2019

Using Social Network Analysis to Examine the Connections within a Noyce Community’s Facebook Group

Amanda Jensen
South Dakota State University

USING SOCIAL NETWORK ANALYSIS TO EXAMINE THE CONNECTIONS WITHIN A NOYCE COMMUNITY'S FACEBOOK GROUP

BY

AMANDA JENSEN

A thesis submitted in partial fulfillment of the requirements for the

Master of Science

Major in Mathematics

South Dakota State University

2019

USING SOCIAL NETWORK ANALYSIS TO EXAMINE THE CONNECTIONS WITHIN A NOYCE COMMUNITY'S FACEBOOK GROUP

AMANDA JENSEN

This thesis is approved as a creditable and independent investigation by a candidate for the Master of Science in Mathematics degree and is acceptable for meeting the thesis requirements for this degree. Acceptance of this thesis does not imply that the conclusions reached by the candidate are necessarily the conclusions of the major department.

Sharon Vestal, Ph.D.

Thesis Advisor Date

Kurt Cogswell, Ph.D.

Head, Department of Mathematics Date

Dean, Graduate School Date

This is for “Uncle Lon” I wish you were here to see this, you would have loved it.

ACKNOWLEDGMENTS

This thesis project was funded by NSF DUE-1439789.

TABLE OF CONTENTS

ABBREVIATIONS

Early Career Teacher (ECT)

National Science Foundation (NSF)

Rural Enhancement of Math And Science Teachers (REMAST)

South Dakota State University (SDSU)

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

USING SOCIAL NETWORK ANALYSIS TO EXAMINE THE CONNECTIONS WITHIN A NOYCE COMMUNITY'S FACEBOOK GROUP

AMANDA JENSEN

2019

One of the successes of the Rural Enhancement of Mathematics And Science Teachers (REMAST) Scholarship Program at South Dakota State University is the community we have built. This community has been built through a summer conference and a closed Facebook group. As we near the end of our Phase II Noyce funding, we are using social network analysis to examine the connections within the REMAST Facebook group. What we learn in this research project will be useful to other Noyce projects as it is a model for developing a strong professional learning community.

In order to determine information about the connections within the group of alumni and scholars, we created adjacency matrices based on cohort number, the number of semesters of funding, and subject area. From these adjacency matrices, we used the software programs Gephi and MATLAB to generate data and to create images that illustrate friendships between group members. Looking at the images, we made conjectures about the Facebook group and analyzed these conjectures using various attributes, such as degrees, eigenvector centralities, and triangles, of the data.

CHAPTER 1

BACKGROUND

This project uses the basics of graph theory and applies it to a social network graph. This process is commonly known as social network analysis. Mathematics is used to look at the significance of some of the phenomena that can happen in social networks. The following definitions will be used throughout the thesis. A **graph** is a way of specifying relationships among a collection of nodes. A **node** is an object, the set of nodes in a graph is denoted as *V*. An **edge** is a link that connects a pair of nodes. A **directed graph** consists of a set of nodes together with a set of directed edges where the direction of the edge is important. The graphs we focus on in this project are **undirected**, meaning there is no direction on the edges.

A graph is **connected** if for every pair of nodes, there is a path between them. A **path** is a sequence of nodes such that each consecutive pair in the sequence is connected by an edge. In a **social network**, the nodes are people and the edges represent a social interaction between the people. The **degree** of node *m* is the number of edges associated with *m*, d_m. When we say that two nodes are **adjacent**, we mean that they are connected with an edge. An **adjacency matrix** is a matrix, *A*, such that

$$A = [a_{mn}], where\ a_{mn} = \begin{cases} 1 \text{ if node } m \text{ is adjacent to node } n \\ 0 \text{ otherwise} \end{cases}.$$

A **triangle** in a graph occurs when three nodes and three edges form a triangle.

The **clustering coefficient** of a node, *m*, is the probability that two random friends of *m* are friends with each other. The formula for the clustering coefficient for node m is:

$$C_m = \frac{\text{the number of edges that exist between node } m'\text{s friends}}{\binom{d_m}{2}}.$$

The **betweenness centrality** for node *m* is the sum of the number of shortest paths from *k* to *n* that go through *m*, $p_{kn}(m)$, divided by the number of shortest paths from *k* to *n*, p_{kn}:

$$C_B = \sum_{k \neq m \neq n \in V} \frac{p_{kn}(m)}{p_{kn}}.$$

Betweenness can also be described as the extent that other nodes depend on *m* as a transmitter of information. **Eigenvector centrality** describes the influence of a node in the network, the calculation will be discussed in more detail in Chapter 3.

The graph in this project is the Facebook group made up of students, alumni, and faculty who have received funding or are involved with the National Science Foundation (NSF) Robert Noyce Scholarship Program at South Dakota State University (SDSU). At SDSU, this program is named the Rural Enhancement of Math And Science Teachers (REMAST) Scholarship Program. REMAST scholarships are available to students entering their junior or senior year who are pursuing secondary certification in Biology, Chemistry, Mathematics, or Physics. Students who receive REMAST are required to teach in a high needs school district one year for each semester they received the scholarship. A high needs school district is a district that meets at least one of the following criteria: a) a high percentage of individuals from families with incomes below the poverty line; b) a high percentage of secondary school teachers not teaching in the content area in which they were trained to teach; or c) a high teacher turnover rate [12].

The first NSF grant was awarded in September 2007 and the first scholarships were given out in the Fall 2008 semester. There were no scholarships awarded in the 2014-2015 school year as we were between NSF awards. When our Phase II NSF grant was funded, it included funding for research for this project. We began awarding Phase II scholarships in the Fall 2015 semester. There have been nine cohorts of scholarship recipients since REMAST began in 2008. We will focus on the first eight because the ninth cohort is in their first year of funding so they have not had as much interaction with the REMAST group.

The REMAST Facebook group was started during the Fall 2009 semester. The purpose of the group was to create a space for students and faculty to share joys, concerns, ideas, and seek support throughout their teaching career. The group has 65 members: three co-PIs, one former faculty, one local teacher, one person from the South Dakota Department of Education, and 59 scholarship recipients. All members are encouraged to post in the group. The focus we had for this project was to examine the Facebook friendships between the members of the group and see what connections exist. For all the social network data analysis we removed the following people from the Facebook group roster: those not directly involved with REMAST, the four members of Cohort 9, two recipients who are no longer teaching and are not active in the group, and the three co-PIs. After these exclusions we were left with 53 scholars and alumni that were included in our social network graphs.

The motivation behind this project stems from the resilience of the teachers that received REMAST during their pre-service education. There have been 56 REMAST students who graduated with teaching certification in math or science, three either graduated without certification or did not complete a degree. Of the 56 students, 43 are teaching full-time or involved in some aspect of education during the 2018-2019 school year. There are three students currently enrolled full-time in graduate programs. We have several who have completed Master's degrees while teaching full-time and some are currently pursuing them while teaching full-time. Table 1 shows the number of REMAST alumni that have completed graduate degrees or are seeking graduate degrees. The year in the first column is the school year that students in each cohort first received funding.

TABLE 1: Illustrates how many recipients have completed or are seeking graduate degrees in each cohort.

Cohort (year)	**Number of People**	**Number who completed Master's degree**	**Number seeking graduate degree**
1 (2008-2009)	11	6	1
2 (2009-2010)	5	2	0
3 (2010-2011)	7	3	0
4 (2011-2012)	12	2	2
5 (2012-2013)	8	0	4
6 (2015-2016)	8	0	1
7 (2016-2017)	3	0	0
8 (2017-2018)	2	0	0

The REMAST program has been a very successful Noyce program. One of the reasons for this success is the tight-knit community that we have created. This community consists of the cohorts of students who have received the scholarship, the faculty members involved, and the Facebook group. Our main goal in this thesis is to see if we can mathematically measure the connections within the Facebook group.

CHAPTER 2

CONJECTURES

The Facebook friends data was organized by three different characteristics: cohort number, number of semesters of funding received, and content major. We chose to view the data by these characteristics because we wanted to see if people who had common characteristics were more connected with each other. In this chapter we discuss conjectures that were made based on social network graphs and the image of the adjacency matrix for the REMAST Facebook group.

Facebook is a social media site structured such that two people can only be friends if both people authorize it. Because of this built-in structure of Facebook, our social network graphs are all undirected. If the social media we were focusing on was Twitter or Instagram this would not necessarily be the case. The social network graphs representing Twitter or Instagram would be directed graphs. The 53 people in the group were given an ID number and then were also identified by cohort number, semesters of funding, and content area.

The data was organized as an adjacency matrix which consists of zeros and ones. If a 1 is present in row *m* and column *n*, then person *m* and person *n* are Facebook friends; if there is a 0 then person *m* and person *n* are not Facebook friends. An important characteristic of an adjacency matrix of an undirected graph is that the matrix, by construction, is symmetric. A **symmetric matrix** is a square matrix such that $A = A^T$.

The adjacency matrix was used in creating various images. We used MATLAB to create a color-coded image of the matrices and we used Gephi to create a social network graph. MATLAB is a software that is used to analyze data, develop algorithms, and create models and applications [11]. It is matrix-based which is the main reason we chose to use it to perform calculations and create images. Gephi is a software used to visualize and analyze networks. It can be used to find patterns or trends in the networks [5].

From these images we made conjectures about what we noticed. For each pairing of cohorts or semesters of funding, the corresponding adjacency matrix was extracted from the original adjacency matrix with all 53 members. Prior to creating a graph in Gephi, the adjacency matrix was run through an R program that extracted a node list and an edge spreadsheet. The node list spreadsheet contained the group members with their given ID number, their content major, their cohort number, and the number of semesters of funding. The rows in the edge spreadsheet listed all pairs of Facebook friends with a weight of 1. The R code was originally written for a directed graph by Ian Morton, former graduate research assistant for the REMAST program. We modified the code to fit the needs of an undirected graph since that is what we are working with.

COHORTS

We will first look at the images made from the adjacency matrix when it is sorted by cohort. There were eight cohorts in our data. Recall that Cohort 9 is this year’s group of scholars and we are excluding them from the analysis. The number of people in the cohorts varied from year to year, which is shown in Table 2. The colors in the third column correspond to node colors for each cohort in Figure 2 (the social network graph).

TABLE 2: This shows how many people are in each cohort and the color of their nodes in Figure 2.

Cohort	**Number of People**	**Node Color**
1	11	Blue
2	5	Green
3	7	Orange
4	9	Lime Green
5	8	Red
6	8	Light Blue
7	3	Pink
8	2	Purple

A color-coded image of the adjacency matrix, A_C, can be seen in Figure 1. The zeros are black, the ones are red, and yellow is the diagonal. The diagonal is not significant but is used as a visual aid to help see the symmetry in the matrix. If the matrix was folded along the diagonal, the part above the diagonal would land exactly on top of the portion below the diagonal.

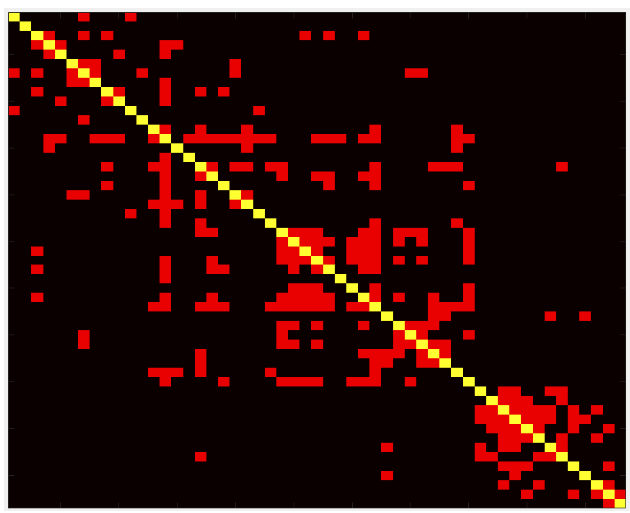


FIGURE 1: Matrix image of A_C

Figure 2 is the social network graph showing the friendships between all members of the REMAST Facebook group. This graph was made from A_C, the nodes are scholarship recipients in the Facebook group and the edges represent Facebook friendships between the recipients. Figures 1 and 2 are both mathematical representations of A_C. We have included both because different characteristics are more apparent in each image.

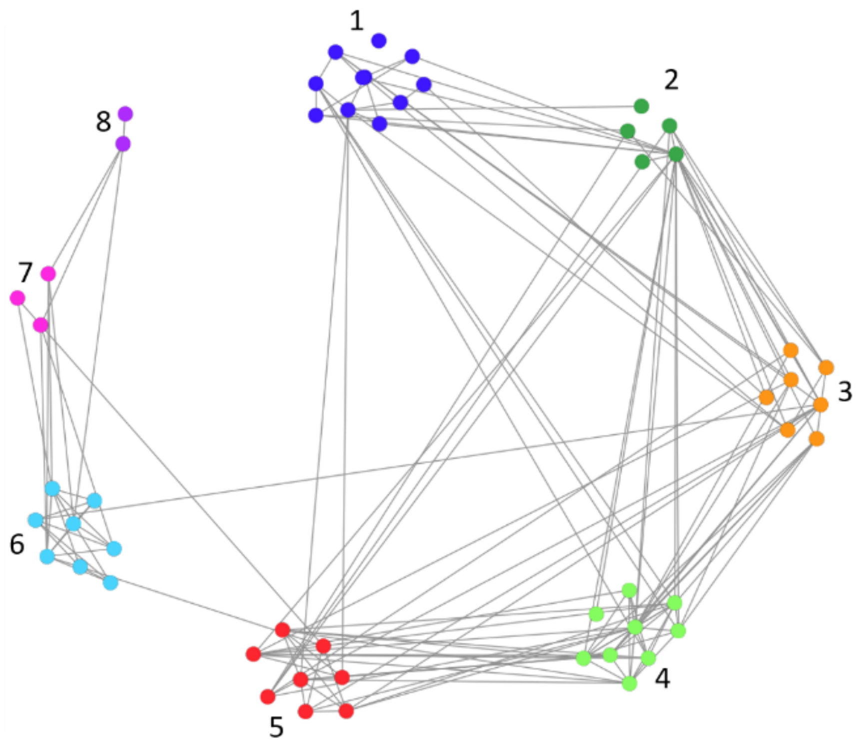


FIGURE 2: Social network graph of A_C

Figure 3 was obtained by squaring A_C in MATLAB and then creating the image from the matrix. We call this matrix $M_C = [m_{mn}]$ because it is a Mutual Friends Matrix. A **Mutual Friends Matrix** is a matrix where the diagonal entries of the matrix correspond to the number of Facebook friends that node has within the group—this number is also the degree of that node in the entire network. The entry in row *m* column *n* represents the number of mutual Facebook friends between person *m* and person *n*. The symmetry of A_C allows us to interpret M_C this way, if we were working with directed graphs M_C would not have the same meaning. The colorbar to the right of the MATLAB image illustrates which values each color represents in the cells in M_C. If a zero is present off the diagonal of the matrix, then those two people do not have any mutual Facebook friends and the color of that entry is black. According to the colorbar from MATLAB, the brighter the color the more mutual friends exist between the two corresponding people.

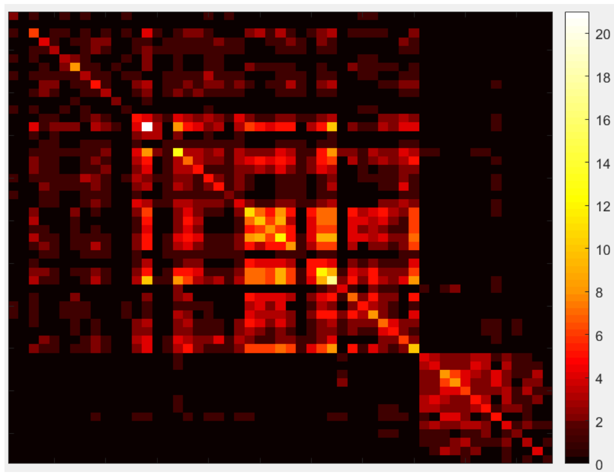


FIGURE 3: Matrix image of M_C, the colorbar on the right shows the value of each cell.

Conjecture 1: *Cohorts 4 and 5 are the most connected in the network.*

These connections are visible in Figures 1, 2, and 3. We will discuss what we noticed in Figures 2 and 3. In Figure 2 there are more edges between the lime green nodes, Cohort 4, and the red nodes, Cohort 5 than the other pairs of cohorts. As a reminder, each edge represents a Facebook friendship between two people. In Figure 3 there are a lot of bright colors in the region of the matrix that contains Cohorts 4 and 5. These bright colors indicate there are large numbers of mutual friends between the members of these two cohorts. We will examine this conjecture in Chapter 3 by comparing the number of degrees, the distribution of the eigenvector centralities, and the number of triangles present between Cohorts 4 and 5.

Conjecture 2: *There exists a distinct gap between the recipients in the Phase I and Phase II REMAST grants.*

The members of Cohorts 1 through 5 were funded under the first NSF grant, which we refer to as Phase I. The last three cohorts were funded by the second NSF grant, Phase II. Notice in Figure 1, the matrix image of A_C, there are large black regions on the right and bottom sides of the image. Within these large regions of black, we see three lone red squares, which represent the three friendships that exist between recipients in Phase I and Phase II. These three friendships are easy to see in Figure 2 as they are the only three edges connecting the left side of the network to the right side of the network. We will provide further evidence of this conjecture in Chapter 3.

Conjecture 3: *Cohort 1 has the most connections throughout the Phase I grant.*

Notice there is a lot of dull red spread across the top of Figure 3—this represents the number of mutual friends Cohort 1 has with the other cohorts in Phase I. It is also visible in Figure 2 because there are edges connecting Cohort 1 to each of the other cohorts in Phase I. We will investigate in Chapter 3 by comparing the total number of degrees for each cohort in Phase I.

