Homework 2

Nathan Bie

1. Consider a connected k-regular undirected network

(i) Show that the uniform vector $\vec{1} = (1,1,1,\dots,1)$ is an eigenvector of the adjacency matrix with eigenvalue k.

(ii) Find the Katz centralities of all nodes in the network as a function of k.

(iii) You should find that, like the eigenvector centrality, the Katz centrality of all nodes are the same. Name a centrality measure that could give different centralities for different nodes in a regular network.

(i) Firstly, we understand that a k-regular undirected network has nodes all with degree k. If we consider the adjacency matrix A in this case, for each row there is k 1's. Therefore, the calculation $A\vec{1} = k\vec{1}$. This means that $\vec{1}$ is an eigenvector. This is specific form of $\vec{x} = \frac{1}{k} A \vec{x}$.

(ii) The Katz centrality of the nodes is defined as

$\vec{x} = (I - \alpha A)^{-1}\vec{1}$. We use the finding $A\vec{x} = k\vec{x}$, which was $A\vec{1} = k\vec{1}$.

We see that $(I - \alpha A)^{-1} = \frac{1}{I - \alpha A}$. We can use series expansion to get $(I - \alpha A)^{-1} = I + \alpha A + \alpha^2 A^2 + \cdots$. So:

$$\vec{x} = (I + \alpha A + \alpha^2 A^2 + \cdots)\vec{1}$$

$$\Leftrightarrow \vec{x} = (\vec{1} + \alpha A\vec{1} + \alpha^2 A^2 \vec{1} + \cdots) = \frac{1}{1 - \alpha\lambda 1} = \frac{1}{1 - \alpha k 1}$$

We therefore get $\vec{x} = (1(I - \alpha k))^{-1}$, a function of k.

We see that this is a vector where are entries are the same.

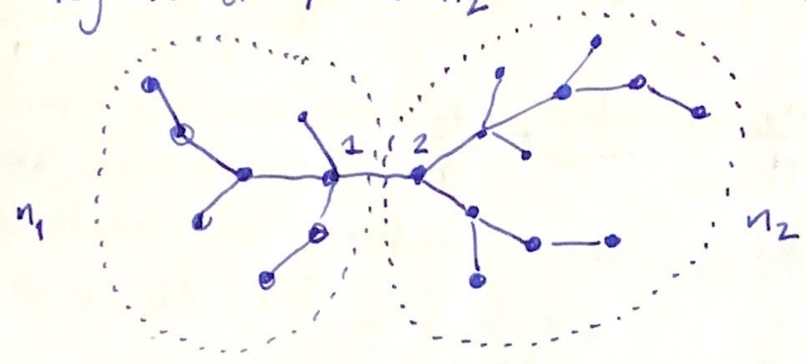
(iii) In the text we see that PageRank, Katz, degree, and eigenvector centralities are similar and all, in this case, ~~are~~ yield, entries that are all the same.

we therefore need to consider the other metrics, such as the closeness and betweenness centralities.

Since we are considering a k-regular network, my suspicion is that betweeness will yield uniform centralities, since all nodes are on the same number of paths, since they are all the same degree.

However, I think that the distances between various nodes are not uniform. As the major piece of the closeness centrality, this brings me to think that this would be a metric that is good.

2. Consider an undirected tree of n nodes. A particular edge in the tree joins nodes 1 and 2 and divides the tree into two disjoint regions of n_1 and n_2 nodes as sketched here:



Show that the closeness centralities c_1 and c_2 of the two nodes, defined as $c_i = \frac{1}{l_i} = \frac{n}{\sum_j d_{ij}}$, $l = \frac{1}{n}\sum_j d_{ij}$, are related by

$$\frac{1}{c_1} + \frac{n_1}{n} = \frac{1}{c_2} + \frac{n_2}{n}$$

The closeness centrality of node i intuitively is the inverse of the mean shortest path, or distance, to every other node j.

what we notice is that, in considering the two portions of the graph, ~~when we~~ consider any distance from node 1 to the right side nodes is one more than the distance from node 2 to any of them. The similar observation, switching nodes 1 and 2, holds for the left side.

Therefore, we see that, since we are summing to n_1 and n_2 on the left and right sides, we add n_1 and n_2 to the total distance respectively.

$$c_1 = \frac{n}{\sum_j d_{ij}}, \quad c_2 = \frac{n}{\sum_j d_{2j}}.$$

so $\frac{1}{c_1} + \frac{n_1}{n} = \frac{\sum_j d_{ij}}{n} + \frac{n_1}{n}$, $\quad \frac{1}{c_2} + \frac{n_2}{n} = \frac{\sum_j d_{2j}}{n} + \frac{n_2}{n}$

$$\Rightarrow \frac{\sum_j d_{ij} + n_1}{n}, \quad \frac{\sum_j d_{2j} + n_2}{n}. \quad \xrightarrow{\text{rear.}}$$

we add as mentioned, and see

$$\sum_i d_{1j} + n_1 = \sum d_{2j} + n_2 \iff \frac{\sum d_{1j} + n_1}{n} = \frac{\sum d_{2j} + n_2}{n}$$

$$\iff \frac{1}{c_1} + \frac{n_1}{n} = \frac{1}{c_2} + \frac{n_2}{n}.$$

3. Consider an undirected tree of n nodes. Suppose that a particular node in the tree has degree k, so its removal would divide the tree into k disjoint regions, and suppose that the sizes of those regions are $n_1, \ldots, n_k$.

(i) Show that the unnormalized betweenness centrality x of the node, defined by $x_i = \sum_{st} \frac{n^i_{st}}{g_{st}}$, where n^i_{st} is the number of paths from s to t through i and g_{st} is the number paths from s to t, is $x = n^2 - \sum_{m=1}^{k} n_m^2$.

