Jung H Choi

HW 8

12/08/16

**Honor code:** The Virginia Tech honor pledge for assignments is as follows:

**“I have neither given nor received unauthorized assistance on this assignment.”**

**1. (100 points) Network analysis using igraph and igraphdata.**

**1a. Load the karate network in R, and describe the dataset in your own words, in 2-3 lines. Feel free to use help(karate).**

The structure of this network consists of name, citation, and author which is led by John A and Mr. Hi. The vertices of each graph consist of name, faction, color, and label with edges attribute to the weight. There are 34 vertices and 78 undirected edges.

**1b. Is this a directed network? How many vertices and edges are there?**

> kar = karate

> is.directed(kar)

[1] FALSE

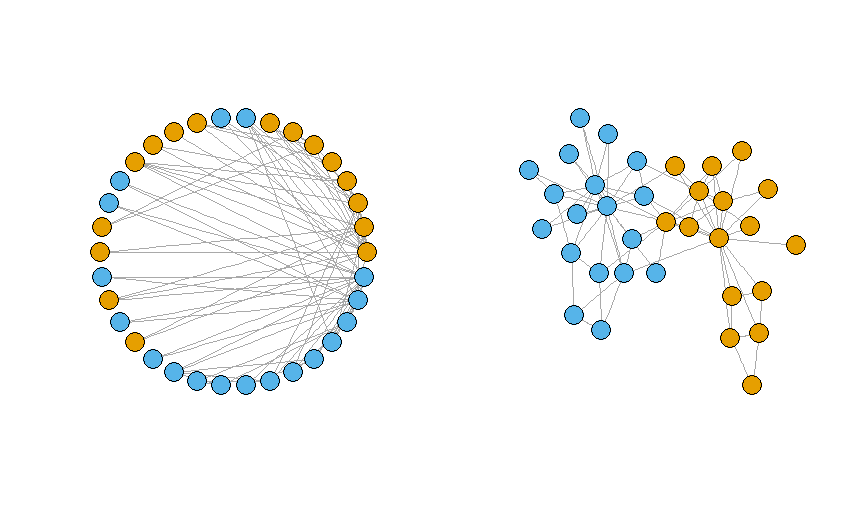
> vcount(kar)

[1] 34

> ecount(kar)

[1] 78

**1c. Plot the network using the circle layout and the graphopt layout.**



**1d. What are the vertex attributes for this network? Plot the network using the graphoptlayout and different vertex colors according to the vertex attribute Faction.**

> vertex\_attr(kar)

$Faction

[1] 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2

$name

[1] "Mr Hi" "Actor 2" "Actor 3" "Actor 4" "Actor 5" "Actor 6" "Actor 7" "Actor 8" "Actor 9"

[10] "Actor 10" "Actor 11" "Actor 12" "Actor 13" "Actor 14" "Actor 15" "Actor 16" "Actor 17" "Actor 18"

[19] "Actor 19" "Actor 20" "Actor 21" "Actor 22" "Actor 23" "Actor 24" "Actor 25" "Actor 26" "Actor 27"

[28] "Actor 28" "Actor 29" "Actor 30" "Actor 31" "Actor 32" "Actor 33" "John A"

$label

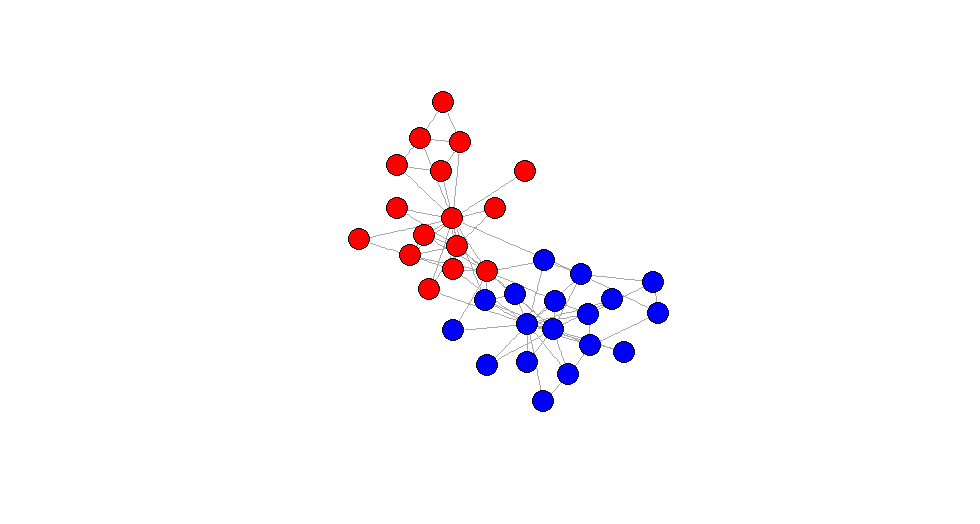
[1] "H" "2" "3" "4" "5" "6" "7" "8" "9" "10" "11" "12" "13" "14" "15" "16" "17" "18" "19" "20"

