COM 530500 **Network Science Homework #2**

Due: Thursday, November 11, 2021

No late homework will be accepted.

班級: 資應所

姓名: 鄭程哲

學號: 110065512

Problem 1.(40%) Consider a k-regular undirected network (i.e., a network in which every vertex has degree k) with number of nodes n.

- (a) (10%) Show that the n-vector $\mathbf{1}_n = [1, 1, \ldots, 1]$ is an eigenvector of the adjacency matrix with corresponding eigenvalue k.

- (b) (10%) By making use of the fact that eigenvectors are orthogonal (or otherwise), show that there is no other eigenvector that has all elements positive. [*Note: The Perron-Frobenius theorem says that the eigenvector with all elements positive has the largest eigenvalue, and hence the eigenvector $\mathbf{1}_n$ gives, by definition, the eigenvector centrality of our k-regular network and the centralities are the same for every vertex.*]

- (c) (10%) Find the **Katz centralities** of all vertices in a k-regular network.

- (d) (10%) Find a centrality measure that can give different centralities for different vertices in a regular network. Please provide a specific example.

Solution:

- (a) The result of adjacency matrix multiplying n-vector 1_n is equal to the row sum, which means the degree of each node, i.e., k. Therefore, n-vector 1_n is an eigenvector of the adjacency matrix with corresponding wigenvalue k.

- (b) Eigenvectors are orthogonal if and only if their dot product is zero. If there is other eigenvector v has all elements positive, $1_n \cdot v$ must bigger than 0.

- (c) Katz centralities of all vertices $x = \beta(I - \alpha A)^{-1}1$, where 1 is the vector $(1, 1, 1, ..., 1)^T$, $\alpha \in [0, \frac{1}{k})$, and β is defined by the user.

- (d) It's hard to answer this question. I wrote a python program (code-prob1d.ipynb) to see the centrality value of regular network with $n \in [3, 8], k \in [2, n)$. There is very tiny difference between measures. Take $n = 4, k = 2$ for example, PageRank gives different centralities for different vertices, i.e. 0.25, 0.24999999999999994, 0.24999999999999994, 0.25, respectively. Another example is shown in Figure 1, when $n = 7, k = 2$, harmonic centralities equal to 0.33333333, 0.41666667, 0.41666667, 0.41666667, 0.41666667, 0.33333333, 0.33333333, respectively. This example shows obvious difference between the centralities of vertices. However, this network isn't connected.
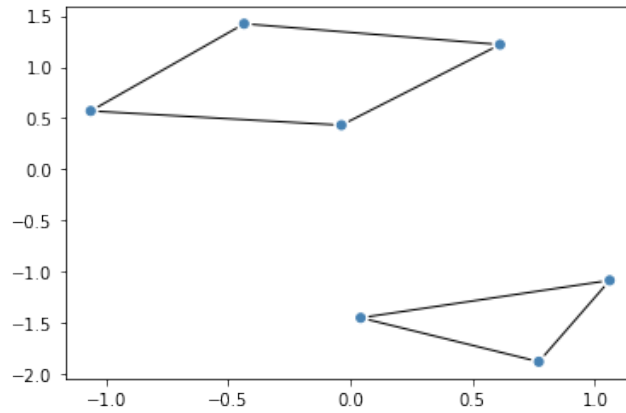
$\diamond$

Figure 1: 2-regular network with 7 nodes.

Problem 2.(60%) Please find a real dataset from the Internet. (**Note: You need to cite the dataset in the reference.**) Note that this dataset should be an undirected network, and the total number of nodes should be greater than 500. **Please do not use the same dataset in Homework #1.**

(a) (10%) Briefly introduce this dataset, and list some basic statistical information, such as the number of nodes, number of edges, average clustering coefficient, diameter, average degree, maximum degree, etc.

(b) (10%) Please visualize the dataset by plotting it.

(c) (20%) Please implement the Katz centrality measure (textbook chapter 7.3 [1]) **without** using the katz_centrality function and the katz_centrality_numpy function provided by NetworkX, and find the top 10 nodes ranked by the Katz centrality measure you've written.