Conjecture 4: *There is a clear separation between cohorts in the mutual friends matrix.*

Looking at Figure 3, the colors are brighter closer to the diagonal of M_C; this means that people who are in the same cohort have a lot of the same friends. We will investigate this in Chapter 3 by looking at the number of degrees within each cohort.

SEMESTERS

Now we will look at the images created after sorting the adjacency matrix by the number of semesters of funding of the recipients. Students who applied for the scholarship as soon as they were eligible could have it for a maximum of four semesters. Once selected, students received funding until they graduated unless they decided not to pursue teaching or there was a concern about the student completing the program. Table 3 shows how many students were in each category of funding from one semester to four semesters. The colors in the third column correspond to the colors of the nodes for each semester in the social network graph in Figure 5.

TABLE 3: This shows how many people received funding for the different numbers of semesters and the color of their nodes in Figure 5.

Semester	**Number of People**	**Node Color**
1	3	Pink
2	13	Lavender
3	11	Blue gray
4	26	Mint green

Figure 4 has the same values as the adjacency matrix in Figure 1, but the people (nodes) are sorted in ascending order, with the people who had the scholarship one semester at the top left and those having it four semesters in the bottom right. We will refer to this matrix as A_S since it is the adjacency matrix for the semester data. Figure 5 is a social network graph created with A_S with the different colors representing the number of semesters students received funding as seen in Table 3.

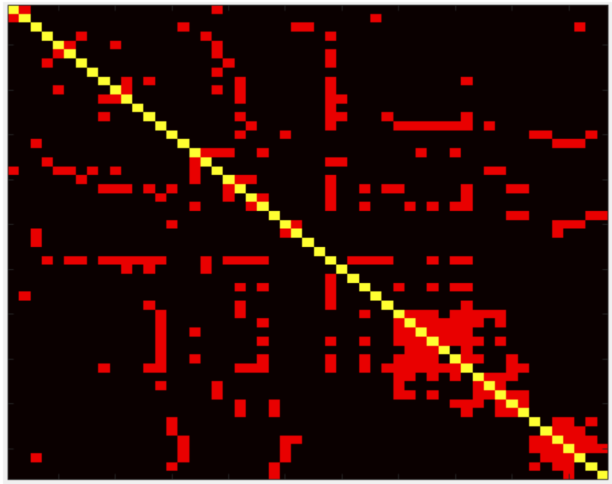


FIGURE 4: Matrix image of A_S

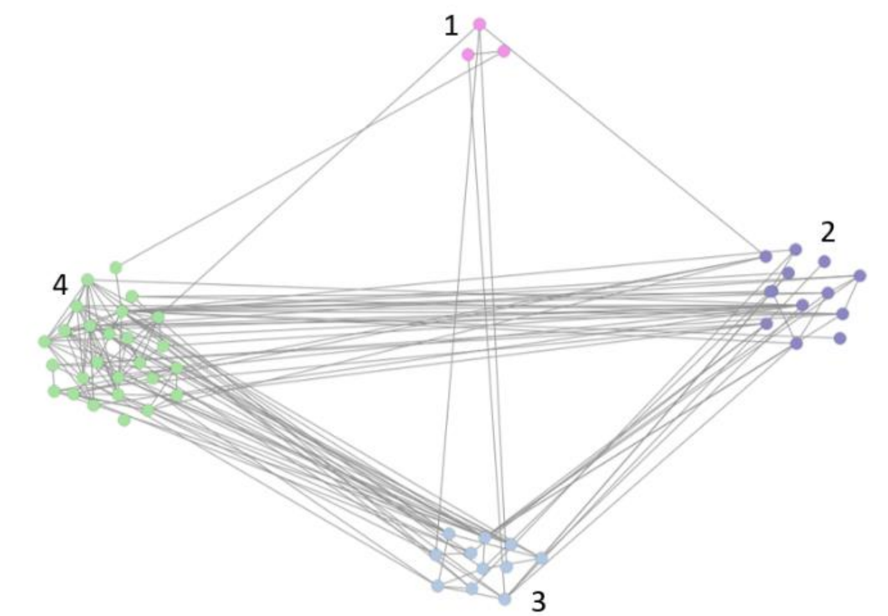


FIGURE 5: Social network graph of A_S

Figure 6 was obtained by squaring A_S and is a mutual friends matrix sorted by semester, M_S. There is some of the same structure present in Figure 3 and Figure 6 because the diagonal entries are the same values but in a different configuration.

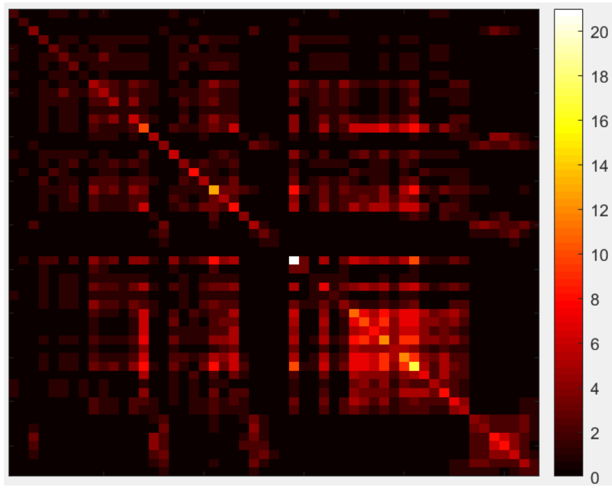


FIGURE 6: Matrix image of M_S, the colorbar on the right shows the value of each cell.

Conjecture 5: *The longer a person received funding the more connected they are within the entire group.*

Notice in Figure 5 how dense the lines are among the people who received funding for four semesters. It is also visible by the amount of lines going from two to four semesters and three to four semesters. Notice in Figure 6 there are brighter colors in the lower right region of the matrix image. Recall this is where the people who received funding for four semesters are located in the matrix image. Brighter colors are also present above and to the left of the lower region. This tells us that the people who received more funding have more mutual friends with the rest of the group. Again, we will study this conjecture in Chapter 3 by comparing the number of degrees between each pair of semesters, the distribution of the eigenvector centralities, and the number of triangles that exist between each pair of semesters.

SUBJECTS

We will briefly discuss what we observed when the adjacency matrix was sorted by the content major of the recipients. As a reminder, to be eligible for the REMAST scholarship a student had to be pursuing secondary certification in Biology, Chemistry, Mathematics, or Physics. Table 4 shows how many recipients were in each content area and their corresponding node color in Figure 7 (the social network graph). Notice that there are more mathematics majors than the other three content areas.

TABLE 4: This shows how many people are in each content area and the color of their nodes in Figure 7.

Content	**Number of People**	**Node Color**
Biology	20	Kelly green
Chemistry	6	Orange
Mathematics	26	Light blue
Physics	1	Purple

Figure 7 shows the social network graph when sorted by content major. We did not make any formal conjectures for this configuration of the network. We did observe that the mathematics majors are more connected than the others. We attribute this to the large amount of classes these students take together. The SDSU program for Secondary Certification specialization in Mathematics is unique because there are four mathematics courses specifially designed for and taken by students in this program. The biology and chemistry majors don’t necessarily have this same opportunity. There is a Science Methods course that biology, chemistry, and physics majors take together.

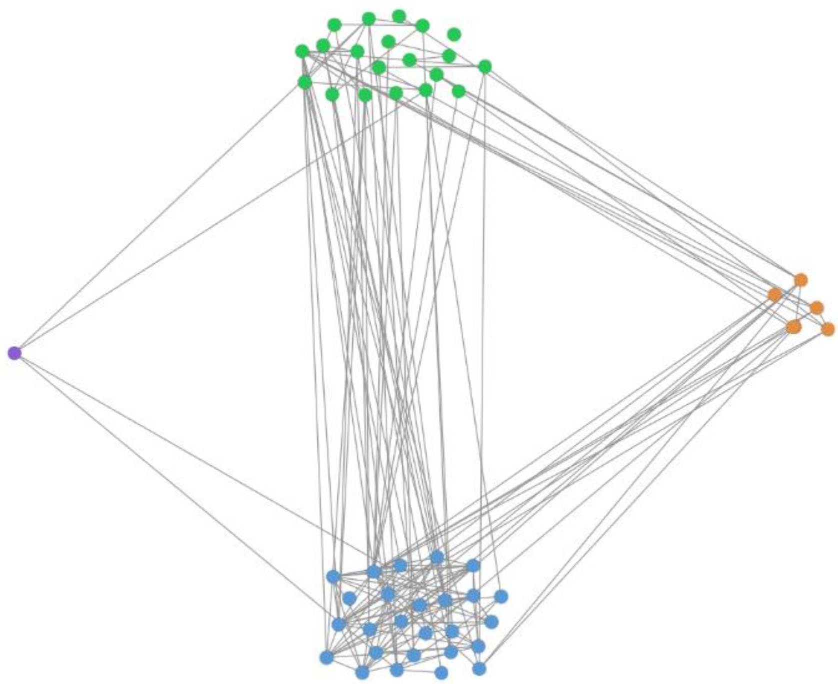


FIGURE 7: Social network graph of the content majors

CHAPTER 3

CONJECTURE RESULTS

Conjecture 1*: Cohorts 4 and 5 are the most connected in the network.*

As a reminder, this conjecture was made from the area in the white box in Figure 8. The area in the white box is the region of M_C where the people in Cohorts 4 and 5 are located. We noticed there were a lot of bright colors in this region, leading us to believe that these two cohorts are more connected than the others. We investigated this conjecture by comparing the number of degrees between pairs of cohorts, by looking at eigenvector centralities of certain pairs of cohorts, and by finding the proportion of triangles in cohorts and between cohorts.

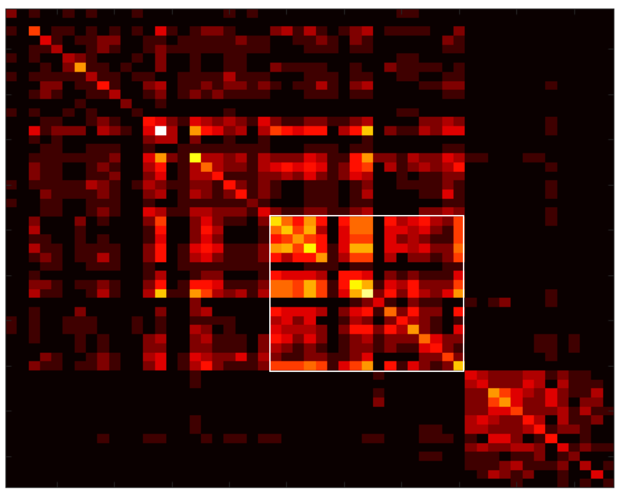


FIGURE 8: Matrix image of M_C with a white box around Cohorts 4 and 5.

Social network graphs have various measures of centrality: betweenness, clustering coefficient, degree, and eigenvector centrality. We chose not to use the clustering coefficient because it depends on the size of the group and the number of friends that they have in that group. Recall that the clustering coefficient is the probability that a person's friends are friends with each other. For example, one member of Cohort 3 had a clustering coefficient of 0 in Cohorts 3 and 5. This means that none of her friends were friends with each other in that pair of cohorts. However, the same member had a clustering coefficient of 1 in Cohorts 2 and 3, meaning all her friends were friends with each other in that pair of cohorts. Since the clustering coefficient can change drastically for individuals depending on the pair of cohorts being examined, we decided against using it to help verify this conjecture.

We also observed that a person's betweenness centrality becomes large when that person is part of a bridge between cohorts or is friends with a person who is part of the bridge. An edge between nodes *m* and *n* is a **bridge** if when it is deleted *m* and *n* lie in two different **components**, or connected parts of the graph. For example, in Figure 9 the edge that is circled is considered a bridge between Cohorts 5 and 6. If the circled edge was removed, then the light blue nodes would be their own component and the red nodes would be another component. The betweenness centrality of a node measures the extent that other nodes depend on it for information. Again, the betweenness centrality changes dramatically for a node, depending on the cohorts that are being paired. Therefore, we decided it would not be a good centrality to use to determine connectedness of cohorts.

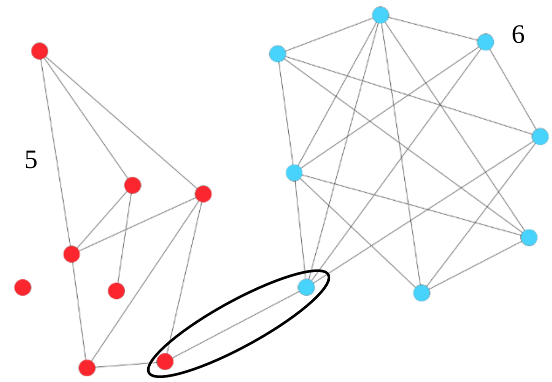


FIGURE 9: Social network graph of Cohorts 5 and 6. The circled edge is a bridge between the two cohorts.

In order to determine which pair of cohorts was most connected, one of the measures we used was the number of degrees within cohorts and between pairs of cohorts. The number of degrees between a pair of cohorts is the sum of the degrees of every node in both cohorts. The degree of a node is not affected by the size of the group or by certain relationships within the group so it is the most direct measure when talking about the connections in a group. Recall the degree of a node is the number of edges associated with that node. In addition, degree is easily understood and is easily captured in the adjacency matrix.

To conduct the degree analysis we created what we call a degree matrix, $D_C = [d_{ij}]$. The number in each cell, d_{ij}, of the degree matrix was taken directly from the data spreadsheet generated by Gephi. This number represents how many degrees exist between cohorts *i* and *j*. The diagonal entries represent how many degrees are present within a single cohort. Table 5 represents the degree matrix, D_C, with the cohort numbers in the first row and first column. If there is a zero in d_{ij}, then there were no friendships between cohorts *i* and *j* so we did not collect any data from Gephi for those pairs.

TABLE 5: Degree matrix D_C, each entry shows how many degrees are between each pair of cohorts.

Cohort	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	22	40	42	74	46	0	0	0
2	40	4	34	62	32	0	0	0
3	42	34	10	76	40	48	0	0
4	74	62	76	46	104	0	0	0
5	46	32	40	104	20	58	22	0
6	0	0	48	0	58	36	48	40
7	0	0	0	0	22	48	0	6
8	0	0	0	0	0	40	6	2

We then discovered that the information we got from D_C wasn't what we needed. In order to get more information from the degree data, we created an edge matrix, E_C, from D_C, using the following equation:

$$E_C = [e_{ij}], where\ E_C = \begin{cases} e_{ii} = \dfrac{d_{ii}}{2} \\ e_{ij} = \dfrac{d_{ij} - (d_{ii} + d_{jj})}{2} \end{cases} \quad (1)$$

We must divide by two in both calculations because each edge is counted twice in calculating the degree since each edge is associated with two nodes. It is straightforward to calculate e_{ii}. The calculation for e_{ij} is a little more involved. Since these entries are telling us how many edges exist between cohort *i* and cohort j, it includes the edges that are within cohorts *i* and *j* exclusively. In the formula you can see that these are removed by subtracting d_{ii} and d_{jj}. Each entry in E_C tells us how many edges exist between cohorts *i* and j. Table 6 shows E_C, with the cohort numbers in the first row and first column. Notice that if a zero was present in D_C it is still zero in E_C.

TABLE 6: Edge matrix E_C, each entry shows how many edges connect each pair of cohorts.

Cohort	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	11	7	5	3	2	0	0	0
2	7	2	10	6	4	0	0	0
3	5	10	5	10	5	1	0	0
4	3	6	10	23	19	0	0	0
5	2	4	5	19	10	1	1	0
6	0	0	1	0	1	18	6	1
7	0	0	0	0	1	6	0	2
8	0	0	0	0	0	1	2	1

There are 306 total degrees in the entire REMAST group, dividing by two gives us 153 edges. If we sum the lower triangular entries of E_C we do get 153. A **lower triangular matrix** is a matrix where all the entries above the main diagonal are zero:

$$L = \begin{bmatrix} a & 0 & 0 \\ b & c & 0 \\ d & e & f \end{bmatrix}.$$

We only need to sum the lower triangular entries because it is a symmetric matrix.

Notice that the largest diagonal entries in both D_C and E_C are d_{44} and e_{44} respectively. This tells us that Cohort 4 has the most connections within itself. When we look at the off-diagonal entries, notice that the largest numbers in both matrices are d_{45} and e_{45}. The value 23 for e_{44} tells us that there are 23 edges within Cohort 4; and e_{45} being 19 gives the number of edges (friendships) between Cohort 4 and Cohort 5. The last number that is important in this analysis is e_{55}, which is 10. Although it is not one of the larger numbers on the diagonal of E_C, it is important in this analysis because it gives us the number of friendships present within Cohort 5. These entries have been highlighted in yellow in Table 6. This highlighted portion represents the edge matrix for Cohorts 4 and 5. By summing the lower triangular portion of the edge matrix for Cohorts

4 and 5, we get that there are 52 friendships between and among Cohorts 4 and 5. The next largest number of friendships between a pair of cohorts is 38 for Cohorts 3 and 4.

We determined it would also be valuable to examine the relative frequency of edges that are present in the cohort pairs. To organize this information we created a relative frequency matrix, P_C, using Equation 2, where n_i corresponds to the number of people in cohort *i*. The values for n_i can be found in Table 2 from Chapter 1.

$$P_C = [p_{ij}] \text{, where } P_C = \begin{cases} p_{ii} = \dfrac{e_{ii}}{\binom{n_i}{2}} \\ p_{ij} = \dfrac{e_{ij}}{n_i * n_j} \end{cases}. \quad (2)$$

The relative frequencies shown in Table 7 give a better indication of the true connection between cohorts because the relative frequency takes into account the different sizes of the cohorts. For example, p_{88} has a relative frequency of 1 because there are two members in Cohort 8 and they are friends with each other on Facebook. Notice that the second largest relative frequency in P_C is 0.6389 located in p_{44}, meaning about 64% of the possible edges (friendships) exist within Cohort 4. The entries highlighted in yellow are the entries that correspond to Cohort 4, Cohort 5, and Cohort 4 with Cohort 5. Notice that these entries are some of the largest in the matrix. If we find the relative frequency for the edges in Cohorts 4 and 5 together, we do the following calculation:

$$\frac{52}{\binom{17}{2}} = 0.3824.$$

This calculation shows that about 38% of the possible edges in Cohorts 4 and 5 exist. We use the same calculation we would for a diagonal entry of the relative frequency matrix because we are treating Cohorts 4 and 5 as a single group in this calculation.

