Assignments

Lectures 6, 7: Centrality Measures &
Lectures 8, 9: Spectral Graph Theory

1. **Laplacian of a Graph**
   Let G be a *simple graph*[1] whose adjacency matrix is represented by $\boldsymbol{A}$. The Laplacian matrix (aka the Graph Laplacian) $\boldsymbol{L}$ for the graph G is defined as $\boldsymbol{L} = \boldsymbol{D} - \boldsymbol{A}$, where $\boldsymbol{D}$ is a diagonal matrix whose i^{th} entry corresponds to the degree of the i^{th} node. In other words,

   $$\boldsymbol{D}[i][i] = \sum_{j=1}^{n} \boldsymbol{A}[i][j], \text{ and } \boldsymbol{D}[i][j] = 0 \text{ if } i \neq j$$

   a. For any G, show that 0 is an Eigenvalue of $\boldsymbol{L}$. What is the corresponding unit length Eigenvector?
   b. Suppose G is a disconnected graph with two connected components. Show that the multiplicity of the Eigenvalue 0 is at least 2. What are the corresponding Eigenvectors? Now, prove the following generalization: if G is a union of k connected components, then the Eigenvalue 0 has a multiplicity $\geq k$.
   c. What is the spectrum of the Laplacian of a clique with n vertices?
   d. Prove that for an undirected and unweighted graph G,

   $$\sum_{i=1}^{n} D[i][i] = \sum_{j=1}^{n} \lambda_j^2 - 2e$$

   where λ_j are the Eigenvalues of the Laplacian of G and e is the number of edges.

2. Consider a *path* graph P with n nodes, such that node i is connected to node i-1 and i+1, except for the nodes 1 and n which are only connected to nodes 2 and n-1 respectively.
   a. Compute the eigenvector centrality of each of the nodes of P.
   b. Compute the spectrum of P.
   c. Now consider a directed path P' where there is an edge from node i to node i+1. There is no edge from node n. Compute the spectrum and the eigenvector centralities of the nodes for P'.
   d. Can you generalize the result of (2c): in a weakly connected directed acyclic graph, a node has non-zero eigenvector centrality if and only if its out-degree is 0.

3. The second eigenvector of a graph is used for spectral clustering in the following way: If you order the vertices as per the value of the second eigenvector component, then the nodes having a component greater than the median (of the components) is in one cluster and rest are in another cluster. While answering the following questions assume that the median of the second eigenvector components is 0 so that you can cluster the network just by looking at the sign of the second eigenvector component.
   a. Would it be the case that in a bipartite network, nodes in one partition will have the same sign for the second eigenvector component? Why or why not?
   b. What behavior of the 2^{nd} eigenvector component would you expect for a graph which has two connected components? Explain why would you expect so?
   c. Suppose that in a connected graph there is a node with a very high degree, but a very low clustering coefficient. What can you say about the second eigenvector component of such a node?

[1] A simple graph does not have self-loops and multiple edges between the same pair of nodes. However, a simple graph can be directed and un-directed, as well as weighted and unweighted.

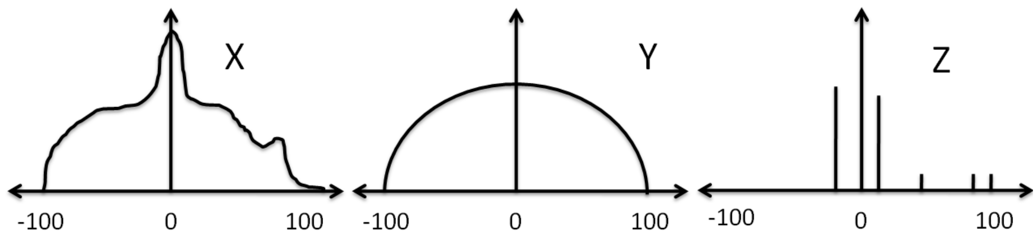
4. **The HITS Algorithm**

Jon Kleinberg came up with an alternative algorithm for ranking the popularity of the webpages, which is known as the HITS (Hyperlink Induced Topic Search). According to this algorithm, every node has two scores: the hub score h_i and the authority score z_i. These scores are defined as follows:

$$h_i = \sum_{j=1}^{n} z_j a_{ij} \qquad\qquad z_i = \sum_{j=1}^{n} h_j a_{ji}$$

where a_{ij} is 1 if there is an edge from node i to node j, else 0. Let $\boldsymbol{A}$ be the adjacency matrix of a graph G, and $\mathbf{h}$ and $\mathbf{z}$ be the vectors representing the hub scores and the authority scores of the nodes of G. How are the vectors $\mathbf{h}$ and $\mathbf{z}$ related to the matrix $\boldsymbol{A}$? (Hint: These are eigenvectors of some matrices closely related to $\boldsymbol{A}$).

5. Shown below are the spectra of the adjacency matrix of three real world networks X, Y and Z.



a. Which of these networks have a large number of structurally similar nodes and why?
b. For which of these networks, you can expect to effectively explain the structure of the real world systems underlying that network in terms of only a few governing factors?
c. Which of them have a topology very similar to a Poisson (or ER) random graph?

6. **Airline Network**

Consider an Airline network, where the nodes are the airports and the weights on the edges represent the average number of passengers travelling between two airports daily. You may assume that the number of passengers travelling from airport 1 to airport 2 is same as that of the passengers travelling from airport 2 to airport 1, so that the network essentially is an undirected, but weighted network. Further assume that the degree of a node in a weighted network is sum of the weights of the edges connected to that node. Suppose that you are planning to introduce a new flight.

How good are the following strategies from the point of view of making profits? Clearly explain your answers.

a. Connecting two airports that has high degree centralities
b. Connecting two airports that has high betweenness centrality
c. Connecting an airport with low shortest path centrality but high degree centrality to one that has high shortest path centrality but low degree centrality
d. Connecting an airport with high eigenvector centrality to one with low eigenvector centrality
e. Connecting two airports with low betweenness centrality and low shortest path centrality.

Suppose that a new airport has been constructed to which no flights are available at this point. You want to introduce a few new flights to/from this airport to existing ones. Which existing airports would you choose to maximize your profit? You may be ignore socio-economic and geographical factors, and focus only on the topological properties of the network.