(ii) Hence or otherwise calculate the betweenness of the ith node from the end of a "line graph" of n nodes, such as o—o—o—o

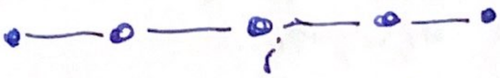
(i) Firstly, we know that in a tree, there is only 1 shortest path between any two nodes. Betweenness centrality intuitively captures how often a node sits on the shortest path between all pairs of nodes in the graph/network.

The total number of shortest paths in the tree of size n is n^2.

When node in question is removed, there's the k disjoint regions, with total of $\sum_{m=1}^{k} n_m$ nodes in them. Each of these has one shortest path to the node, and however these go ~~Ma~~ the node but can proceed to any of the other nodes, which gives $\sum_{m=1}^{k} n_m^2$

Therefore ~~there are~~ we get $n^2 - \sum_{m=1}^{k} n_m^2$ ~~paths~~ as unnormalized betweenness.

(ii) Now considering a line path such as:



~~The betweeness centrality of node i depends on if the number of nodes is even or odd~~

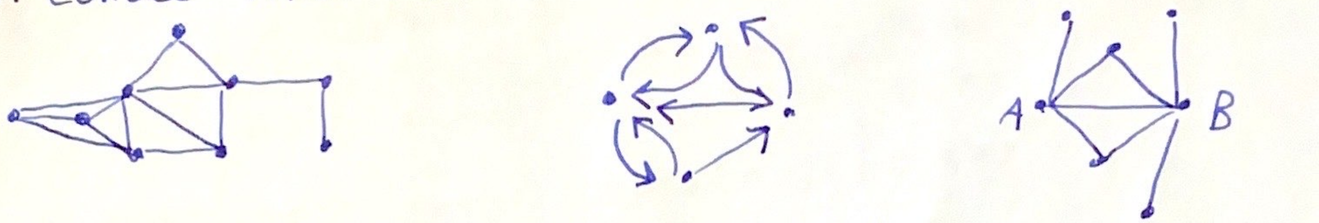
For any node i, we can use the theorem just proven.

There will be two regions if node i is removed.
Note that there is the special cases of $i=1$ (the "first" node in the line) and $i=n$ (the last node), where there aren't two resulting regions, but the formula still works.

using $x = n^2 - \sum_{m=1}^{k} n_m^2$, we get

$$x = n^2 - (i-1)^2 - (n-i)^2$$

4. Consider these three networks.



(i) Find a 3-core in the first network.
(ii) what is the reciprocity of the second network.
(iii) what is the cosine similarity of nodes A and B in the third network.

(i) A 3-core in the network is the leftmost portion

since the definition of a 3-core is a set of nodes where each is joined by an edge to at least 3 others.

(ii) The definition of reciprocity is given as the number of edges that are reciprocated divided by the total number of edges. This concept mainly applies in directed networks. There are 8 total edges, and 6 are reciprocated, so reciprocity $r = \frac{6}{8}$.

(iii) cosine similarity is defined $\sigma_{ij} = \cos\theta = \dfrac{\sum_k A_{ik} A_{kj}}{\sqrt{\sum_k A_{ik}^2}\sqrt{\sum_k A_{jk}^2}}$

Intuitively, this measures the level to which two nodes share the same neighbors in the network. We note that this network is unweighted and simple.

For A and B, we have $\sigma_{AB} = \dfrac{n_{AB}}{\sqrt{k_A k_B}} = \dfrac{3}{\sqrt{4 \cdot 5}} = \dfrac{3}{\sqrt{20}} \cong 0.671$.

5. Consider the following study.

(ii)

		WOMEN				
		Black	Hispanic	white	Other	Total
MEN	Black	0.258	0.016	0.035	0.013	0.323
	Hispanic	0.012	0.157	0.058	0.019	0.247
	white	0.013	0.023	0.306	0.035	0.377
	Other	0.005	0.007	0.024	0.016	0.053
	Total	0.289	0.204	0.423	0.084	

using $Q = \sum_r (e_r - a_r^2)$, what are e_r and a_r for each type?
using the numbers, calculate the modularity of the network with respect to ethnicity.

we consider the definitions of e_r and a_r.

$e_r = \frac{1}{2m} \sum_{ij} A_{ij} \delta_{g_i, r} \delta_{g_j, r}$, which is the fraction of edges that join nodes of type r.

$a_r = \frac{1}{2m} \sum_i k_i \delta_{g_i, r}$, which is the fraction of edges attached to nodes of type r.

For Black: $e_B = 0.258$, $a_B = \frac{0.323 + 0.289}{2} = 0.306$

For white: $e_W = 0.306$, $a_W = \frac{0.377 + 0.423}{2} = 0.4$

For Hispanic: $e_H = 0.157$, $a_H = \frac{0.247 + 0.204}{2} = 0.2255$

For Other: $e_O = 0.016$, $a_O = \frac{0.053 + 0.084}{2} = 0.0685$

we can calculate

$Q_B = 0.258 - 0.306^2 = 0.1644$ $\quad Q_W = 0.306 - 0.4^2 = 0.146$

$Q_H = 0.157 - 0.2255^2 = 0.106$ $\quad Q_O = 0.016 - 0.0685^2$
$= 0.0114$

→ Next page →

once we have these numbers, we follow the formula to get the modularity $Q = \sum_r (e_r - a_r^2)$.

$$Q = 0.1644 + 0.106 + 0.146 + 0.0114 = 0.4278.$$