TABLE 7: Relative frequency matrix P_C, each entry shows the relative frequency of edges that are present in each pair of cohorts.

Cohort	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	.2	.1273	.0649	.0303	.0227	0	0	0
2	.1273	.2	.2857	.1333	.1	0	0	0
3	.0649	.2857	.2381	.1587	.0893	.0179	0	0
4	.0303	.1333	.1587	.6389	.2639	0	0	0
5	.0227	.1	.0893	.2639	.3571	.0156	.0417	0
6	0	0	.0179	0	.0156	.6429	.25	.0625
7	0	0	0	0	.0417	.25	0	.3333
8	0	0	0	0	0	.0625	.3333	1

Now we will compare the eigenvector centralities between five pairs of cohorts. We chose to focus on the eigenvector centralities because it describes the importance of a specific node in a network, taking into consideration the connections the node has throughout the network. We also chose this centrality because it can be used as a valid node centrality when normalized using the Euclidean norm and multiplied by $\sqrt{2}$ [13]. The **Euclidean norm** is

$$\|\boldsymbol{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}.$$

The fact that the eigenvector centrality is a node centrality means that the only way a node should have a centrality measure of 1 is if it is the center of a star shaped graph, shown in Figure 10. None of our social network graphs have that structure so it is important that we normalize this centrality measure properly.

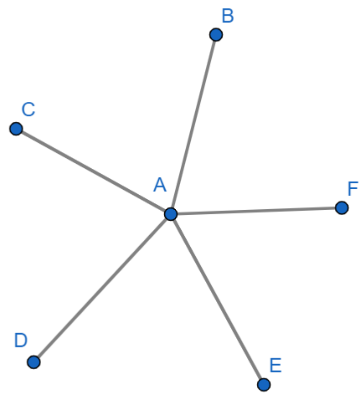


FIGURE 10: Example of a star graph

A MATLAB program was written and used to find and normalize the eigenvector centralities, shown in Figure 11. An adjacency matrix was imported into MATLAB. Then the eigenvalues and corresponding eigenvectors were found and stored in matrices. The program then determines the maximum eigenvalue's position and takes the absolute value of the entries of the eigenvector corresponding to the maximum eigenvalue. This makes the eigenvector non-negative. The eigenvector is then normalized with the Euclidean norm and then multiplied by $\sqrt{2}$. The resulting eigenvector contains the eigenvector centralities for the nodes of the social network graph being analyzed [2]. The program Gephi also calculated the eigenvector centralities but we were unable to find how the centrality measure was normalized. Therefore, we chose to use the eigenvector centralities generated by the MATLAB program as we know it is a valid calculation.

```
[b, c] = eig(a); %finds the eigenvectors, b, and the eigenvalues, c, of
    %adjacency matrix a
[val,pos]=max(max(abs(c))); %finds the position of the maximum eigenvalue

max_eig = abs(b(:,pos)); %takes the absolute value of the eigenvector
   %associated with the maximum eigenvalue
max_eig = max_eig/norm(max_eig,2); %normalizes according to the Euclidean
    %norm

nodecen= sqrt(2)*max_eig; %makes it a node-centrality
```

FIGURE 11: MATLAB code used to calculate eigenvector centralities [2].

To analyze the eigenvector centralities we used Desmos, a free online graphing calculator created “to help every student learn math and love learning math” [3]. We made boxplots of the eigenvector centralities for the following pairs of cohorts: 2 and 3, 2 and 4, 3 and 4, 3 and 5, and 4 and 5. These pairs of cohorts were chosen because they had the largest average eigenvector centralities. The boxplots are shown in Figure 12. The five-number summary for the eigenvector centralities is in Table 8. The rows in Table 8 are color coded to correspond with the color of the boxplot in Figure 12 (i.e. the red row in the table corresponds to the red boxplot). We do not include the range because each cohort pair’s minimum value was zero, which means the range is the same as the maximum value. We chose to use boxplots instead of a statistical test to analyze the eigenvector centralities because our data is not independent. Inferential statistics relies on the data being independent to get valid results, and since network data is not independent “false positives” or “false negatives” are often given [7]. We created stacked boxplots to compare the data for these five pairs.

TABLE 8: Five-number summary from the boxplots in Figure 12. The colors of the row matches the color of the boxplot.

Cohort Pair	Min	1st Quartile	Median	3rd Quartile	Max	Inter-quartile Range	Mean	# of people
2 and 3	0	0.1712	0.3205	0.50145	0.7467	0.33025	0.3442	12
2 and 4	0	0.0573	0.38205	0.5129	0.5827	0.4556	0.3142	14
3 and 4	0	0.0844	0.34535	0.47165	0.5937	0.38725	0.2923	16
3 and 5	0	0.1705	0.2337	0.5076	0.657	0.3371	0.3035	15
4 and 5	0	0.167	0.3059	0.4649	0.5063	0.2979	0.3013	17

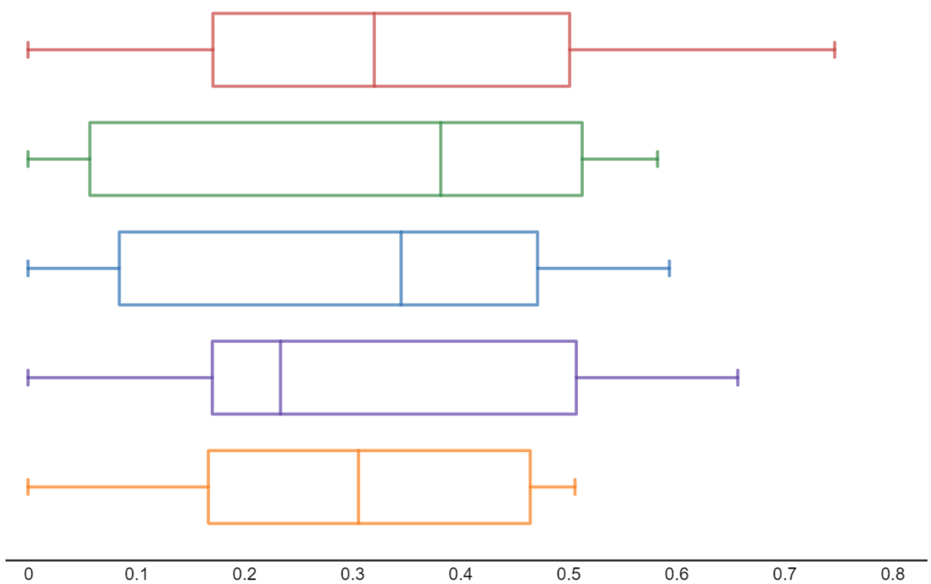


FIGURE 12: Stacked boxplots of the eigenvector centralities from five pairs of cohorts.

We will focus on the inter-quartile range and the maximum for each pair of cohorts. Both of these values are lowest for Cohorts 4 and 5, the orange row in Table 8 and orange boxplot in Figure 12. The next smallest inter-quartile ranges were, 0.33025 and 0.3371, from Cohorts 2 and 3 and Cohorts 3 and 5 respectively. These two pairs happen to have the two highest maximums as well, so overall their eigenvector centralities are more spread out, as seen in the boxplots. The smaller inter-quartile range for Cohorts 4 and 5 tells us that these centrality measures are closer together in this cohort pairing.

To further help visualize what is happening, the social network graphs of these three pairings are shown in Figures 13-15. There are a lot more edges in Figure 15 than there are in Figure 13 or Figure 14. There are also fewer nodes in Figure 15 that have zero, one, or two edges associated with them. Notice in Figure 13 how the node that appears to be in the center of the triangle has many edges coming out of it. This node is very significant in this group and increases the eigenvector centrality of the nodes who share an edge with it.

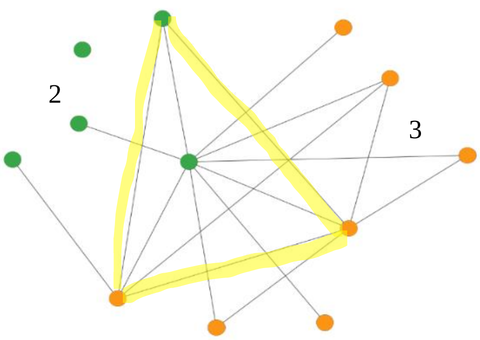


FIGURE 13: Social network graph of Cohorts 2 and 3

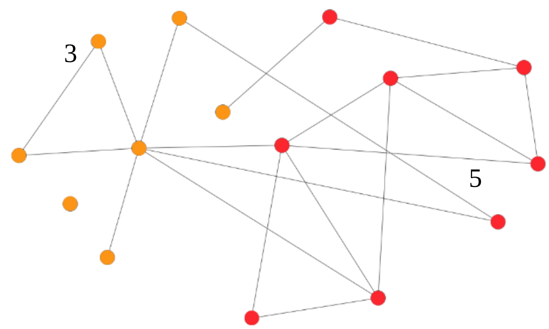


FIGURE 14: Social network graph of Cohorts 3 and 5

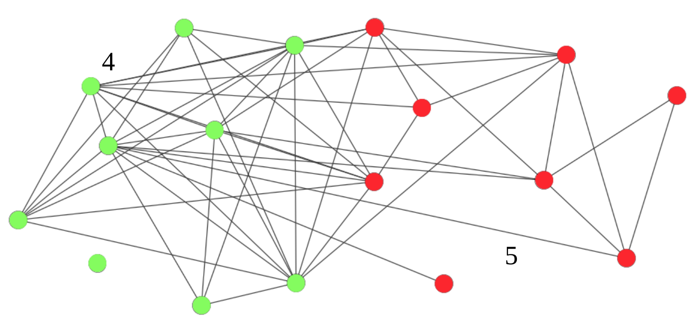


FIGURE 15: Social network graph of Cohorts 4 and 5

To provide additional evidence for Conjecture 1, we looked at the number of triangles that were present in all pairs of cohorts and the individual cohorts. As a reminder, a triangle occurs in a graph when three nodes and three edges form a triangle. An example of this is highlighted in Figure 13. To conduct this analysis, we created a triangle matrix, $T_C = [t_{ij}]$, shown in Table 9. The data for the number of triangles in each pair was generated by Gephi. Each entry, t_{ij}, tells us how many triangles exist in the graph of cohorts *i* and *j*. Triangles are easiest to see in the social network graphs because it is a relationship between more than two nodes.

TABLE 9: Triangle matrix, T_C, each entry tells us how many triangles exist in each pair of cohorts.

Cohort	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	1	4	3	35	6	0	0	0
2	4	0	9	39	5	0	0	0
3	3	9	1	47	7	13	0	0
4	35	39	47	32	73	0	0	0
5	6	5	7	73	4	16	4	0
6	0	0	13	0	16	12	16	12
7	0	0	0	0	4	16	0	0
8	0	0	0	0	0	12	0	0

There are 156 triangles in the entire network. To further analyze this information, we created a ratio matrix, $R_C = [r_{ij}]$. The entries in this matrix tell us the ratio of triangles within cohorts *i* and *j* to the total number of triangles in the network. This tells us the percentage of triangles in a pair of cohorts. Each entry in Table 10, R_C, was calculated by dividing the entries in T_C by 156.

TABLE 10: Ratio matrix, R_C, each entry tells us the ratio of triangles in each cohort pair to the total triangles in the network.

Cohort	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**
1	.0064	.0256	.0192	.2244	.0385	0	0	0
2	.0256	0	.0577	.25	.0321	0	0	0
3	.0192	.0577	.0064	.3013	.0449	.0833	0	0
4	.2244	.25	.3013	.2051	.4679	0	0	0
5	.0385	.0321	.0449	.4679	.0256	.1026	.0256	0
6	0	0	.0833	0	.1026	.0769	.1026	.0769
7	0	0	0	0	.0256	.1026	0	0
8	0	0	0	0	0	.0769	0	0

We noticed that r_{45} had the largest ratio of .4679, and the next closest was .3013 in r_{34}. Recall that Cohorts 3 and 4 also had the second highest number of edges between cohorts after Cohorts 4 and 5. This means that roughly 47% of the triangles in the whole group come from Cohorts 4 and 5. This percentage is significant because only 32% of the people in the entire group are in these cohorts. In addition to this, 4 of the 17 people in Cohorts 4 and 5 have 100% of their triangles within these cohorts. Also, only 4 of the 17 people have less than 50% of their triangles within Cohorts 4 and 5. When looking closely at Figure 15, it is difficult to find all 73 triangles that exist because the group is so connected. We have three forms of evidence that show Cohorts 4 and 5 are the most connected: degree, eigenvector centrality, and triangles.

In addition to the social network analysis evidence that Cohorts 4 and 5 are the most connected, there is anecdotal evidence as well. Of the 17 members of these two cohorts, 12 scholars were mathematics majors. Also, 12 of the 17 had four semesters of funding. In fact 10 of the 17 students were both mathematics majors and had funding for the maximum amount of time. This is illustrated in the Venn diagram in Figure 16. These students had several classes together and would be more likely to spend time with each other and become friends. The two other students who received funding for all four semesters are both biology majors and share some of the same friends in the REMAST group.

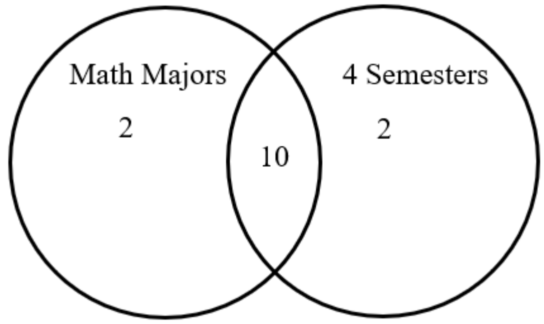


FIGURE 16: Venn diagram showing the overlap between math majors and 4 semesters of funding for Cohorts 4 and 5.

Conjecture 2: *There exists a distinct gap between the Phase I and Phase II REMAST grants.*

To illustrate what we saw in Figure 1 that led us to this conjecture, Figure 17 has a white box around the black region that shows the few friendships between Phase I and Phase II. This conjecture did not require as much analysis as Conjecture 1 because there are so few connections that exist between the Phase I and Phase II REMAST grants. We believe the year off between grants may have contributed to this gap as the cohorts from Phase I do not interact much with the cohorts in the Phase II grant. Additionally, the scholars in Cohorts 5 and 6 would have had very few if any classes together because of the year-long break in providing scholarships.

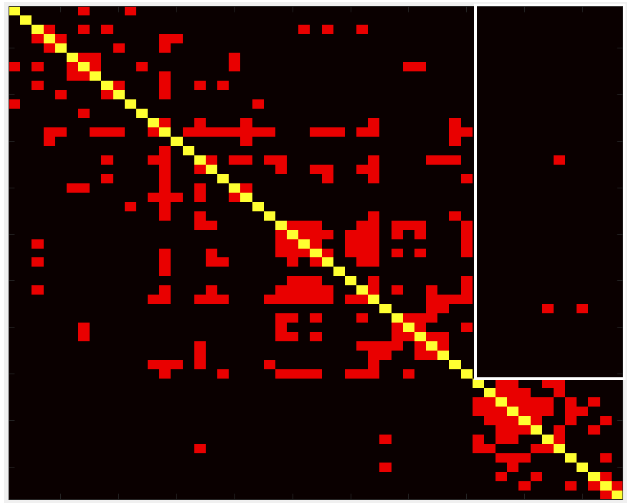


FIGURE 17: The white box surrounds the region of A_C where the friendships between Phase I and Phase II are located.

There are only three friendships between the two phases, shown by the three red squares in the white box in Figure 17. One of the friendships is between members of Cohorts 3 and 6. The two met at the REMAST conference that is held every summer in Brookings and they both teach science. The other two friendships are between a member of Cohort 5 and two members of Cohort 6 and Cohort 7. These friendships transpired as a result of the three members being interested in running. Two of them are alumni of the SDSU Cross Country team. The two members of Cohort 6 and Cohort 7 are not friends themselves, but both teach math. The member of Cohort 5 who is friends with both of them teaches science.

Figure 18b shows the social network graph of Semesters 1 and 4 with the nodes colored by cohort. We include this image here because it has two components. The components exist because of the separation of the Phase I and Phase II grants. The smaller component contains recipients of the Phase II grant. The bottom, larger component contains recipients of the Phase I grant.

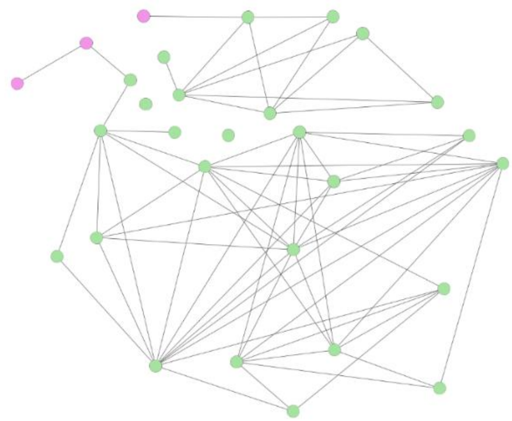


FIGURE 18a: Social network graph of Semester 1 and Semester 4

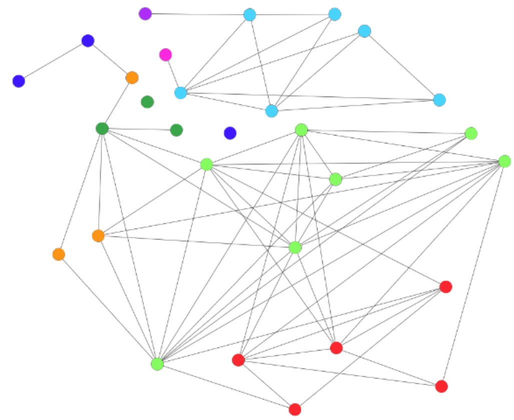


FIGURE 18b: Figure 18a colored by cohort number

Conjecture 3: *Cohort 1 has the most connections throughout the Phase I grant.*

This conjecture was made from the dull red that is in the white box in Figure 19. This conjecture was also made shortly after we created these images, before we had spent a lot of time familiarizing ourselves with the data and the meaning of the images. To investigate this conjecture we used the edge matrix, E_C, developed for Conjecture 1.

