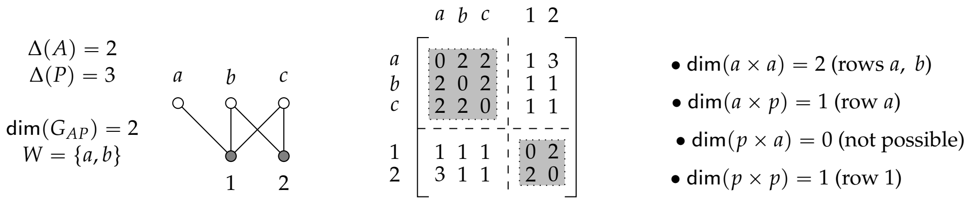


Figure 16. A G_{AP} where $|A| > |P|$ and $\dim(p \times a) = 0$ due to twin pair $\{b, c\}$.

Proof. If G_{AP} has at least one twin a pair, by the definition of a twin pair, the rows in $a \times p$ indexed by the twin pair have duplicate entries; and there exist duplicate columns in the $p \times a$ block so $\dim(p \times a) = 0$.

When $\dim(p \times a) = 0$, there must exist duplicate columns in this block. Only the vertices in A can be twins; and any twin pair has identical rows in the $a \times p$ block; so twin vertices have identical columns in the $p \times a$ block and $\dim(p \times a) = 0$. If there are no twin vertices, then there are no identical columns in the $p \times a$ block so it is resolvable and $\dim(p \times a) \neq 0$. □

Twin a vertices also generate duplicate column combinations in the $a \times a$ block because, except for the single zero, their rows have duplicate entries and the $a \times a$ block is symmetric along its diagonal. With their need for unique neighborhoods, no two papers can be twins. It follows then that for a G_{AP}, the $a \times p$ block is always resolvable when using row block resolving. When $\dim(p \times a) = 0$, no W for the G_{AP} can contain *only* p vertices. However, p vertices can be included with a vertices in a minimal W where $\dim(G_{AP})$ is determined by $\dim(a \times a)$ and $\dim(a \times p)$

Proposition 9.5. *Suppose that a G_{AP} has authors set A and representative papers set P where $G_{AP} \not\cong P_3$ and G_{AP} is not a subdivided K_n. If $\Delta(A) > \Delta(P)$ then $\dim(a \times a) \leqslant \dim(p \times p)$; and if $\Delta(A) < \Delta(P)$ then $\dim(a \times a) > \dim(p \times p)$.*

Proof. Let $\Delta(A) = x$ and $\Delta(P) = y$. From Proposition 5.1, when $\Delta(A) \neq \Delta(P)$, $\Delta(A)$ generates at least a K_x subgraph in G_P and $\Delta(P)$ generates at least a K_y subgraph in G_A. Corollary 5.3 states that the K_n subgraph generated by the smaller maximum partite set degree cannot exceed the K_n subgraph generated by the other maximum partite set degree when G_{AP} is not a subdivided K_n.

When G_{AP} contains twins, other than the single zero in each $a \times a$ block row, the rows of twin a vertices contain identical distances that create blocks of identical column combinations where the other a vertices have columns indexed by the twins. Thus, when $x > y$, $\dim(a \times a)$ in this case can exceed the metric dimension of the K_y double subblock and reach equality with $\dim(p \times p)$ resulting in $\dim(a \times a) \leqslant \dim(p \times p)$. When $x < y$, because, p vertices cannot be twins, equality does not occur. The $p \times p$ block has at least a K_x double subblock while the $a \times a$ block has at least a K_y double subblock; so $\dim(a \times a) > \dim(p \times p)$. □

Because metric dimension is defined across all of G's vertices, there exists no "efficiency" that could generate $\dim(G_{AP}) < \min\{\dim(G_A), \dim(G_P)\}$. Nor is there any inefficiency that could produce $\dim(G_{AP}) > \max\{\dim(a \times a), \dim(a \times p)\}$ or $\dim(G_{AP}) > \max\{\dim(p \times p), \dim(p \times a)\}$. In other words, because $V(G_{AP})$ is the union of A and P and both $\max\{\dim(a \times a), \dim(a \times p)\}$ and $\max\{\dim(p \times p), \dim(p \times a)\}$ minimally resolve the sets A and P across all of $V(G_{AP})$ then $\dim(G_{AP}) \not> \max\{\dim(a \times a), \dim(a \times p)\}$ and $\dim(G_{AP}) \not> \max\{\dim(p \times p), \dim(p \times a)\}$. This proves Lemma 9.6.

