Investigations in Network Medicine

Nicole DiPasquale, Sebastian Cioaba, Muge Capan

University of Delaware, Department of Mathematics, Newark, DE 19717

Introduction

Network Medicine is a new and evolving field that utilizes mathematical concepts to create biological networks in an attempt to better prevent, diagnose and treat various diseases. This research project focused on applying the concepts of network medicine to better understand and detect the early warning signs of a deteriorating patient. With the help of Christiana Care Hospital, three randomly selected sets of data containing 50 de-identified inpatients was organized into a vector with the end goal being to cluster the patients into groups. For example, Figure 1 illustrates the clustering of a graph network and Figure 2 illustrates the separation of three different types of points into similar groups.

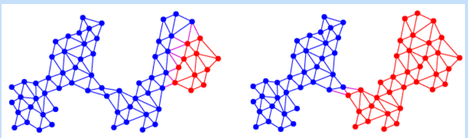


Figure 1

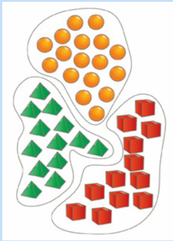


Figure 2

Step 1: Measure of Similarity

There is no proven best clustering method so we chose to use the Euclidean distance, which measures the similarity between two data points and is commonly used in clustering analysis. If two data points, p and q, are given in n-dimensional space, then the distance from p to q is given by

$$\mathrm{d}(\mathbf{p},\mathbf{q}) = \mathrm{d}(\mathbf{q},\mathbf{p}) = \sqrt{(q_1-p_1)^2+(q_2-p_2)^2+\cdots+(q_n-p_n)^2}$$

The Euclidean distance was measured between each patient, represented by the data points, and among their characteristics, represented by the n-dimensions of each point. Each dimension represented a specific static characteristic such as age, gender and comorbidity-related features of the individual patient. These values were then organized into a *m x m* matrix, with *m* representing the number of patients, also known as the similarity matrix.

```
[                0.0  4.358898943540674  8.717797887081348
[  4.358898943540674                0.0                7.0
[  8.717797887081348                7.0                0.0
[7.3484692283495345  5.385164807134504   4.47213595499958
[5.5677643628300215   6.48074069840786  12.68857754044952
[                3.0 3.1622776601683795  9.539392014169456
[  6.082762530298219  8.366600265340756 14.247806848775006
```

Figure 2 shows the first three columns of the similarity matrix.

Work Cited

[1] Barabási, A., N. Gulbahce, and J. Loscalzo. "Network Medicine: A Network-based Approach to Human Disease." *Nat Rev Genet Nature Reviews Genetics* 12.1 (2011): 56-68.

[2] Benson-Putnins, D., M. Bonfardin, M. Magnoni, and D. Martin. "Spectral Clustering and Visualization: A Novel Clustering of Fisher's Iris Data Set." *SIURO SIAM Undergraduate Research Online* 4 (2011): 1-15. Web.

[3] Luxburg, U. "A Tutorial on Spectral Clustering." *Statistics and Computing Stat Comput* 17.4 (2007): 395-416. Web.

[4] Royal College of Physicians. *National Early Warning Score (NEWS): Standardising the assessment of acute- illness severity in the NHS.* Report of a working party. London: RCP, 2012.

[5] Benson-Putnins, D., M. Bonfardin, M. Magnoni, and D. Martin. "Spectral Clustering and Visualization: A Novel Clustering of Fisher's Iris Data Set." (n.d.): n.

Step 2: Laplacian Matrix

There are various types of Laplacian matrices, but for this research project we defined our matrix as

$$\mathbf{L} = \mathbf{D} - \mathbf{S}$$

where **D** is the degree matrix and **S** is the similarity matrix. Because **L** always has row sums of zero, we determined **D** by putting the row sums of **S** on the diagonal and zeros in every other place. Thus, **L** contains only positive numbers on the diagonal while every other place contains the negative value of **S**. The Laplacian matrix is a very special kind of matrix because it is symmetric and positive semi-definite, which means that its eigenvalues are real and non-negative. This makes determining the eigenvalues and eigenvectors of the matrix simpler, which is important for the Fiedler's method.

```
[   35.0756929520998  -4.358898943540674  -8.717797887081348
[  -4.358898943540674  34.753682374592174                -7.0
[  -8.717797887081348                -7.0   56.66571024547491
[-7.3484692283495345  -5.385164807134504   -4.47213595499958
[-5.5677643628300215   -6.48074069840786  -12.68857754044952
[                -3.0 -3.1622776601683795  -9.539392014169456
[  -6.082762530298219  -8.366600265340756 -14.247806848775006
```

Figure 3 shows the first three columns of the Laplacian Matrix.