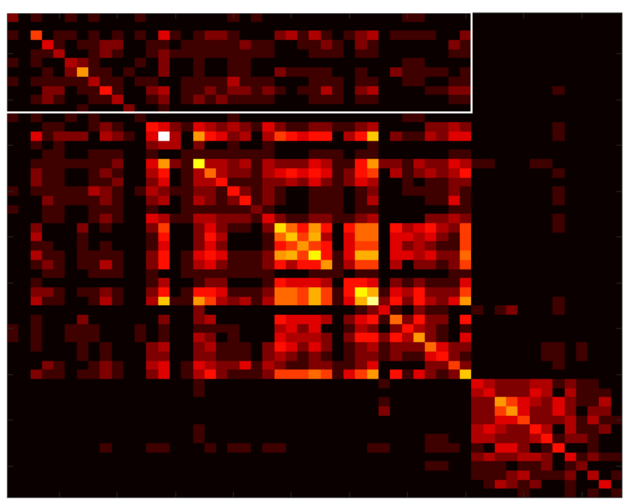


FIGURE 19: The white box surrounds the region of M_C that shows the mutual friends Cohort 1 has with the rest of Phase I.

After looking at Figure 19 closer we determined that our original conjecture was false. We chose to keep it to illustrate that in mathematics, sometimes our conjectures are false. From E_C we have the number of friendships between each pair of cohorts. We will sum the entries of the first five rows until the fifth column. This will tell us how many friendships Cohorts 1 through 5 have within Phase I. These sums are in Table 11.

TABLE 11: Shows the number of friendships each Phase I cohort has with each other.

Cohort	**Friendships in Phase I**
1	28
2	29
3	35
4	61
5	40

Our original conjecture was wrong, as seen in Table 11: Cohort 1 actually has the fewest connections in Phase I. Clearly, Cohort 4 has the most connections. Figure 20 highlights this with a white box surrounding the part of the matrix that represents Cohort 4's mutual friendships in Phase I. Instead of the dull red that was present in the white box in Figure 19, there is more orange, yellow and bright red in the white box in Figure 20. These brighter colors illustrate that the people in Cohort 4 have more mutual friends within Phase I.

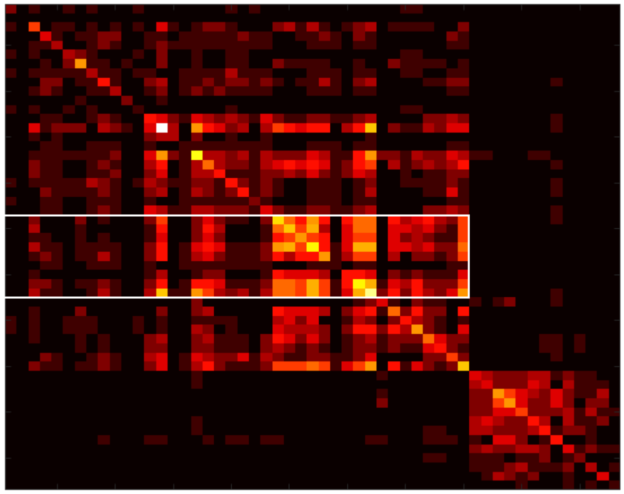


FIGURE 20: The white box surrounds the region of M_C that shows the mutual friends Cohort 4 has with the rest of Phase I.

Another observation we had when looking at which cohort had the most connections was the spread of the connections. Referring back to E_C, Cohort 5 has at least one friendship in every cohort except Cohort 8. Whereas Cohort 4 only has friendships with people in the Phase I grant. So, while Cohort 4 has the most connections in Phase I, Cohort 5 has more widespread connections throughout the entire REMAST program.

Conjecture 4: *There is a clear separation between cohorts in the mutual friends matrix.*

This conjecture was made from the bright colors surrounding the diagonal in Figure 3. To emphasize what we saw we added white boxes around each region that shows the mutual friends each cohort has with itself in M_C, shown in Figure 21. To investigate this conjecture we used Figure 21 as a visual aid and the edge matrix, E_C to help verify the mathematics.

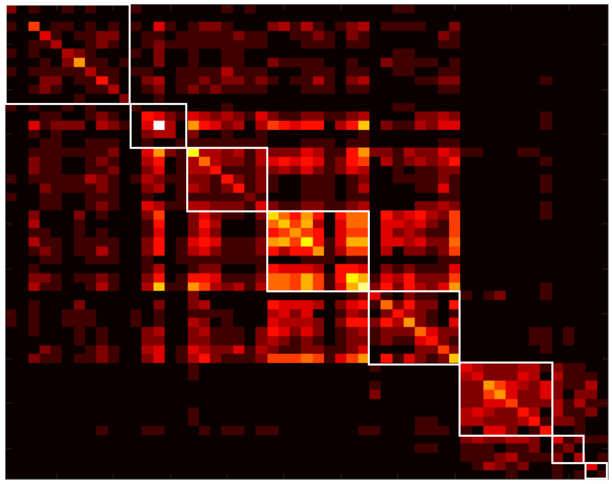


FIGURE 21: The white boxes surround the regions of M_C that show the mutual friends each cohort has with itself.

We focused on the diagonal entries of E_C since they contain the information about the friendships within the cohorts. For Cohorts 1, 4, and 6, the diagonal entry is the largest in their respective rows. This means that there are more friendships in their own cohort than with the other cohorts. Cohort 2 was not very connected with other cohorts (see Table 11) as they had the second fewest friendships in Phase I. Cohort 3 is more connected to the cohorts the year before and the year after than they were with themselves. Cohort 5 is very connected with Cohort 4, as shown by Conjecture 1. If we ignore e_{45}, then e_{55} is the largest number in the fifth row. This means that besides the friendships with Cohort 4, Cohort 5 is most connected with itself. Since Cohorts 7 and 8 are still pretty new to the program they don't have many friendships yet so there isn't a large distinction in their corner of Figure 21.

The observations made from E_C are solidified when looking at P_C, Table 7. The diagonal entries for Cohorts 1, 4, and 6 are still the largest in their respective rows. We also see in Table 7 that the largest frequency in row 8, is the diagonal entry. However, because there are only two people in Cohort 8, the higher frequencies are deceiving. If we look at row 3, Cohort 3 is much more connected with Cohort 2 than they are with Cohort 4. Looking at E_C, it looked like Cohort 3 was equally connected with both Cohorts 2 and 4. A similar observation can be made for row 5, containing the information about Cohort 5. If we ignore p_{45}, then Cohort 5 is significantly more connected with itself than the other cohorts.

One thing to note is that Cohort 6 has very little black in their white box, meaning this group shares a lot of mutual friends. In fact, Cohort 6 is the only cohort that is completely connected (see Figure 22). As a reminder, a graph is connected if for every pair of nodes, there is a path between them. If we look at Table 7, the largest entry is .6429 located in p_{66}. This entry corresponds to the relative frequency of edges that exist in Cohort 6, so about 64% of the possible edges are present. Cohort 6 was also the first cohort to receive funding from the Phase II grant. Additionally, all but two members of Cohort 6 received all four semesters of funding and half the members were math majors. There were also consistent mentoring meetings with faculty members for Cohort 6 which likely played a role in the friendships that were formed.

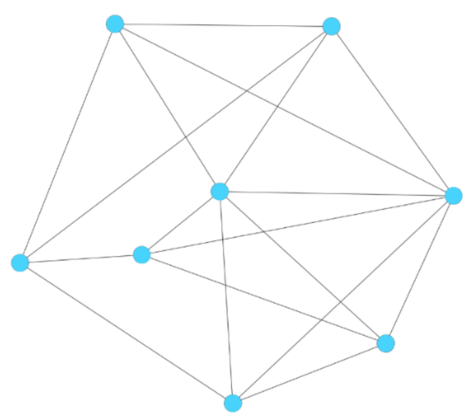


FIGURE 22: Social network graph of Cohort 6

Conjecture 5: *The longer a person received funding the more connected they are within the entire group.*

This is the only conjecture we made from the data that was sorted by the number of semesters a student received funding. This conjecture was made from Figures 5 and 6. Recall in Figure 5 there were a lot of edges within four semesters and connecting four semesters to the other groups of semesters. To analyze what we saw in Figure 6, we placed white boxes around the regions in M_S (the mutual friends matrix sorted by semester) that show the mutual friends within the semester groupings (see Figure 23). Notice the brightest colors in the image are in the largest white box in the bottom right corner. This is the portion of the matrix where the people with four semesters of funding are located. There are also bright colors above and to the left of the large white box.

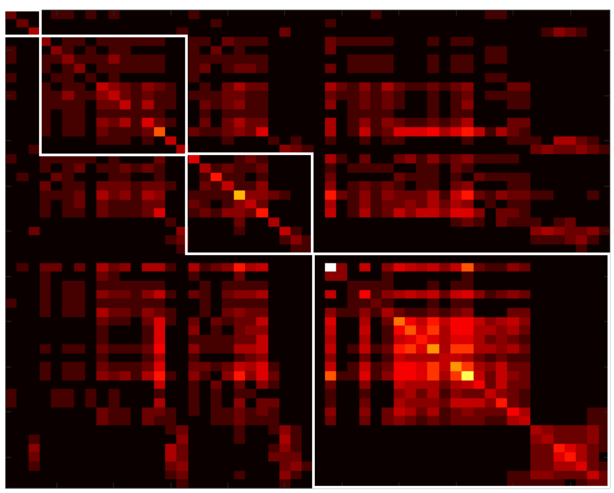


FIGURE 23: The white boxes surround the regions of M_S that show the mutual friends each semester of funding has with itself.

Further investigation of this conjecture was done using the same measures we did for Conjecture 1. We created another degree matrix from the data generated by Gephi, we call this one $D_S = [d_{ij}]$. This matrix is shown in Table 12. Notice that there are very large numbers located in the fourth row.

TABLE 12: Degree matrix, D_S each entry shows how many degrees exist between each pair of semesters.

Semester	**1**	**2**	**3**	**4**
1	2	16	24	128
2	16	12	54	190
3	24	54	16	198
4	128	190	198	122

We used Equation 1 to create the edge matrix that we call $E_S = [e_{ij}]$, shown in Table 13. The numbers in E_S are much smaller than in D_S. Notice the largest numbers in E_S are e_{24}, e_{34}, and e_{44} with values of 28, 30, and 61 respectively. These numbers show us that the people who received four semesters of funding have more friendships across the entire network. We believe this is because the people who received four semesters of funding had more opportunities to spend time with each other and become friends. About 40% of the friendships come from within the group of people who received all four semesters of funding.

TABLE 13: Edge matrix, E_S each entry shows how many edges connect each pair of semesters.

Semester	**1**	**2**	**3**	**4**
1	1	1	3	2
2	1	6	13	28
3	3	13	8	30
4	2	28	30	61

We also looked at the eigenvector centralities between the following semester pairs: 2 and 3, 2 and 4, and 3 and 4. We chose to omit the semester pairs including one semester of funding because there were only three people in this group and they were not very connected. Desmos was used to create boxplots of the different sets of eigenvector centralities. These groups have more people than the cohort pair groups. In addition, there were more people in these semester pairs that had an eigenvector centrality of zero. This fact affected the first quartile of all three pairings. Figure 24 shows the boxplots of the three semester pairings. Table 14 shows the five number summary from the boxplots. We included mode because of the amount of zeros in the data. The rows are color coded according to the color of the boxplot.

TABLE 14: Five-number summary from the boxplots in Figure 24. The colors of the row matches the color of the boxplot.

Semester Pair	min	1st Quartile	median	3rd Quartile	max	inter-quartile range	mean	mode	# of people
2 and 3	0	0.05415	0.14125	0.3574	0.6686	0.30325	0.215625	0	24
2 and 4	0	0	0.0528	0.2745	0.5112	0.2745	0.15124	0	39
3 and 4	0	0.00425	0.1247	0.3036	0.5213	0.29935	0.16502	0.0005	37

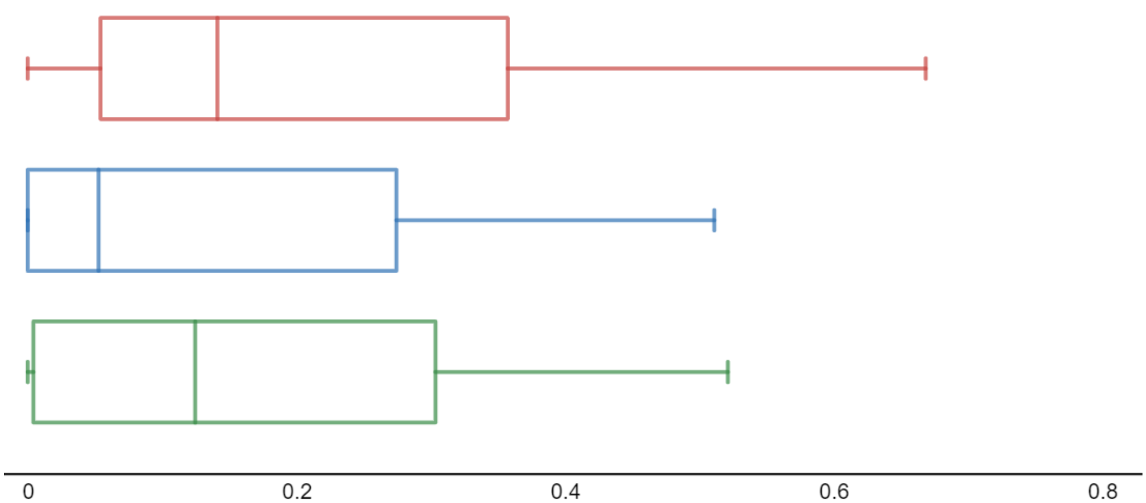


FIGURE 24: Stacked boxplots of eigenvector centralities from three pairs of semesters of funding.

Notice that the first quartile for the semester pair of 2 and 4 is 0. There were 11 people who had an eigenvector centrality of zero in this pairing. The presence of zeros in semesters 2 and 4 made itself apparent with a very low first quartile and a much smaller median than the other two pairs. The interquartile ranges for the semester pairs are much closer together than the interquartile ranges for the cohort pairs. Thus, it is difficult to make any conclusions based on the interquartile ranges for this data.

TABLE 15: Triangle matrix, T_S each entry shows how many triangles exist in each pair of semesters.

Semester	**1**	**2**	**3**	**4**
1	0	0	0	67
2	0	0	5	103
3	0	5	0	103
4	67	103	103	67

We also looked at the number of triangles present in the semester pairs. A triangle matrix, $T_S = [t_{ij}]$, was created in Table 15. Similar to the investigation of Conjecture 1, we created a ratio matrix, $R_S = [r_{ij}]$, to get more information about what the triangle data means. There are still 156 triangles present in the whole group. To find the entries of R_S, we divided the entries of T_S by 156. The ratio matrix is shown in Table 16.

TABLE 16: Ratio matrix, R_S each entry shows the ratio of triangles in each semester pair to the total triangles in the group.

Semester	**1**	**2**	**3**	**4**
1	0	0	0	.4295
2	0	0	.0321	.6603
3	0	.0321	0	.6603
4	.4295	.6603	.6603	.4295

It is very clear from the fourth row of R_S that almost all the triangles that exist in the social network graph come from students who received four semesters of funding. An interesting thing to note is that r_{14} and r_{44} have the same value. This is because there are no triangles formed between the people who had one semester of funding and those who had four semesters of funding, see Figure 18a. With the triangle analysis we have two forms of evidence that show students who received four semesters of funding are the most connected.

CHAPTER 4

FURTHER INVESTIGATION

The goal of this project was to mathematically measure the connections within the Facebook group. This was done by analyzing the friendships that exist between 53 members of the REMAST Facebook group. In addition to the friendships, we also collected qualitative data about the Facebook posts that were made in the group from July 2017-June 2018. We wanted to keep the focus for the interactions on people who are directly involved with REMAST. We included the three co-PIs for the post data, giving us 56 members for which we tracked interactions. The interactions we tracked were: when posts were made, who made the post, who commented on it, and who had a reaction to the post. The analysis of the interactions will be included in a different paper.

We also gave each post a qualitative code that corresponded to the content of the post. The qualitative codes we used come from the Early Career Teacher (ECT) Resilience Framework. We conducted a survey in 2015 among people who had graduated from SDSU with teaching certification in all disciplines. Students were asked if they were part of REMAST or not and were asked questions from parts of the ECT Resilience Framework. The questions came from the following domains: Policies & Practices, Teachers' Work, School Culture, Relationships, and Teacher Identity. The codes we used for the tracked Facebook posts came directly from the ECT Resilience Framework. We did this so we could see if there was any relation between the Framework domains and the Facebook posts [8].

We would also like to see if there are connections that could be made between the qualitative and quantitative data. For example, if someone authors a post, is it usually their Facebook friends who interact with it? Another thing we want to look into is if the attendance of the summer conference is related to the activity in the Facebook group. We think there may be a link between the people who consistently attend the conference and who actively interact in the Facebook group.

APPENDIX A

Adjacency matrix, part 1, sorted by cohort number.