Lemma 9.6. *Suppose bipartite G_{AP} has authors set A and representative papers set P. Concerning the DM for G_{AP}, if $\dim(a \times a) \neq \dim(p \times p)$ then $\dim(odd) \leqslant \max\{\dim(a \times a), \dim(p \times p)\}$.* □

Theorem 9.7 states that if $\dim(a \times a) \neq \dim(p \times p)$ then only three DM blocks need to be resolved.

Theorem 9.7. *For G_{AP} with authors set $a \in A$ and representative paper set $p \in P$, with respect to block resolving G_{AP}'s DM, if $\dim(a \times a) \neq \dim(p \times p)$ then resolving only three blocks determines $\dim(G_{AP})$.*

Proof. Let $\dim(a \times a) = x$ and $\dim(p \times p) = y$ where $x \neq y$. It is a given that finding G_{AP} requires determining $\dim(a \times a)$ and $\dim(p \times p)$. From Lemma 9.6 $\{\dim(a \times p), \dim(p \times a)\} \leqslant \max\{\dim(a \times a), \dim(p \times p)\}$. First let $x < y$. As the goal is to find a minimal value for $\dim(G_{AP})$, resolving $a \times p$ in this case gives the minimal value of a possible W for all vertices in G_{AP}. If $\dim(a \times p) \leqslant \dim(a \times a)$ then $\dim(G_{AP}) = \dim(a \times a)$; and if $\dim(a \times a) < \dim(a \times p) \leqslant \dim(p \times p)$ then $\dim(G_{AP}) = \dim(a \times p)$. If $a \times p$ contains identical rows making $\dim(p \times a) = 0$, and indicating the existence of twins in G_{AP}, then $\dim(G_{AP}) = \max\{\dim(a \times a), \dim(a \times p)\}$. Now let $x > y$. Because $p \times a$ is the extension of $p \times p$, and using the same logic as when $x < y$, $\dim(G_{AP}) = \max\{\dim(p \times p), \dim(p \times a)$. In either case for x and y, only three blocks need to be resolved in order to find $\dim(G_{AP})$. ☐

The following proposition follows from Proposition 9.3 that shows the existence of the K_n double subblocks.

Proposition 9.8. *Suppose G_{AP} has authors set A and representative papers set P. If $\dim(G_A) = \dim(G_P)$, then $\dim(G_{AP}) = \dim(G_A) = \dim(G_P)$.*

Proof. Suppose $\dim(G_A) = \dim(G_P) = x$ so $\dim(even) = x$. Let $x < y$. If $\dim(odd) = x$ then it is a given that $\dim(G_{AP}) = x$. If $\dim(a \times p) = x$ and $\dim(p \times a) = y$, then $\dim(G_{AP}) = x$ because x number of a vertices minimally resolve the graph. It follows that if $\dim(a \times p) = y$ and $\dim(p \times a) = x$ then G_{AP} is minimally resolved by x number of p vertices so $\dim(G_{AP}) = x$. If $\dim(odd) = y$ then a combination of a and p vertices whose numbers total x can minimally resolve G_{AP} giving $\dim(G_{AP}) = x = \dim(G_A) = \dim(G_P)$. ☐

Proposition 9.9 is focused on using maximum degree and $r_\square$ to find $\dim(G_{AP})$. Recall that $R_{r_\square}$ indicates the degree diagram row indicated by the $r_\square$ type and $\ell(R_{r_\square})$ denotes the length of this row. The table in Figure 1 is referenced in the following proof of Proposition 9.9.

Proposition 9.9. *Assume that G_{AP} has authors set A and representative papers set P. If $\Delta(A) = \Delta(P)$, $r_\square(A) = r_\square(P)$ and the $r_\square$ corner type for both A and P is the same, then $\dim(G_{AP}) = \dim(G_A) = \dim(G_P)$.*

Proof. Assume $\Delta(A) = \Delta(P) = x$. Let $G_{AP} \cong P_n$ where $n = |A| + |P|$ and n is odd. Because $\Delta(A) = \Delta(P)$ and $r_\square(A) = r_\square(P)$ with the same corner type, then $|A| \geqslant 5$ and $|P| \geqslant 4$, $\Delta(A) = \Delta(P) = 2$, $r^r_\square(A) = r^r_\square(P) = 2$ and it is a given that $\dim(G_{AP}) = \dim(G_A) = \dim(G_P) = 1$.