(d) (10%) Please find the top 10 nodes by two other different centrality measures (you can use any packages and functions).

(e) (5%) Are the top 10 nodes ranked by different centrality measures in (c) and (d) the same? Explain why?

(f) (5%) Is there a best centrality measure for ranking this dataset? Explain why?

Solution:

(a) The dataset I used is about real-world animal interaction network[2]. The vertices in the network represent voles while an edge was inserted into the network whenever two voles were caught in at least one common trap over the primary trapping sessions being considered. Basic statistical information is shown below:

- Number of nodes: 1481
- Number of edges: 4569
- Average clustering coefficient: 0.522543
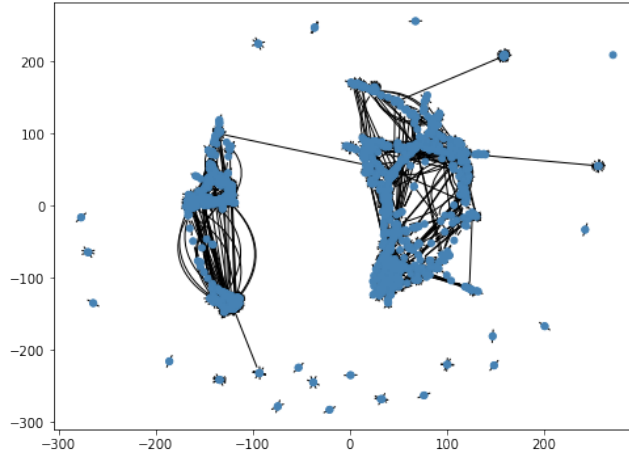- Diameter: 18
- Average degree: 6.17

Figure 2: Dataset visualization.

- Maximum degree: 39

(b) Visualization is shown in Figure 2.

(c) I used the formula taught in class, i.e. $x(t) = \alpha A x(t-1) + \beta$. Iterate it and get the converged Katz centrality.

Top-10 nodes: [13, 62, 358, 93, 1023, 107, 321, 22, 221, 32]

(d) Eigenvector centrality top-10 nodes: [89, 193, 231, 111, 221, 196, 3, 13, 61, 22]

Closeness centrality top-10 nodes: [956, 664, 759, 758, 936, 1298, 498, 662, 663, 499]

Harmonic centrality top-10 nodes: [358, 196, 62, 61, 93, 13, 778, 192, 562, 355]

Betweenness centrality top-10 nodes: [778, 932, 928, 358, 999, 1023, 196, 973, 818, 1157]

(e) The first table below shows the number of same nodes compared to Katz centrality in top-10 and top-100 respectively. These values are quite low and the most similar one is still under 50%. Furthermore, I compare the other four different centrality measures (directly using igraph API) and the result is shown in the second table. The value in each cell represents the number of the same nodes in top-10 and top-100. As we can see from the table, different centrality measures don't have similar results. However, I printed out the centrality values of each method and noticed that these centrality values are quite close. I think one of reason is that the central nodes are too close to each other and there is no significant difference between them. As the figure shown above, most of the nodes are crowded in the center.

	Top-10	Top-100
Eigenvector	3	34
Closeness	0	15
Harmonic	4	48
Betweenness	2	49

	Eigenvector	Closeness	Harmonic	Betweenness
Eigenvector		0 / 8	3 / 44	1 / 19
Closeness			0 / 40	0 / 30
Harmonic				3 / 53
Betweenness				

(f) As I described in (e), most of the centrality values are too close to distinguish even using different measures. Therefore, I think there is no best centrality measure for ranking this dataset.

$\diamondsuit$

References

[1] Newman, Mark. *Networks: An Introduction.* Oxford University Press, 2010.

[2] Ryan A. Rossi and Nesreen K. Ahmed. *The Network Data Repository with Interactive Graph Analytics and Visualization.* `http://networkrepository.com/` 2015