Cohort	Subject	Semester	ID #	36	3	47	14	21	33	24	62	27	23	28	35	65	22	6
1	1	1	36	0	0	0	0	0	0	1	0	0	0	1	0	0	0	0
1	1	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	47	0	0	0	1	0	0	1	0	1	0	0	0	0	0	0
1	3	3	14	0	0	1	0	1	0	0	0	0	0	0	0	0	1	1
1	1	2	21	0	0	0	1	0	0	0	0	0	1	0	0	0	1	0
1	2	2	33	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
1	2	3	24	1	0	1	0	0	1	0	1	0	0	0	1	0	0	0
1	1	2	62	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
1	3	3	27	0	0	1	0	0	0	0	0	0	1	0	0	0	1	0
1	2	2	23	0	0	0	0	1	0	0	0	1	0	0	0	0	1	0
1	1	1	28	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	2	35	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	2	2	65	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
2	3	4	22	0	0	0	1	1	0	0	1	1	1	0	0	1	0	0
2	1	4	6	0	0	0	1	0	0	0	0	0	0	0	0	0	0	0
2	3	4	40	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3	1	3	17	0	0	0	0	0	0	0	0	1	0	0	0	1	1	0
3	3	4	57	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
3	3	3	50	0	0	0	0	0	0	0	0	1	0	0	0	0	1	0
3	3	2	12	0	0	0	0	0	1	1	0	0	0	0	0	0	1	0
3	1	2	31	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
3	1	4	2	0	0	0	0	0	0	0	0	0	0	1	0	0	1	0
3	2	4	42	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
4	3	4	49	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	61	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	32	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	63	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
4	3	3	25	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
4	1	2	37	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
4	3	4	9	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	54	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
4	3	4	51	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
5	1	3	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	4	29	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	4	34	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	3	4	41	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
5	1	4	19	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	4	26	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	2	55	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
5	3	2	56	0	0	0	0	0	0	0	0	0	0	0	0	0	1	0
6	3	4	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	4	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	4	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	3	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	2	43	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	4	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	3	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	3	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Adjacency matrix, part 2, sorted by cohort number.

Cohort	Subject	Semester	ID #	40	17	57	50	12	31	2	42	49	61	32	63	25	37	9
1	1	1	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	47	0	0	0	0	0	0	0	0	0	0	1	0	1	0	0
1	3	3	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	2	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	2	33	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	2	3	24	0	0	0	0	1	0	0	0	0	0	0	0	0	0	0
1	1	2	62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	27	0	1	0	1	0	0	0	0	0	0	0	0	0	0	0
1	2	2	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	28	0	0	0	0	0	0	1	0	0	0	0	0	0	0	0
2	1	2	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	65	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
2	3	4	22	1	1	1	1	1	1	1	1	0	0	0	1	1	1	0
2	1	4	6	0	0	0	0	0	1	0	0	0	0	0	0	0	0	0
2	3	4	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	3	17	0	0	1	0	1	1	0	1	1	0	0	0	0	0	0
3	3	4	57	0	1	0	0	0	0	0	0	1	0	0	1	1	0	0
3	3	3	50	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
3	3	2	12	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0
3	1	2	31	0	1	0	0	1	0	0	0	0	0	0	0	0	0	0
3	1	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	4	42	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	49	0	1	1	0	0	0	0	0	0	1	1	1	0	0	0
4	3	4	61	0	0	0	0	0	0	0	0	1	0	1	1	1	0	1
4	3	4	32	0	0	0	0	0	0	0	0	1	1	0	1	0	0	1
4	3	4	63	0	0	1	0	0	0	0	0	1	1	1	0	1	0	1
4	3	3	25	0	0	1	1	0	0	0	0	0	1	0	1	0	0	0
4	1	2	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	9	0	0	0	0	0	0	0	0	0	1	1	1	0	0	0
4	3	4	54	0	0	1	0	0	0	0	0	1	1	1	1	1	0	0
4	3	4	51	0	1	1	1	0	0	0	1	1	1	1	1	1	0	1
5	1	3	30	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	4	29	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
5	3	4	34	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
5	3	4	41	0	0	0	0	0	0	0	0	1	1	0	1	0	0	0
5	1	4	19	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	1	4	26	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	2	2	55	0	1	0	0	0	0	0	1	0	0	0	0	0	0	0
5	3	2	56	0	0	0	1	0	0	0	0	1	1	1	1	0	0	1
6	3	4	48	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	4	4	10	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	4	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	4	60	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	3	45	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	3	4	13	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
6	1	2	43	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	2	8	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	3	4	38	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
7	1	3	52	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	1	1	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
8	3	3	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Adjacency matrix, part 3, sorted by cohort number.

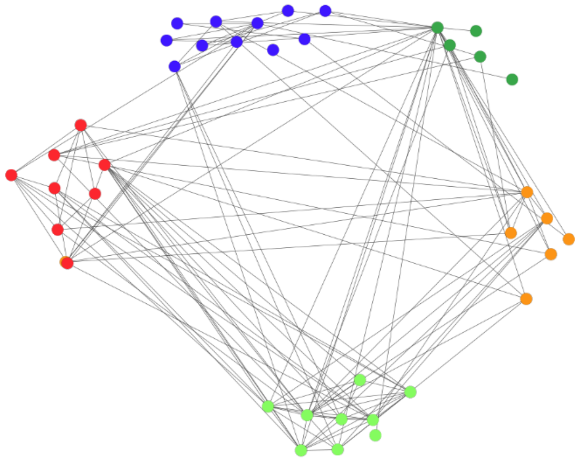
Cohort	Subject	Semester	ID #	54	51	30	29	34	41	19	26	55	56	48	10	1	4	60
1	1	1	36	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	4	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	47	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	14	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	2	21	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	2	33	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	3	24	0	0	0	0	1	1	0	0	0	0	0	0	0	0	0
1	1	2	62	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	3	3	27	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	2	2	23	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
1	1	1	28	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	2	35	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	2	65	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
2	3	4	22	1	1	0	0	0	0	0	0	1	1	0	0	0	0	0
2	1	4	6	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0
2	3	4	40	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	3	17	0	1	0	0	0	0	1	1	1	0	0	0	0	0	0
3	3	4	57	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
3	3	3	50	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
3	3	2	12	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	2	31	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	1	4	2	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
3	2	4	42	0	1	0	0	0	0	0	0	1	0	0	0	0	0	0
4	3	4	49	1	1	0	1	1	1	0	0	0	1	0	0	0	0	0
4	3	4	61	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0
4	3	4	32	1	1	0	0	0	0	0	0	0	1	0	0	0	0	0
4	3	4	63	1	1	0	1	0	1	0	0	0	1	0	0	0	0	0
4	3	3	25	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0
4	1	2	37	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
4	3	4	9	0	1	0	0	0	0	0	0	0	1	0	0	0	0	0
4	3	4	54	0	1	0	1	0	0	1	0	0	1	0	0	0	0	0
4	3	4	51	1	0	0	0	0	0	1	1	1	1	0	0	0	0	0
5	1	3	30	0	0	0	0	0	0	1	1	0	0	0	0	0	0	0
5	3	4	29	1	0	0	0	1	1	1	0	0	0	0	0	0	0	0
5	3	4	34	0	0	0	1	0	1	0	0	0	1	0	0	0	0	0
5	3	4	41	0	0	0	1	1	0	1	1	0	0	0	0	0	0	0
5	1	4	19	1	1	1	1	0	1	0	1	0	0	0	0	0	0	0
5	1	4	26	0	1	1	0	0	1	1	0	0	0	0	0	0	0	0
5	2	2	55	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0
5	3	2	56	1	1	0	0	1	0	0	0	0	0	0	0	0	0	0
6	3	4	48	0	0	0	0	0	0	0	0	0	0	0	0	1	1	0
6	4	4	10	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
6	3	4	1	0	0	0	0	0	0	0	0	0	0	1	1	0	1	1
6	3	4	4	0	0	0	0	0	0	0	0	0	0	1	1	1	0	1
6	1	4	60	0	0	0	0	0	0	0	0	0	0	0	1	1	1	0
6	1	3	45	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
6	3	4	13	0	0	1	0	0	0	0	0	0	0	1	0	1	1	0
6	1	2	43	0	0	0	0	0	0	0	0	0	0	1	1	0	0	0
7	1	2	8	0	0	0	0	0	0	0	0	0	0	0	0	1	1	1
7	3	4	38	0	0	1	0	0	0	0	0	0	0	0	0	0	1	0
7	1	3	52	0	0	0	0	0	0	0	0	0	0	0	0	1	0	0
8	1	1	20	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
8	3	3	18	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0

Adjacency matrix, part 4, sorted by cohort number.

Cohort	Subject	Semester	ID #	45	13	43	8	38	52	20	18
1	1	1	36	0	0	0	0	0	0	0	0
1	1	4	3	0	0	0	0	0	0	0	0
1	3	3	47	0	0	0	0	0	0	0	0
1	3	3	14	0	0	0	0	0	0	0	0
1	1	2	21	0	0	0	0	0	0	0	0
1	2	2	33	0	0	0	0	0	0	0	0
1	2	3	24	0	0	0	0	0	0	0	0
1	1	2	62	0	0	0	0	0	0	0	0
1	3	3	27	0	0	0	0	0	0	0	0
1	2	2	23	0	0	0	0	0	0	0	0
1	1	1	28	0	0	0	0	0	0	0	0
2	1	2	35	0	0	0	0	0	0	0	0
2	2	2	65	0	0	0	0	0	0	0	0
2	3	4	22	0	0	0	0	0	0	0	0
2	1	4	6	0	0	0	0	0	0	0	0
2	3	4	40	0	0	0	0	0	0	0	0
3	1	3	17	0	0	1	0	0	0	0	0
3	3	4	57	0	0	0	0	0	0	0	0
3	3	3	50	0	0	0	0	0	0	0	0
3	3	2	12	0	0	0	0	0	0	0	0
3	1	2	31	0	0	0	0	0	0	0	0
3	1	4	2	0	0	0	0	0	0	0	0
3	2	4	42	0	0	0	0	0	0	0	0
4	3	4	49	0	0	0	0	0	0	0	0
4	3	4	61	0	0	0	0	0	0	0	0
4	3	4	32	0	0	0	0	0	0	0	0
4	3	4	63	0	0	0	0	0	0	0	0
4	3	3	25	0	0	0	0	0	0	0	0
4	1	2	37	0	0	0	0	0	0	0	0
4	3	4	9	0	0	0	0	0	0	0	0
4	3	4	54	0	0	0	0	0	0	0	0
4	3	4	51	0	0	0	0	0	0	0	0
5	1	3	30	0	1	0	0	1	0	0	0
5	3	4	29	0	0	0	0	0	0	0	0
5	3	4	34	0	0	0	0	0	0	0	0
5	3	4	41	0	0	0	0	0	0	0	0
5	1	4	19	0	0	0	0	0	0	0	0
5	1	4	26	0	0	0	0	0	0	0	0
5	2	2	55	0	0	0	0	0	0	0	0
5	3	2	56	0	0	0	0	0	0	0	0
6	3	4	48	0	1	1	0	0	0	0	0
6	4	4	10	0	0	1	0	0	0	0	0
6	3	4	1	1	1	0	1	0	1	0	0
6	3	4	4	1	1	0	1	1	0	0	0
6	1	4	60	1	0	0	1	0	0	1	0
6	1	3	45	0	0	1	0	0	1	0	0
6	3	4	13	0	0	1	0	0	0	0	0
6	1	2	43	1	1	0	0	0	0	0	0
7	1	2	8	0	0	0	0	0	0	1	0
7	3	4	38	0	0	0	0	0	0	0	0
7	1	3	52	1	0	0	0	0	0	1	0
8	1	1	20	0	0	0	1	0	1	0	1
8	3	3	18	0	0	0	0	0	1	0	0

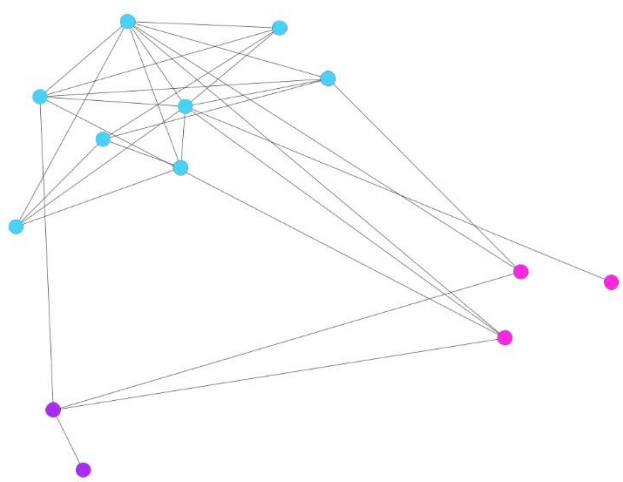
Phase I: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	9.652381	0	0	0.0084
3	1	1	4	0	0	0	0	0
47	1	3	3	6	34.44204	0.133333	2	0.1382
14	1	3	3	4	12.959325	0.166667	1	0.0717
21	1	1	2	3	0.75	0.666667	2	0.061
33	1	2	2	3	0.333333	0.666667	2	0.0247
24	1	2	3	8	99.502619	0.107143	3	0.0728
62	1	1	2	3	17.459113	0.333333	1	0.0567
27	1	3	3	5	8.703175	0.3	3	0.1199
23	1	2	2	3	0.5	0.666667	2	0.0657
28	1	1	1	2	4.215873	0	0	0.0059
35	2	1	2	1	0	0	0	0.0078
65	2	2	2	5	1.05119	0.8	8	0.1637
22	2	3	4	21	286.185673	0.166667	35	0.4315
6	2	1	4	3	1.503968	0	0	0.0379
40	2	3	4	1	0	0	0	0.0463
17	3	1	3	12	59.716855	0.318182	21	0.2978
57	3	3	4	7	2.393347	0.761905	16	0.2981
50	3	3	3	5	2.15864	0.6	6	0.1852
12	3	3	2	5	36.63621	0.4	4	0.1005
31	3	1	2	5	12.079167	0.5	5	0.1106
2	3	1	4	2	29.563492	0	0	0.0469
42	3	2	4	4	0	1	6	0.1518
49	4	3	4	11	17.08786	0.527273	29	0.3822
61	4	3	4	10	7.309811	0.622222	28	0.3608
32	4	3	4	8	5.713742	0.714286	20	0.3029
63	4	3	4	12	31.309426	0.560606	37	0.4315
25	4	3	3	8	9.182234	0.571429	16	0.2985
37	4	1	2	1	0	0	0	0.0463
9	4	3	4	5	0	1	10	0.2115
54	4	3	4	12	31.843091	0.515152	34	0.4234
51	4	3	4	17	83.736223	0.360294	49	0.5152
30	5	1	3	2	0	1	1	0.0356
29	5	3	4	7	3.201471	0.619048	13	0.2293
34	5	3	4	5	10.90257	0.5	5	0.1346
41	5	3	4	8	34.711156	0.392857	11	0.2084
19	5	1	4	7	27.506227	0.380952	8	0.1977
26	5	1	4	5	16.490156	0.5	5	0.1345
55	5	2	2	6	15.369444	0.6	9	0.1713
56	5	3	2	10	27.830186	0.555556	25	0.3623



Phase II: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
48	6	3	4	4	0.95	0.666667	4	0.3634
10	6	4	4	4	2.216667	0.5	3	0.3915
1	6	3	4	8	11.883333	0.392857	11	0.6311
4	6	3	4	8	17.25	0.357143	10	0.6101
60	6	1	4	6	9.716667	0.533333	8	0.5109
45	6	1	3	5	5.216667	0.4	4	0.4388
13	6	3	4	4	0.95	0.666667	4	0.3634
43	6	1	2	4	1.583333	0.166667	1	0.2975
8	7	1	2	4	4.266667	0.666667	4	0.3778
38	7	3	4	1	0	0	0	0.1165
52	7	1	3	3	3.133333	0.333333	1	0.2474
20	8	1	1	4	11.833333	0.166667	1	0.2253
18	8	3	3	1	0	0	0	0.043



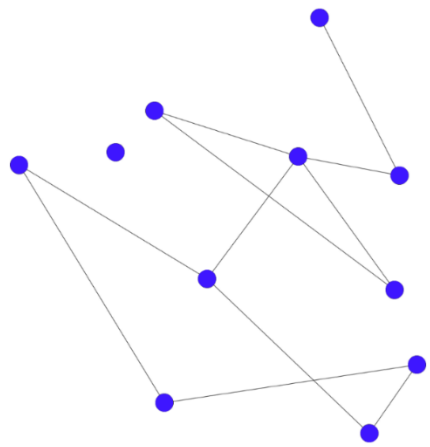
R code used to generate Gephi social network graphs and data tables

```r
#Set storage location by copying folder address and replacing backslashes with forward slashes.
folder = "//moray.jacks.local/sharon.vestal/Documents and Settings/My Documents/REMAST/REMAST social network" #tells which folder to us
setwd(folder)
data = read.csv("Updated_Facebook_Data_No_Faculty.csv") #reads the csv file
data2 = read.csv("Updated_Facebook_Data_No_Faculty.csv",header=F)
Facebook2 =  cbind(data2[,4], data2[,5:57]) #tells which columns to look at in the csv file


#Export Node List
NodeList = data[1:4] #captures the id number, cohort number, semesters of funding, and content area and creates a node list
write.csv(NodeList, row.names=FALSE, "NodeListNoFaculty.csv")

#Export Facebook Edges NoFaculty
Facebook = cbind(data[,4], data[,5:57])
Facebook_Edges_NoFaculty = data.frame(Source=integer(),Target=integer(),Weight=integer()) #this loop identifies which pair of node
for (row in 1:nrow(Facebook)) {   #this loop identifies which pair of nodes are friends with each other and creates the edge list
  for (column in 2:ncol(Facebook)) {
    if (!is.na(Facebook[row,column])) {
      if ((Facebook[row,column])>0 && Facebook2[1,column]!=Facebook[row,1])
        {Facebook_Edges_NoFaculty = rbind(Facebook_Edges_NoFaculty, c(Facebook[row,1], Facebook2[1,column], Facebook[row,column]))}}
  }
}
names(Facebook_Edges_NoFaculty) = c("Source", "Target", "Weight")
write.csv(Facebook_Edges_NoFaculty, row.names=FALSE, "Facebook_Edges_NoFaculty.csv")
```