Recall that for any G_{AP}, $\sum \deg(a) = \sum \deg(p)$. Because $\Delta(A) = \Delta(P) = x$, the degree diagrams for both A and P have the same top row. Because $r_\square(A) = r_\square(P)$ and both $r_\square$ have the same corner type, then for both diagrams, the rows above and below $R_{r_\square}$ plus row $R_{r_\square}$ have the same general structure as shown by the table in Figure 1. Deviation from similarity is controlled by the fact that $\sum \deg(a) = \sum \deg(p)$. Thus, A and P must have very similar or identical large degree structures. From Proposition 5.1 there exists a K_x subgraph in the projection graphs. The similarity of the large degree structures forces the degree structures of the projection graphs to deviate from the K_x subgraph structure in similar ways. Hence, if $\Delta(A) = \Delta(P)$ and $r_\square(A) = r_\square(P)$ and the $r_\square$ corner type for both A and P is the same, then $\dim(G_{AP}) = \dim(G_A) = \dim(G_P)$. ☐

10. Possible Future Work

10.1. Social Aspects

Regarding the departments in this study, when overall papers are considered, the physics departments generate 3.5 times as many papers as either the mathematics or biology departments. Compared to the average biology and average mathematics G_{AP}, the average physics G_{AP} in Figure 8 displays large author vertex degree reflecting greater variety in the combinations of research groups based on *representative* papers. For two of the actual physics departments, this study revealed that the representative papers outnumbered authors by 150% indicating a greater variety in research group author inclusion. Could the physics style of departmental collaboration result in a higher paper output?

Compared to the results in [24], the percent of vertices in this note's G_A large components is much less. At what investigation level, does the order of the large component get close to the range 60% to 90+% inclusion found in other studies? Would exploring institutional collaboration create the large components with higher percents of inclusion?

10.2. Network Analysis Aspects

Most collaboration studies focus exclusively on the G_A. As previously recognized in other studies, excluding papers from a collaboration network analysis results in information loss. The fact that different hubs are identified between some of this study's G_A compared to the related G_{AP}, and the fact that the interpretation of the large degree structure in G_a is not well defined, emphasizes the need for paper inclusion since the G_{AP} is the actual social network. These aspects provide good foundation for Conjecture 6.1 that states determining the predicting model for collaboration network evolution, and new edge formation, requires the study of both authors and papers. Foundation for the conjecture is also found by the metric dimension analysis in this note showing that $\dim(G_A)$ alone cannot be reliably used to predict $\dim(G_{AP})$ where G_{AP} is the actual social network.

The challenges faced with paper inclusion in the collaboration studies based on the large databases are significant. As done in this study, utilizing representative papers reduces the number of papers; but the large number of authors on many papers poses a potential problem in defining the representative papers. Can a co-authorship "core" for papers be identified? Is there some other strategy that allows for paper inclusion without hindering collaboration network analysis due to enormous network order and size?

DM block resolving provides an accurate and more computationally efficient method for finding $\dim(G)$ that clearly works for bipartite graphs. However, many networks are not bipartite. Can a method be developed, perhaps by a labeling methodology, that creates clear blocks in any graph's DM?

References

1. T.R. Anderson, R.K.S. Hankin, P.D. Killworth (2008) Beyond the Durfee square: Enhancing the h-index to score total publication output. *Scientometrics* 76(3). 577-588.
2. G.E. Andrews, K. Eriksson (2004) *Integer Partitions*. Cambridge Univ. press, Cambridge,UK.
3. A. Anuradha, B. Amutha (2019) A study on metric dimension of some faimiles of graphs. *AIP Conference Proceedings, 2019* 2112 (1).
4. M. Baca, E.T. Baskoro, A.N.M. Salman, S>W> Saputro, D. Suprijanto (2011) The metric dimension of regular bipartite graphs. *Bull. Math. Soc. Sco. math. Roumanie Tome* 54(102). No.1. 15-28.
5. C.G. Blanchard, J.V. Becker, A.R. Bristow (1979) Attitudes of Southern Women: Selected Group Comparisons. *Psychology of Women Quarterly* 1(2). 160-171.
6. M.R. Carey, D.S. Johnson (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W. H. Freeman & Co., New York, NY.
7. D. Cartwright, F. Harary (1956) Structural balance: a generalization of Heider's Theory. *The Psychological Review* 63(5). 277-293.
8. G. Chartrand, L. Ehoh, M.A. Johnson, O.R. Oellermann (2000) Resolvability in graphs and the metric dimension of a graph. *Discrete Applied Mathematics* 105.. 99-113.
9. G. Chartrand, L. Lesniak, P. Zhang (2016) *Graphs & Digraphs, 6th Ed*. CRC Press, Boca Raton, FL.
10. C.J.Cummins, R.C. King (1987) Young diagrams, supercharacters of OSp(M/N) and modification rules. *Jrnl. Phys. A: Math. Gen.* 20. 3103-3120.
11. J. Diaz, O. Pottohen, M. Serna, E.J. van Leeuwen (2017) Complexity of metric dimension on planar graphs. *Jrnl. Computer and System Sciences*. 83. 132-158.
12. R. Diestel (2017) *Graph Theory, 5th Ed*. Springer: Graduate Texts in Mathematics series, Berlin, Germany.
13. F. Harary, R.Z. Norman (1953) Graph theory as a mathematical model in social science. *Bull. de Institut de recherches économiques et sociales* (1960) 26(8).
14. F. Harary, R.A. Melter (1976). On the metric dimension of a graph, *Ars Combinatoria* 2. 191–195.