Step 3: Fiedler's Method

After determining the eigenvalues and eigenvectors of the Laplacian matrix, Fiedler showed that the eigenvector corresponding to the second smallest eigenvalue, called the Fiedler vector, can be used to divide the data points into clusters. By looking at the Fiedler vector, the clusters are formed by placing the rows with the same sign into one cluster and the others into a different cluster.

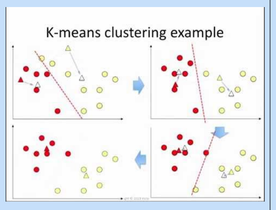
```
(35.84495274290143, [(-0.3341344680416061, -0.4020682350127139,
 0.02337149045875912, 0.013759881284023043, -0.1076561722812856
7, 0.8443655923845704, -0.037638088791747466)], 1)
```

Figure 4 shows Fiedler's Vector

Additional Methods

K-Means

This method is a widely used technique of clustering, but some believe that it is outdated and imprecise and therefore we did not use this method.



Pre-determine how many clusters the data should have by setting *k* equal to the desired number → Divide the data values into the desired number of clusters and calculate the cluster centers → Reassign the data values to the cluster with the closest center, and then recalculate the cluster center

Spectral Clustering

Although a more recent approach to clustering, many believe that spectral clustering outperforms other clustering techniques like the K- Means algorithm. This algorithm is more complex and would be used to further the project.

Construct similarity graph and weighted adjacency matrix → Compute the first *k* eigenvectors and place them into a matrix → Apply the K-Means algorithm to the matrix

Result 1

Characteristic	Cluster 1	Cluster 2
Disposition	0.125 [0.342]	0.088 [0.288]
Sex	0.625 [0.5]	0.412 [0.5]
Age	67.125 [6.228]	62.147 [18.794]
RRT	0.75 [0.856]	0.35 [0.597]
Code	0.125 [0.5]	0.029 [0.171]
XICU	0.688 [1.352]	0.265 [0.511]
CI_Cur	7.375 [1.996]	3.294 [2.355]
CI_HX	9.813 [2.536]	3 [2.934]

Table 1 shows the means and standard deviations between the two clusters characteristics for data 1 with statically significant differences highlighted in yellow.

Result 2

Characteristic	Cluster 1	Cluster 2
Disposition	0.130 [0.344]	0.111 [0.320]
Sex	69.913 [16.200]	69.074 [11.906]
Age	1.739 [1.176]	1.778 [1.05]
RRT	0.391 [0.499]	0.370 [0.565]
Code	0.043 [0.209]	0 [0]
XICU	0.391 [0.722]	0.333 [0.679]
CI_Cur	5.826 [3.950]	4.741 [2.581]
CI_HX	6 [4.359]	4.963 [3.368]

Table 2 shows the means and standard deviations between the two clusters characteristics for data 2

Result 3

Characteristic	Cluster 1	Cluster 2
Disposition	0.182 [0.395]	0.107 [0.315]
Sex	59.591 [0.908]	60.393 [17.203]
Age	1.409 [1.221]	1.46 [0.999]
RRT	0.591 [0.908]	0.393 [0.629]
Code	0 [0]	0.036 [0.189]
XICU	0.409 [0.666]	0.393 [0.875]
CI_Cur	3.909 [3.294]	3.607 [2.544]
CI_HX	3.364 [2.854]	3.75 [2.927]

Table 3 shows the means and standard deviations between the two clusters characteristics for data 3

Future Research

We believe that this project could be furthered by including dynamic characteristics of the patients rather than just focusing on static data. Dynamic values, such as vital signs, could give the researcher a better idea of the patient's deterioration because it would illustrate how the patient was changing over time.