APPENDIX B

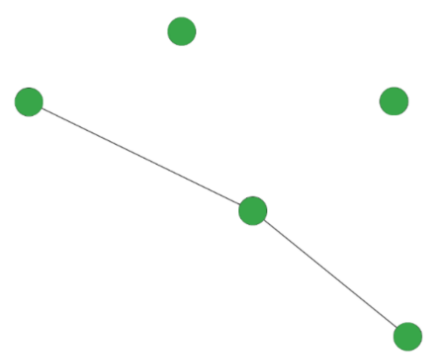
Cohort 1: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	8	0	0	0.3837
3	1	1	4	0	0	0	0	0
47	1	3	3	3	21	0	0	0.5692
14	1	3	3	2	6	0	0	0.3089
21	1	1	2	2	1	0	0	0.2051
33	1	2	2	2	0	1	1	0.5368
24	1	2	3	4	24	0.166667	1	0.8086
62	1	1	2	2	0	1	1	0.5368
27	1	3	3	2	6	0	0	0.3089
23	1	2	2	2	1	0	0	0.2051
28	1	1	1	1	0	0	0	0.1531



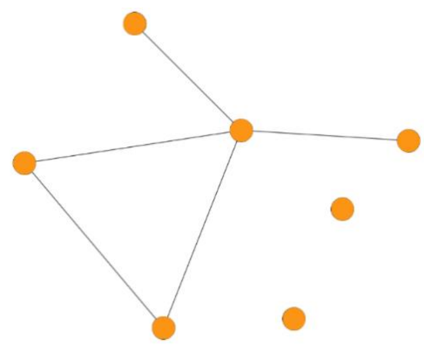
Cohort 2: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
35	2	1	2	0	0	0	0	0
65	2	2	2	1	0	0	0	0.7071
22	2	3	4	2	1	0	0	1
6	2	1	4	0	0	0	0	0
40	2	3	4	1	0	0	0	0.7071



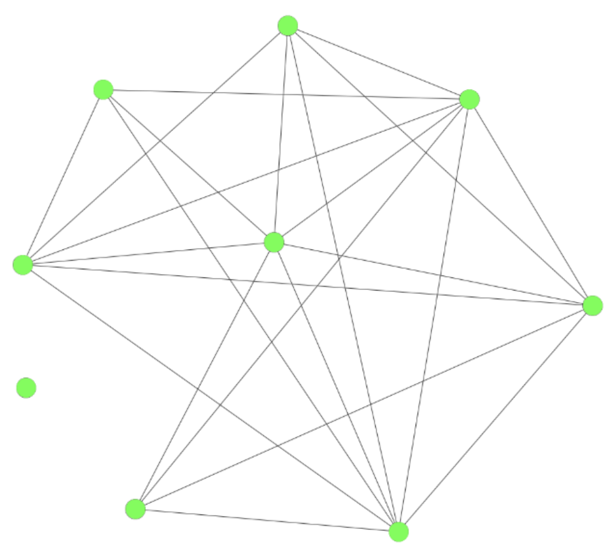
Cohort 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
17	3	1	3	4	5	0.166667	1	0.8992
57	3	3	4	1	0	0	0	0.3838
50	3	3	3	0	0	0	0	0
12	3	3	2	2	0	1	1	0.6696
31	3	1	2	2	0	1	1	0.6696
2	3	1	4	0	0	0	0	0
42	3	2	4	1	0	0	0	0.3838



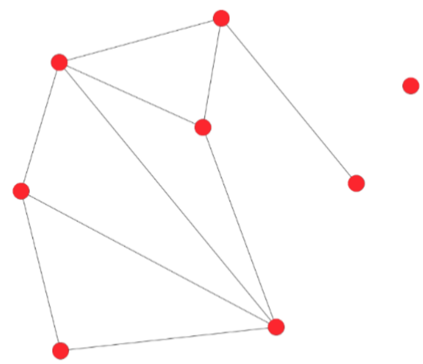
Cohort 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
49	4	3	4	5	0	1	10	0.4617
61	4	3	4	7	1.333333	0.761905	16	0.5698
32	4	3	4	6	0.5	0.866667	13	0.5157
63	4	3	4	7	1.333333	0.761905	16	0.5698
25	4	3	3	4	0	1	6	0.3748
37	4	1	2	0	0	0	0	0
9	4	3	4	4	0	1	6	0.3748
54	4	3	4	6	0.5	0.866667	13	0.5157
51	4	3	4	7	1.333333	0.761905	16	0.5698



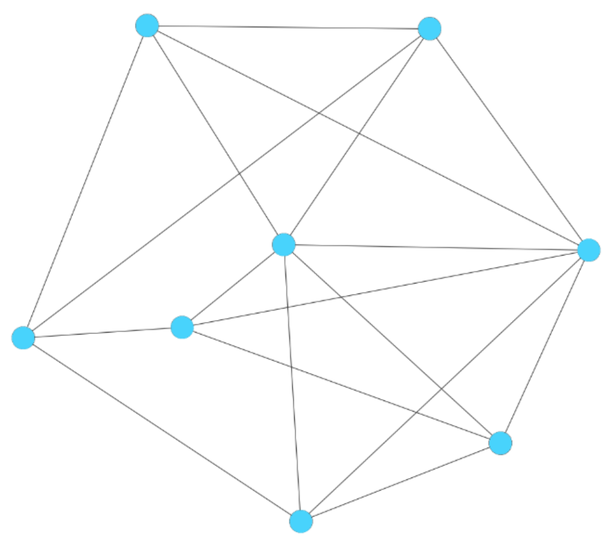
Cohort 5: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
30	5	1	3	2	0	1	1	0.3912
29	5	3	4	3	1.666667	0.666667	2	0.5757
34	5	3	4	3	5	0.333333	1	0.4431
41	5	3	4	4	4.833333	0.5	3	0.7086
19	5	1	4	4	3.333333	0.5	3	0.6963
26	5	1	4	3	1.166667	0.666667	2	0.5596
55	5	2	2	0	0	0	0	0
56	5	3	2	1	0	0	0	0.138



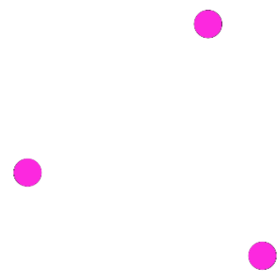
Cohort 6: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
48	6	3	4	4	0.5	0.666667	4	0.4544
10	6	4	4	4	1	0.5	3	0.4575
1	6	3	4	6	2.583333	0.533333	8	0.6301
4	6	3	4	6	2.583333	0.533333	8	0.6301
60	6	1	4	4	0.25	0.833333	5	0.4689
45	6	1	3	4	1	0.5	3	0.4575
13	6	3	4	4	0.5	0.666667	4	0.4544
43	6	1	2	4	1.583333	0.166667	1	0.3932



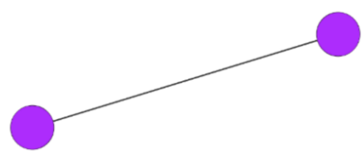
Cohort 7: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
8	7	1	2	0	0	0	0	0
38	7	3	4	0	0	0	0	0
52	7	1	3	0	0	0	0	0



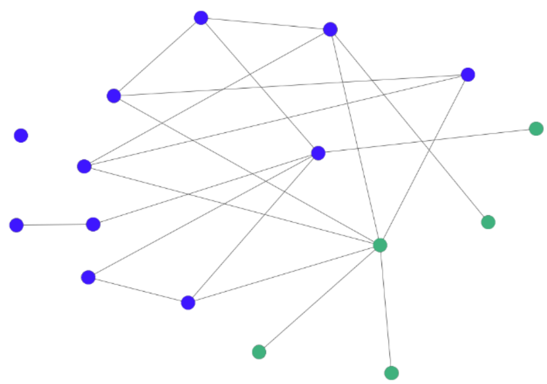
Cohort 8: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
20	8	1	1	1	0	0	0	1
18	8	3	3	1	0	0	0	1



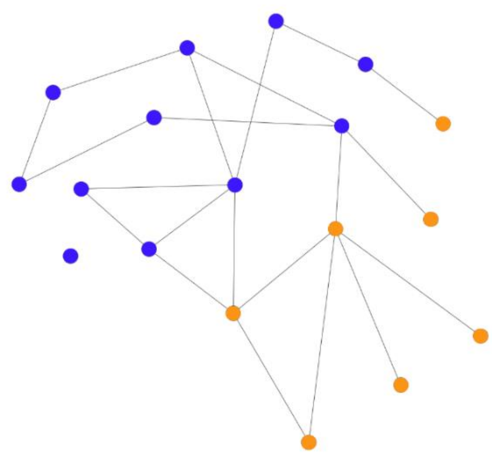
Cohorts 1 and 2: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	13	0	0	0.0925
3	1	1	4	0	0	0	0	0
47	1	3	3	3	18.5	0	0	0.3516
14	1	3	3	4	17.5	0.166667	1	0.4925
21	1	1	2	3	1	0.666667	2	0.4904
33	1	2	2	2	0	1	1	0.1848
24	1	2	3	5	38.5	0.1	1	0.3013
62	1	1	2	3	22.5	0.333333	1	0.3527
27	1	3	3	3	4.5	0.333333	1	0.4506
23	1	2	2	3	0.5	0.666667	2	0.4812
28	1	1	1	1	0	0	0	0.0261
35	2	1	2	1	0	0	0	0.0851
65	2	2	2	1	0	0	0	0.2153
22	2	3	4	7	40	0.142857	3	0.7622
6	2	1	4	1	0	0	0	0.1391
40	2	3	4	1	0	0	0	0.2153



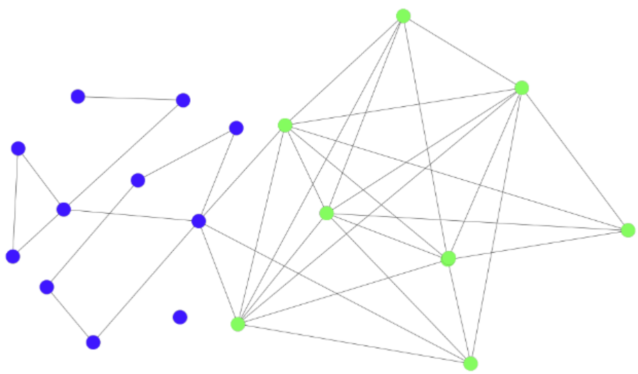
Cohorts 1 and 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	28	0	0	0.2254
3	1	1	4	0	0	0	0	0
47	1	3	3	3	36	0	0	0.3503
14	1	3	3	2	7.5	0	0	0.135
21	1	1	2	2	1	0	0	0.0834
33	1	2	2	3	2.5	0.666667	2	0.5096
24	1	2	3	5	55.5	0.2	2	0.6463
62	1	1	2	2	0	1	1	0.3599
27	1	3	3	4	38.5	0	0	0.3438
23	1	2	2	2	6.5	0	0	0.133
28	1	1	1	2	15	0	0	0.0777
17	3	1	3	5	39	0.1	1	0.514
57	3	3	4	1	0	0	0	0.16
50	3	3	3	1	0	0	0	0.107
12	3	3	2	4	26.5	0.333333	2	0.6308
31	3	1	2	2	0	1	1	0.3564
2	3	1	4	1	0	0	0	0.0242
42	3	2	4	1	0	0	0	0.16



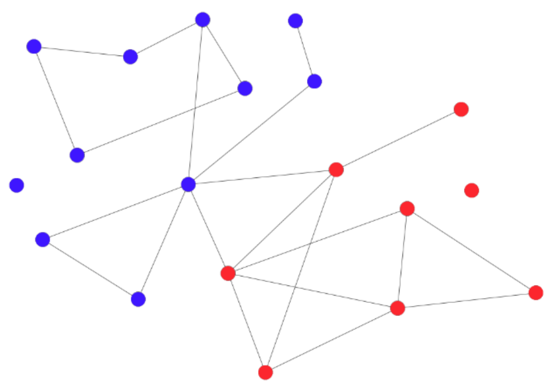
Cohorts 1 and 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	16	0	0	0.0078
3	1	1	4	0	0	0	0	0
47	1	3	3	6	93.2	0.133333	2	0.2594
14	1	3	3	2	14	0	0	0.0438
21	1	1	2	2	1	0	0	0.0085
33	1	2	2	2	0	1	1	0.0091
24	1	2	3	4	56	0.166667	1	0.0466
62	1	1	2	2	0	1	1	0.0091
27	1	3	3	2	14	0	0	0.0438
23	1	2	2	2	1	0	0	0.0085
28	1	1	1	1	0	0	0	0.0013
49	4	3	4	5	0	1	10	0.4399
61	4	3	4	7	1.283333	0.761905	16	0.5451
32	4	3	4	7	25.5	0.666667	14	0.5259
63	4	3	4	7	1.283333	0.761905	16	0.5451
25	4	3	3	5	10	0.7	7	0.3964
37	4	1	2	0	0	0	0	0
9	4	3	4	4	0	1	6	0.353
54	4	3	4	7	15.45	0.714286	15	0.532
51	4	3	4	7	1.283333	0.761905	16	0.5451



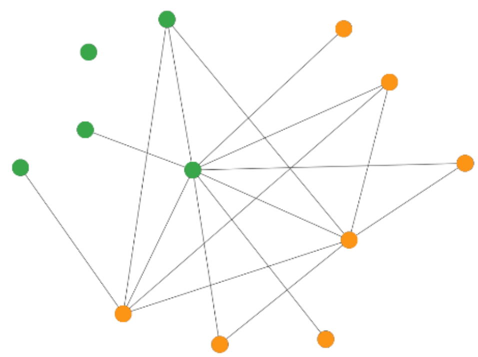
Cohorts 1 and 5: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	15	0	0	0.1698
3	1	1	4	0	0	0	0	0
47	1	3	3	3	49	0	0	0.1896
14	1	3	3	2	13	0	0	0.0592
21	1	1	2	2	1	0	0	0.0229
33	1	2	2	2	0	1	1	0.2171
24	1	2	3	6	87	0.133333	2	0.5618
62	1	1	2	2	0	1	1	0.2171
27	1	3	3	2	13	0	0	0.0592
23	1	2	2	2	1	0	0	0.0229
28	1	1	1	1	0	0	0	0.0473
30	5	1	3	2	0	1	1	0.2575
29	5	3	4	3	1.666667	0.666667	2	0.484
34	5	3	4	4	20	0.333333	2	0.5263
41	5	3	4	5	39.833333	0.4	4	0.6956
19	5	1	4	4	8.333333	0.5	3	0.5146
26	5	1	4	3	6.166667	0.666667	2	0.4091
55	5	2	2	0	0	0	0	0
56	5	3	2	1	0	0	0	0.1467



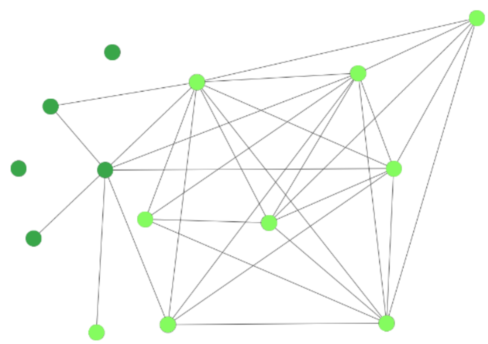
Cohorts 2 and 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
35	2	1	2	0	0	0	0	0
65	2	2	2	3	0	1	3	0.4478
22	2	3	4	9	28.833333	0.194444	7	0.7467
6	2	1	4	1	0	0	0	0.1273
40	2	3	4	1	0	0	0	0.1712
17	3	1	3	6	4.833333	0.466667	7	0.6509
57	3	3	4	2	0	1	1	0.3205
50	3	3	3	1	0	0	0	0.1712
12	3	3	2	3	0	1	3	0.4478
31	3	1	2	5	9.333333	0.5	5	0.5551
2	3	1	4	1	0	0	0	0.1712
42	3	2	4	2	0	1	1	0.3205



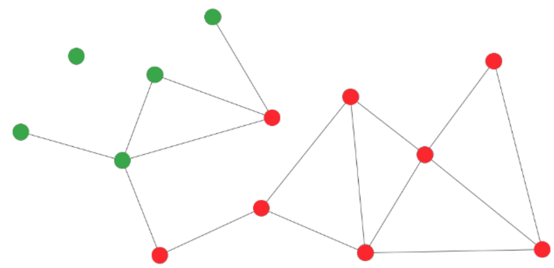
Cohorts 2 and 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
35	2	1	2	0	0	0	0	0
65	2	2	2	2	0	1	1	0.1488
22	2	3	4	7	20.5	0.333333	7	0.3651
6	2	1	4	0	0	0	0	0
40	2	3	4	1	0	0	0	0.0573
49	4	3	4	5	0	1	10	0.4137
61	4	3	4	7	1.333333	0.761905	16	0.5129
32	4	3	4	6	0.5	0.866667	13	0.4588
63	4	3	4	8	5.583333	0.678571	19	0.5625
25	4	3	3	5	0.75	0.9	9	0.399
37	4	1	2	1	0	0	0	0.0573
9	4	3	4	4	0	1	6	0.3324
54	4	3	4	7	3.25	0.761905	16	0.5174
51	4	3	4	9	11.083333	0.555556	20	0.5827