15. s. Hartung, A. Nichterlein (2013) ON the parameterized and approximation hardness of metric dimension. *2013 IEEE Conference on Computational Complexity* Stanford Univ. (USA). 266–276.
16. T. Heinz, P. Shapira, J.D. Rogers, J.M. Senker (2009) Organizational and institutional influences on creativity in scientific research. *Research Policy* 38. 610-623.
17. D. Janežič, A. Miličević, S. Nikolić, N. Trinajstić (2015) *Graph-Theoretical Matrices in Chemistry, 2nd Ed*. CRC Press, Taylor & Francis Group, Boca Raton, FL.
18. S. Kyvik, I. Reyert (2017) Research collaboration in groups and networks: differences across academic fields. *Scientometrics* 113. 951-967.
19. L. Lovász (2010) *Graphs and geometry*. American math. Society. Volume 65.
20. R. Merris (2003) *Combinatorics, 2nd Ed*. Wiley Interscience, John Wiley & Sons, Inc., Hoboken, NJ.
21. R. Merris (2001) *Graph Theory*. Wiley Interscience, John Wiley & Sons, Inc.,Hoboken, NJ.
22. R.K. Merton (1968) The Matthew effect in science. *Science*. New Series AAAS. 159(3810) 56-63.
23. S. Milgram (1967) The small-world problem. *psychology Today* 1(1, May) 61-67.
24. M. Newman (2001) The structure of scientific collaborations networks. *PNAS* 98(no. 2). 404-409.
25. M. Girvan, M.E.J. Newman (2002) Community structure in social and biological networks. *PNAS-06* 99(12). 7821-7826.
26. M. Newman (2004) Coauthorship networks and patterns of scientific collaboration. *PNAS* 101(1). 5200-5205.
27. M.E.J, Newman (2006) Power laws, Pareto distributions and Zipf's law. *arXiv:cond-mat/0412004v3 [cond-mat.stat-mech]*.
28. M. Newman (2018) *Networks, 2nd Ed*. Oxford Press, Oxford, England.
29. D.J. de Solla Price (1965) Networks of scientific papers: the pattern of bibliographic references indicates the nature of the scientific front. *Science* (American Association for the Advancement of Science) 149(3683) 510-515.
30. J. Ruscio, F. Seaman, C. D'Oriano, E. Stremlo, K. Mahalchik (2012) Measuring scholarly impact using modern citation-based indices. *Measurement* 10. 123-146.
31. I. Schriba, S. Farrugia (2011) On the spectrum of threshold graphs. *ISRN Discrete Math.* 2011. 21 pages.
32. P.J. Slater (1975). Leaves of trees (Proc. 6th Southeastern Conference on Combinatorics, Graph Theory, and Computing, Florida Atlantic Univ., Boca Raton, Fla.) *Congressus Numerantium 14, Winnipeg: Utilitas Math.* pp. 549–559
33. R.C. Tillquist, R.M. Frongillo, M.E. Lladser (2923) Getting the lay of the land in discrete space: a survey of metric dimension and its applications. *SIAM Review* 65(4). 919-962.
34. N. Trinajstić (1992) *Chemical Graph Theory*. Taylor and Francis, LLC; Boca Raton, FL.
35. F.J. van Rijnsoever, L.K. hessels, R>I.J. Vandeberg (2008) A resource-based view on the interactions of university researchers. *Research Policy* 37.1255-1266.
36. C.S. Wagner, L. Leydesdorff (2005) Network structure, self-organization, and the growth of international collaboration in science. *Research Policy* 34. 1608-1618
37. J. Wang, F. Tian,Y. Liu, J. Pang, L. Miao (2023) On graphs of order n with metric dimension n-4. *Graphs and Combinatorics* 39 (29) 1-18.
38. D.J. Watts (2003) *Six degrees: the science of a connected age*. W.W. Norton & Co., NYC, NY.
39. D.J. Watts, S.H. Strogatz (1998) Collective dynamics of the 'small-world' networks. *nature* 3393. 440-442.