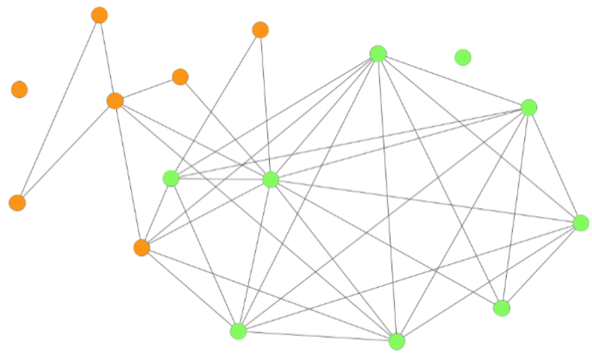
Cohorts 2 and 5: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
35	2	1	2	0	0	0	0	0
65	2	2	2	2	0	1	1	0.0404
22	2	3	4	4	31	0.166667	1	0.0862
6	2	1	4	1	0	0	0	0.0135
40	2	3	4	1	0	0	0	0.0268
30	5	1	3	2	0	1	1	0.3869
29	5	3	4	3	5.833333	0.666667	2	0.5735
34	5	3	4	3	30	0.333333	1	0.4493
41	5	3	4	4	15.666667	0.5	3	0.7049
19	5	1	4	4	6.666667	0.5	3	0.6901
26	5	1	4	3	2.833333	0.666667	2	0.5541
55	5	2	2	3	10	0.333333	1	0.0436
56	5	3	2	2	30	0	0	0.1665



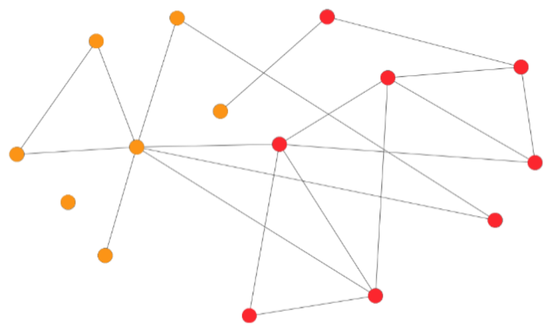
Cohorts 3 and 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
17	3	1	3	6	23	0.333333	5	0.2494
57	3	3	4	6	3.7	0.733333	11	0.4037
50	3	3	3	2	0	1	1	0.1463
12	3	3	2	2	0	1	1	0.0435
31	3	1	2	2	0	1	1	0.0435
2	3	1	4	0	0	0	0	0
42	3	2	4	2	0	1	1	0.1253
49	4	3	4	7	5.45	0.714286	15	0.4692
61	4	3	4	7	1.283333	0.761905	16	0.4741
32	4	3	4	6	0.5	0.866667	13	0.4235
63	4	3	4	8	2.233333	0.714286	20	0.5263
25	4	3	3	6	2.2	0.666667	10	0.391
37	4	1	2	0	0	0	0	0
9	4	3	4	4	0	1	6	0.2997
54	4	3	4	7	0.9	0.809524	17	0.4875
51	4	3	4	11	27.733333	0.436364	24	0.5937



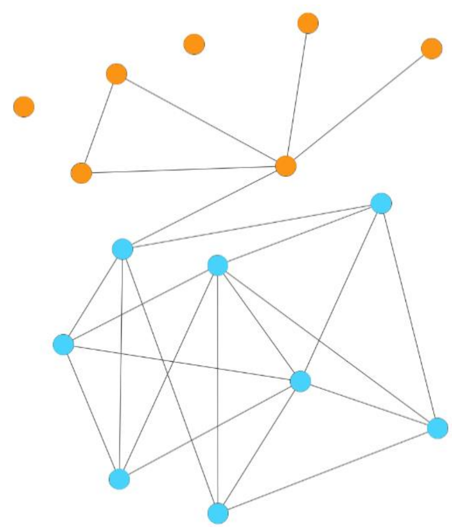
Cohorts 3 and 5: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
17	3	1	3	7	48	0.142857	3	0.6313
57	3	3	4	1	0	0	0	0.1705
50	3	3	3	1	0	0	0	0.0206
12	3	3	2	2	0	1	1	0.2337
31	3	1	2	2	0	1	1	0.2337
2	3	1	4	0	0	0	0	0
42	3	2	4	2	0	1	1	0.2337
30	5	1	3	2	0	1	1	0.3328
29	5	3	4	3	8.5	0.666667	2	0.3853
34	5	3	4	3	22	0.333333	1	0.2618
41	5	3	4	4	19	0.5	3	0.5076
19	5	1	4	5	28	0.4	4	0.657
26	5	1	4	4	13.5	0.5	3	0.575
55	5	2	2	2	0	1	1	0.2337
56	5	3	2	2	12	0	0	0.0763



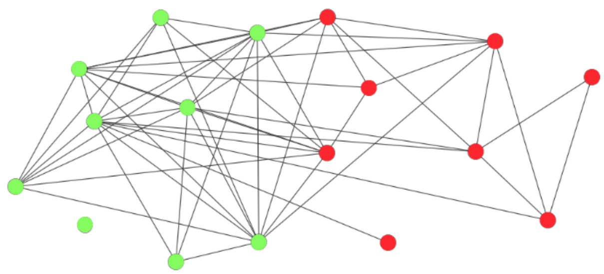
Cohorts 3 and 6: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
17	3	1	3	5	37	0.1	1	0.1123
57	3	3	4	1	0	0	0	0.0241
50	3	3	3	0	0	0	0	0
12	3	3	2	2	0	1	1	0.0307
31	3	1	2	2	0	1	1	0.0307
2	3	1	4	0	0	0	0	0
42	3	2	4	1	0	0	0	0.0241
48	6	3	4	4	3	0.666667	4	0.4531
10	6	4	4	4	6	0.5	3	0.455
1	6	3	4	6	2.583333	0.533333	8	0.6224
4	6	3	4	6	2.583333	0.533333	8	0.6224
60	6	1	4	4	0.25	0.833333	5	0.4623
45	6	1	3	4	6	0.5	3	0.455
13	6	3	4	4	3	0.666667	4	0.4531
43	6	1	2	5	36.583333	0.1	1	0.4138



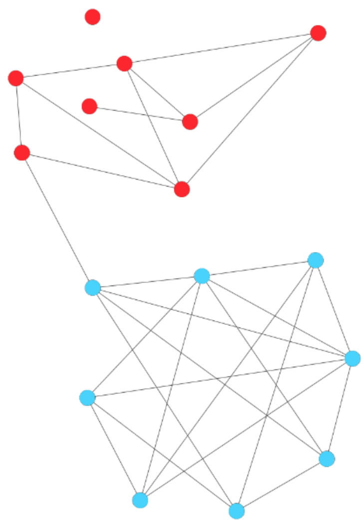
Cohorts 4 and 5: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
51	4	3	4	11	28.887179	0.436364	24	0.4809
61	4	3	4	10	5.657692	0.622222	28	0.5063
63	4	3	4	10	5.657692	0.622222	28	0.5063
54	4	3	4	9	5.914103	0.638889	23	0.4635
49	4	3	4	9	4.674359	0.666667	24	0.4663
32	4	3	4	7	0.4	0.904762	19	0.4039
9	4	3	4	5	0	1	10	0.2982
25	4	3	3	4	0	1	6	0.2511
37	4	1	2	0	0	0	0	0
56	5	3	2	8	4.041026	0.714286	20	0.4264
29	5	3	4	7	3.347436	0.619048	13	0.3431
41	5	3	4	7	6.570513	0.47619	10	0.3059
19	5	1	4	6	10.666667	0.4	6	0.2279
26	5	1	4	4	4.733333	0.5	3	0.1362
34	5	3	4	4	0.45	0.666667	4	0.1978
55	5	2	2	1	0	0	0	0.0617
30	5	1	3	2	0	1	1	0.0467



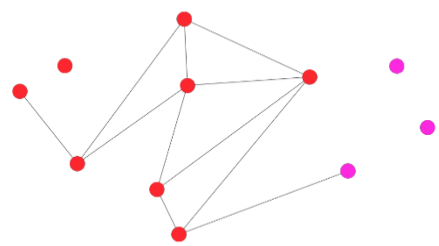
Cohorts 5 and 6: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
30	5	1	3	3	48	0.333333	1	0.1192
29	5	3	4	3	7	0.666667	2	0.0158
34	5	3	4	3	13	0.333333	1	0.0087
41	5	3	4	4	15.5	0.5	3	0.0228
19	5	1	4	4	26	0.5	3	0.0423
26	5	1	4	3	10.5	0.666667	2	0.0395
55	5	2	2	0	0	0	0	0
56	5	3	2	1	0	0	0	0.0019
48	6	3	4	4	0.5	0.666667	4	0.4533
10	6	4	4	4	1	0.5	3	0.4504
1	6	3	4	6	10.75	0.533333	8	0.6244
4	6	3	4	6	10.75	0.533333	8	0.6244
60	6	1	4	4	0.25	0.833333	5	0.4607
45	6	1	3	4	1	0.5	3	0.4504
13	6	3	4	5	49.5	0.4	4	0.4743
43	6	1	2	4	6.25	0.166667	1	0.3918



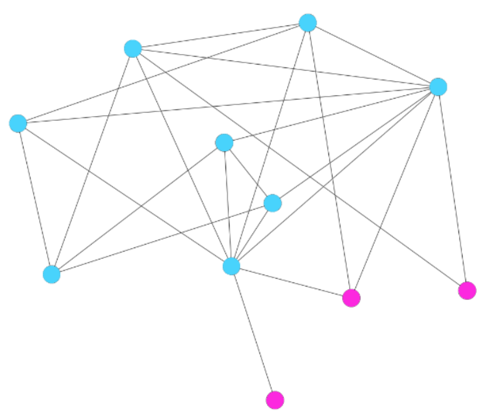
Cohorts 5 and 7: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
30	5	1	3	3	6	0.333333	1	0.4293
29	5	3	4	3	2.333333	0.666667	2	0.5619
34	5	3	4	3	6	0.333333	1	0.4293
41	5	3	4	4	6.166667	0.5	3	0.6946
19	5	1	4	4	6.166667	0.5	3	0.6946
26	5	1	4	3	2.333333	0.666667	2	0.5619
55	5	2	2	0	0	0	0	0
56	5	3	2	1	0	0	0	0.1326
8	7	1	2	0	0	0	0	0
38	7	3	4	1	0	0	0	0.1326
52	7	1	3	0	0	0	0	0



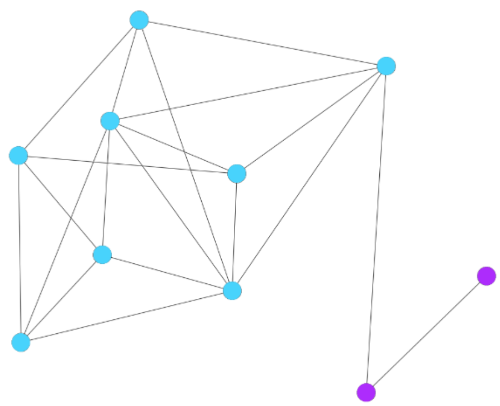
Cohorts 6 and 7: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
48	6	3	4	4	0.95	0.666667	4	0.388
10	6	4	4	4	1.55	0.5	3	0.4068
1	6	3	4	8	10.15	0.392857	11	0.6462
4	6	3	4	8	13.65	0.357143	10	0.6314
60	6	1	4	5	1.116667	0.7	7	0.4843
45	6	1	3	5	4.05	0.4	4	0.4486
13	6	3	4	4	0.95	0.666667	4	0.388
43	6	1	2	4	1.583333	0.166667	1	0.3189
8	7	1	2	3	0	1	3	0.3444
38	7	3	4	1	0	0	0	0.1234
52	7	1	3	2	0	1	1	0.214



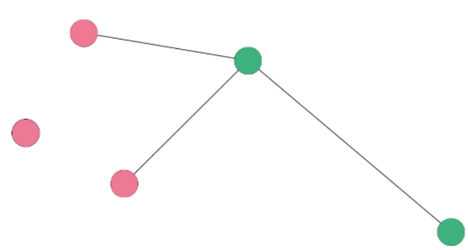
Cohorts 6 and 8: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
48	6	3	4	4	0.5	0.666667	4	0.4474
10	6	4	4	4	2	0.5	3	0.456
1	6	3	4	6	4.583333	0.533333	8	0.626
4	6	3	4	6	4.583333	0.533333	8	0.626
60	6	1	4	5	14.25	0.5	5	0.4874
45	6	1	3	4	2	0.5	3	0.456
13	6	3	4	4	0.5	0.666667	4	0.4474
43	6	1	2	4	1.583333	0.166667	1	0.3873
20	8	1	1	2	8	0	0	0.1095
18	8	3	3	1	0	0	0	0.0235



Cohorts 7 and 8: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

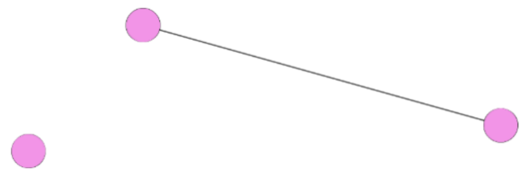
Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
8	7	1	2	1	0	0	0	0.5774
38	7	3	4	0	0	0	0	0
52	7	1	3	1	0	0	0	0.5774
20	8	1	1	3	3	0	0	1
18	8	3	3	1	0	0	0	0.5774



APPENDIX C

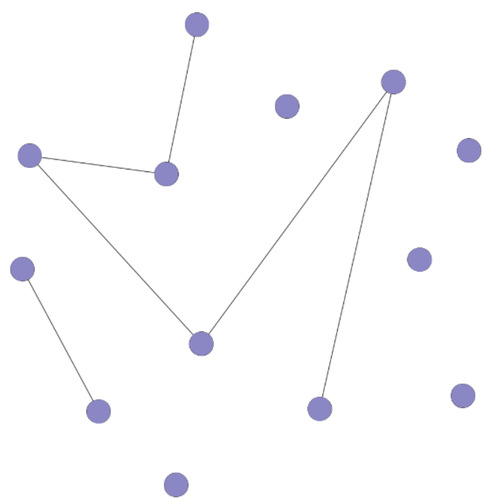
Semester 1: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	1	0	0	0	1
28	1	1	1	1	0	0	0	1
20	8	1	1	0	0	0	0	0



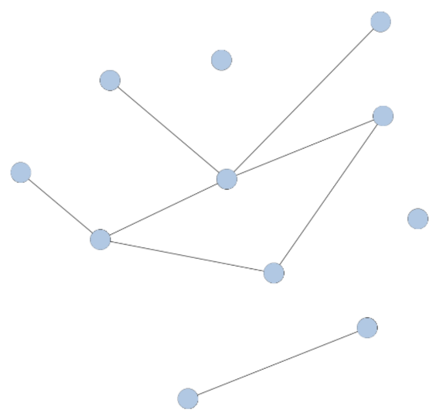
Semester 2: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
21	1	1	2	1	0	0	0	0
33	1	2	2	2	4	0	0	0.591
62	1	1	2	1	0	0	0	0.328
23	1	2	2	1	0	0	0	0
35	2	1	2	0	0	0	0	0
65	2	2	2	2	4	0	0	0.591
12	3	3	2	2	6	0	0	0.737
31	3	1	2	2	6	0	0	0.737
37	4	1	2	0	0	0	0	0
55	5	2	2	1	0	0	0	0.328
56	5	3	2	0	0	0	0	0
43	6	1	2	0	0	0	0	0
8	7	1	2	0	0	0	0	0



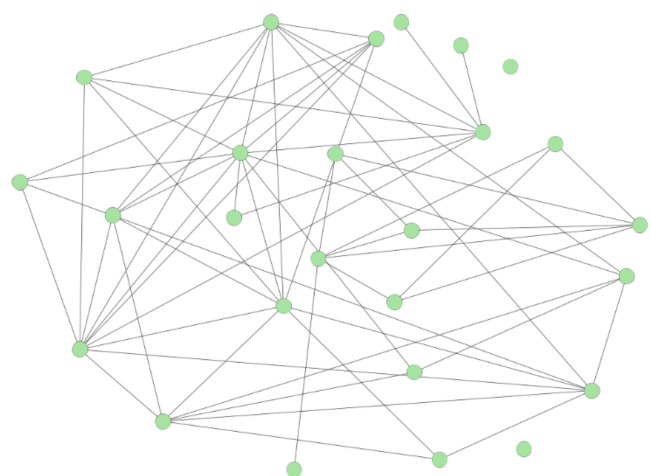
Semester 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
47	1	3	3	4	10	0	0	0.8052
14	1	3	3	1	0	0	0	0.3389
24	1	2	3	1	0	0	0	0.3389
27	1	3	3	3	6.5	0	0	0.6777
17	3	1	3	1	0	0	0	0.2852
50	3	3	3	2	1	0	0	0.5199
25	4	3	3	2	1.5	0	0	0.5577
30	5	1	3	0	0	0	0	0
45	6	1	3	1	0	0	0	0
52	7	1	3	1	0	0	0	0
18	8	3	3	0	0	0	0	0



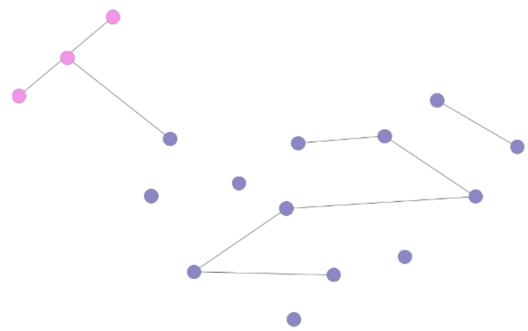
Semester 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
3	1	1	4	0	0	0	0	0
22	2	3	4	7	30.95238	0.333333	7	0.279
6	2	1	4	0	0	0	0	0
40	2	3	4	1	0	0	0	0.0383
57	3	3	4	5	1.178571	0.9	9	0.3155
2	3	1	4	1	0	0	0	0.0383
42	3	2	4	2	0	1	1	0.1081
49	4	3	4	9	10.20714	0.583333	21	0.4877
61	4	3	4	8	3.367857	0.678571	19	0.461
32	4	3	4	6	0.625	0.866667	13	0.3758
63	4	3	4	10	15.54643	0.555556	25	0.5327
9	4	3	4	4	0	1	6	0.2577
54	4	3	4	9	9.92619	0.583333	21	0.4922
51	4	3	4	11	28.60119	0.4	22	0.5094
29	5	3	4	7	3.72381	0.619048	13	0.3736
34	5	3	4	3	0	1	3	0.1637
41	5	3	4	7	5.595238	0.47619	10	0.3324
19	5	1	4	5	1.909524	0.5	5	0.2548
26	5	1	4	3	0.366667	0.666667	2	0.1504
48	6	3	4	3	0	1	3	0
10	6	4	4	3	0	1	3	0
1	6	3	4	5	2	0.6	6	0
4	6	3	4	6	7	0.4	6	0
60	6	1	4	3	0	1	3	0
13	6	3	4	3	0	1	3	0
38	7	3	4	1	0	0	0	0



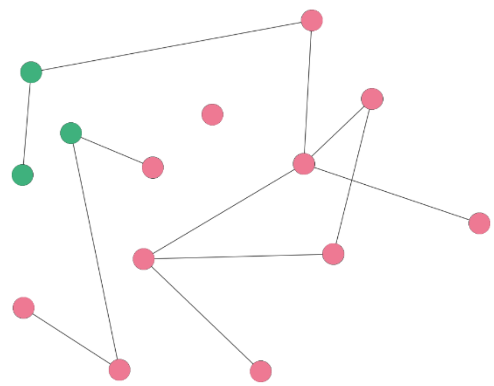
Semesters 1 and 2: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	1	0	0	0	0
28	1	1	1	1	0	0	0	0
20	8	1	1	1	0	0	0	0
21	1	1	2	1	0	0	0	0
33	1	2	2	2	4	0	0	0.591
62	1	1	2	1	0	0	0	0.328
23	1	2	2	1	0	0	0	0
35	2	1	2	0	0	0	0	0
65	2	2	2	2	4	0	0	0.591
12	3	3	2	2	6	0	0	0.737
31	3	1	2	2	6	0	0	0.737
37	4	1	2	0	0	0	0	0
55	5	2	2	1	0	0	0	0.328
56	5	3	2	0	0	0	0	0
43	6	1	2	0	0	0	0	0
8	7	1	2	1	0	0	0	0



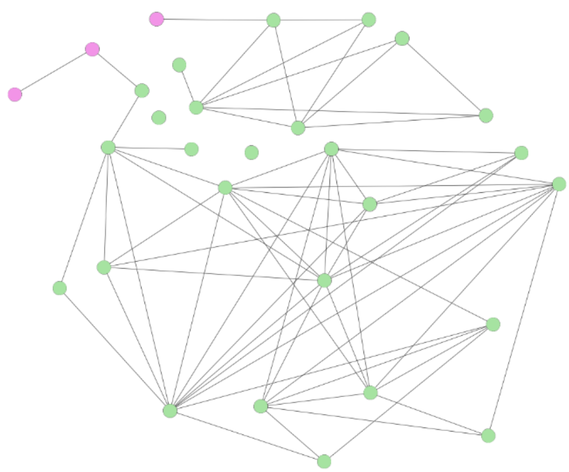
Semesters 1 and 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	2	7	0	0	0.2101
28	1	1	1	1	0	0	0	0.0871
20	8	1	1	2	2	0	0	0
47	1	3	3	4	20	0	0	0.8022
14	1	3	3	1	0	0	0	0.3325
24	1	2	3	2	12	0	0	0.4196
27	1	3	3	3	9.5	0	0	0.647
17	3	1	3	1	0	0	0	0.2682
50	3	3	3	2	1	0	0	0.4904
25	4	3	3	2	2.5	0	0	0.5358
30	5	1	3	0	0	0	0	0
45	6	1	3	1	0	0	0	0
52	7	1	3	2	2	0	0	0
18	8	3	3	1	0	0	0	0



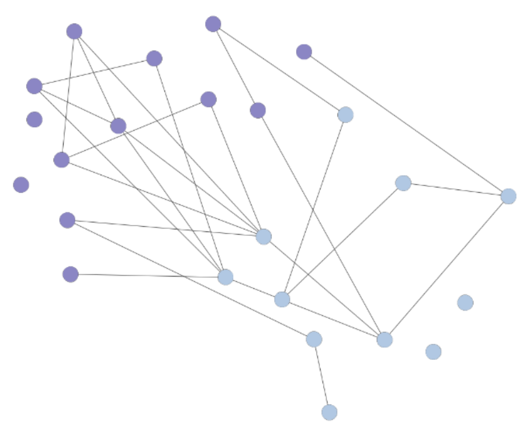
Semesters 1 and 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
36	1	1	1	1	0	0	0	0.0007
28	1	1	1	2	17	0	0	0.0055
20	8	1	1	1	0	0	0	0
3	1	1	4	0	0	0	0	0
22	2	3	4	7	60.952381	0.333333	7	0.2791
6	2	1	4	0	0	0	0	0
40	2	3	4	1	0	0	0	0.0383
57	3	3	4	5	1.964286	0.9	9	0.3155
2	3	1	4	2	32	0	0	0.039
42	3	2	4	2	0	1	1	0.1081
49	4	3	4	9	11.35	0.583333	21	0.4877
61	4	3	4	8	3.367857	0.678571	19	0.461
32	4	3	4	6	0.625	0.866667	13	0.3758
63	4	3	4	10	22.236905	0.555556	25	0.5327
9	4	3	4	4	0	1	6	0.2576
54	4	3	4	9	14.330952	0.583333	21	0.4922
51	4	3	4	11	34.720238	0.4	22	0.5094
29	5	3	4	7	4.295238	0.619048	13	0.3736
34	5	3	4	3	0	1	3	0.1637
41	5	3	4	7	5.880952	0.47619	10	0.3324
19	5	1	4	5	1.909524	0.5	5	0.2548
26	5	1	4	3	0.366667	0.666667	2	0.1504
48	6	3	4	3	0	1	3	0
10	6	4	4	3	0	1	3	0
1	6	3	4	5	3	0.6	6	0
4	6	3	4	6	9	0.4	6	0
60	6	1	4	4	6	0.5	3	0
13	6	3	4	3	0	1	3	0
38	7	3	4	1	0	0	0	0



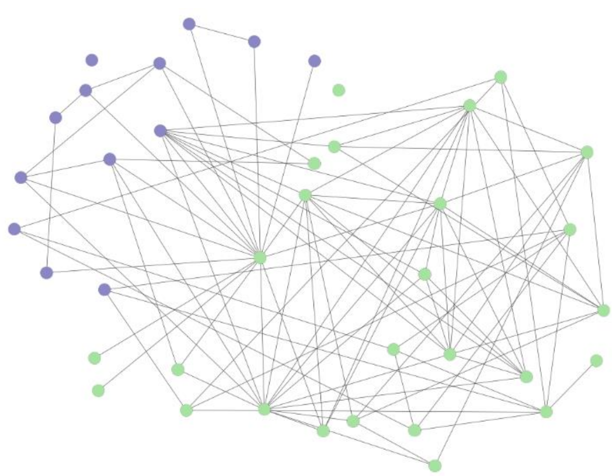
Semesters 2 and 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
21	1	1	2	2	1	0	0	0.0647
33	1	2	2	3	4	0.666667	2	0.3584
62	1	1	2	2	0	1	1	0.2333
23	1	2	2	2	10.5	0	0	0.1209
35	2	1	2	1	0	0	0	0.1304
65	2	2	2	3	0.5	0.666667	2	0.417
12	3	3	2	4	31	0.333333	2	0.5608
31	3	1	2	3	2.5	0.666667	2	0.4726
37	4	1	2	0	0	0	0	0
55	5	2	2	2	0	1	1	0.3116
56	5	3	2	1	0	0	0	0.0436
43	6	1	2	2	34	0	0	0.2108
8	7	1	2	0	0	0	0	0
47	1	3	3	4	45.5	0	0	0.3
14	1	3	3	2	6.5	0	0	0.1047
24	1	2	3	5	41.166667	0.2	2	0.4544
27	1	3	3	4	67.5	0	0	0.3564
17	3	1	3	6	84.833333	0.2	3	0.6686
50	3	3	3	3	22.333333	0	0	0.1521
25	4	3	3	2	5.666667	0	0	0.1298
30	5	1	3	0	0	0	0	0
45	6	1	3	2	18	0	0	0.066
52	7	1	3	1	0	0	0	0.0189
18	8	3	3	0	0	0	0	0



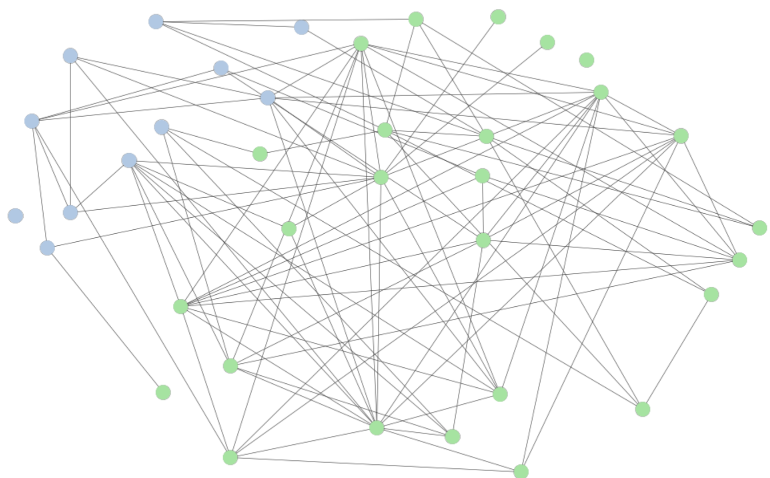
Semesters 2 and 3: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
21	1	1	2	2	0	1	1	0.0491
33	1	2	2	2	0.5	0	0	0.0118
62	1	1	2	2	11.333333	0	0	0.0446
23	1	2	2	2	0	1	1	0.0491
35	2	1	2	0	0	0	0	0
65	2	2	2	4	2.483333	0.666667	4	0.13
12	3	3	2	3	13.666667	0.333333	1	0.0528
31	3	1	2	4	8.781746	0.333333	2	0.0683
37	4	1	2	1	0	0	0	0.0431
55	5	2	2	5	17.384921	0.5	5	0.1385
56	5	3	2	9	26.1	0.638889	23	0.424
43	6	1	2	3	0.666667	0.333333	1	0
8	7	1	2	3	0	1	3	0
3	1	1	4	0	0	0	0	0
22	2	3	4	16	197.637302	0.141667	17	0.3566
6	2	1	4	2	0.333333	0	0	0.025
40	2	3	4	1	0	0	0	0.0431
57	3	3	4	5	1.966667	0.9	9	0.2745
2	3	1	4	1	0	0	0	0.0431
42	3	2	4	3	0	1	3	0.1217
49	4	3	4	10	8.43254	0.6	27	0.4463
61	4	3	4	9	4.215873	0.694444	25	0.4271
32	4	3	4	7	0.4	0.904762	19	0.3668
63	4	3	4	11	32.802381	0.581818	32	0.4952
9	4	3	4	5	0	1	10	0.2691
54	4	3	4	10	21.82381	0.6	27	0.4595
51	4	3	4	14	77.029365	0.362637	33	0.5112
29	5	3	4	7	2.283333	0.619048	13	0.2977
34	5	3	4	4	0.45	0.666667	4	0.1728
41	5	3	4	7	4.75	0.47619	10	0.2608
19	5	1	4	5	2.78254	0.5	5	0.1992
26	5	1	4	3	0.842857	0.666667	2	0.1175
48	6	3	4	4	1.285714	0.666667	4	0
10	6	4	4	4	2.428571	0.5	3	0
1	6	3	4	6	3.428571	0.533333	8	0
4	6	3	4	7	10.428571	0.380952	8	0
60	6	1	4	4	0.47619	0.833333	5	0
13	6	3	4	4	1.285714	0.666667	4	0
38	7	3	4	1	0	0	0	0



Semesters 3 and 4: Gephi data table, Gephi social network graph, and MATLAB eigenvector centrality

Id	Cohort	Subject	Semesters	Degree	Betweenness	Clustering	Triangles	Matlabeig
47	1	3	3	6	29.179927	0.133333	2	0.1666
14	1	3	3	3	34.527632	0	0	0.0646
24	1	2	3	3	5.957576	0.333333	1	0.0666
27	1	3	3	4	4.405263	0.333333	2	0.1167
17	3	1	3	8	56.452913	0.392857	11	0.2784
50	3	3	3	4	1.048485	0.666667	4	0.1624
25	4	3	3	8	8.527197	0.571429	16	0.3489
30	5	1	3	4	225.5	0.166667	1	0.0507
45	6	1	3	4	1.5	0.666667	4	0.0005
52	7	1	3	2	0	1	1	0.0002
18	8	3	3	0	0	0	0	0
3	1	1	4	0	0	0	0	0
22	2	3	4	12	116.806556	0.272727	18	0.362
6	2	1	4	1	0	0	0	0.0078
40	2	3	4	1	0	0	0	0.0436
57	3	3	4	7	2.334957	0.761905	16	0.3457
2	3	1	4	1	0	0	0	0.0436
42	3	2	4	3	0	1	3	0.1398
49	4	3	4	10	17.639314	0.511111	23	0.4189
61	4	3	4	9	6.067026	0.611111	22	0.3981
32	4	3	4	7	5.442747	0.666667	14	0.3194
63	4	3	4	11	22.220884	0.545455	30	0.4741
9	4	3	4	4	0	1	6	0.2062
54	4	3	4	11	56.628423	0.509091	28	0.468
51	4	3	4	14	112.803194	0.384615	35	0.5213
29	5	3	4	7	11.756061	0.619048	13	0.2878
34	5	3	4	4	1.616374	0.666667	4	0.1247
41	5	3	4	8	39.802725	0.392857	11	0.2622
19	5	1	4	7	151.202485	0.380952	8	0.2446
26	5	1	4	5	84.580263	0.5	5	0.1634
48	6	3	4	3	0	1	3	0.0012
10	6	4	4	3	0	1	3	0.0005
1	6	3	4	7	59.833333	0.428571	9	0.0014
4	6	3	4	7	61.833333	0.380952	8	0.0021
60	6	1	4	4	0.333333	0.833333	5	0.0005
13	6	3	4	4	143	0.5	3	0.0067
38	7	3	4	2	39	0	0	0.0064



BIBLIOGRAPHY

[1] U. Brandes, A Faster Algorithm for Betweenness Centrality, *The Journal of Mathematical Sociology*, **25**, 163-177, (2001).

[2] N. Brown, Eigenvector Node Centrality MATLAB program, 2018

[3] Desmos: About Us https://www.desmos.com/about (updated 2019)

[4] D. Easley, J. Kleinberg, *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*, Cambridge University Press, 2010.

[5] Gephi: About https://gephi.org/about/ (updated 2017)

[6] N. Ghali, M. Panda, A.E. Hassanien, A. Abraham, V. Snasel, *Computational Social Networks*, Springer, London, 2012.

[7] R. Hanneman, M. Riddle, *Introduction to Social Network Methods*, University of California, Riverside, 2005.

[8] B. Johnson, B. Down, R. Le Cornu, J. Peters, A. Sullivan, J. Pearce, J. Hunter, Promoting early career teacher resilience: a framework for understanding and acting, *Teachers and Teaching*, **20**, 530-546, (2014).

[9] R. Larson, *Elementary Linear Algebra 7th Edition*, Cengage Learning, 2013

[10] M.M. Makagon, B. McCowan, J.A. Mench, How can social network analysis contribute to social behavior research in applied ethology?, *Applied Animal Behaviour Science.*, **138**, 152-161, (2012).

[11] Mathworks: What is MATLAB? https://www.mathworks.com/discovery/what-is-MATLAB.html (updated 2019)

[12] National Science Foundation. Robert Noyce Teacher Scholarship Program (nsf14508) https://www.nsf.gov/pubs/2014/nsf14508/nsf14508.htm (updated 11/7/06)

[13] B. Ruhnau, Eigenvector Centrality – a node-centrality?, *Social Networks*, **22**, 357-365 (2000).

[14] R. Zafarani, M.A. Abbasi, H. Liu, *Social Media Mining: An Introduction*, Cambridge University Press, 2014.

GLOSSARY

Adjacency matrix: matrix *A* such that

$$A = [a_{mn}], where\ a_{mn} = \begin{cases} 1 \text{ if node } m \text{ is adjacent to node } n \\ 0 \text{ otherwise} \end{cases}$$

Adjacent: two nodes that are connected with an edge

Betweenness centrality: the extent that other nodes depend on *m* as a transmitter of information; for node *m* is the sum of the number of shortest paths from *k* to *n* that go through *m* divided by the number of shortest paths from *k* to *n* [1]

Bridge: An edge between nodes *m* and *n* is a bridge if when it is deleted *m* and *n* lie in two different components [4]

Clustering coefficient: probability that two random friends of *m* are friends with each other [4]

Component: connected part of the graph [4]

Connected: if for every pair of nodes, there is a path between them [4]

Degree: number of edges associated with node *m* [10]

Directed graph: consists of a set of nodes together with a set of directed edges where the direction of the edge is important [4]

Edge: link that connects a pair of nodes [4]

Eigenvector centrality: influence of a node in the network [13]

Euclidean norm: $\|\boldsymbol{v}\| = \sqrt{v_1^2 + v_2^2 + \cdots + v_n^2}$ [9]

Graph: way of specifying relationships among a collection of nodes [4]

Lower triangular matrix: matrix where all the entries above the main diagonal are zero [9]

Mutual Friends Matrix: matrix where the diagonal entries of the matrix correspond to the number of Facebook friends within the group—this number is also the degree of that node in the entire network. The entry in row *m* column *n* represents the number of mutual Facebook friends between person *m* and person *n*.

Node: an object, the set of nodes in a graph is denoted as V [4]

Path: sequence of nodes such that each consecutive pair in the sequence is connected by an edge [4]

Social network: the nodes are people and the edges represent a social interaction between the people [4]

Symmetric matrix: square matrix such that $A = A^T$ [9]

Triangle: when three nodes and three edges form a triangle [4]

Undirected graph: a graph with no direction on the edges [4]