[21] "21" "22" "23" "24" "25" "26" "27" "28" "29" "30" "31" "32" "33" "A"

$color

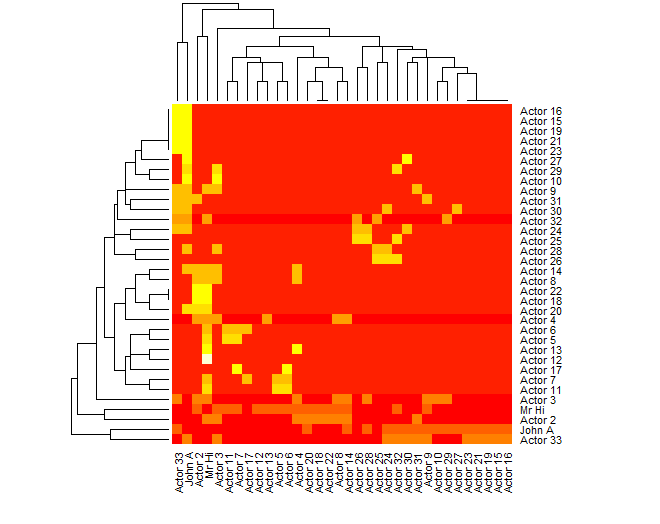
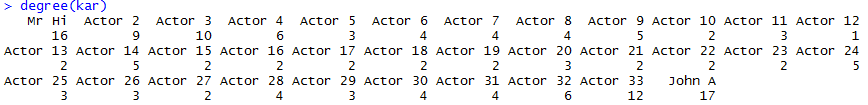
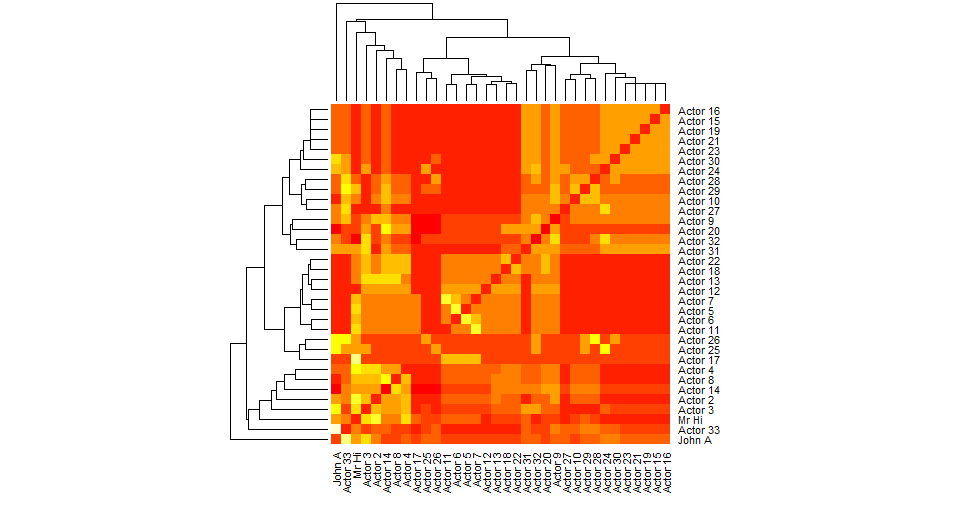
[1] 1 1 1 1 1 1 1 1 2 2 1 1 1 1 2 2 1 1 2 1 2 1 2 2 2 2 2 2 2 2 2 2 2 2

There are Faction, name, label, and color vertex attributes for the network. Among those attributes, Faction contains two categories which represents John A and Mr. Hi’s club.



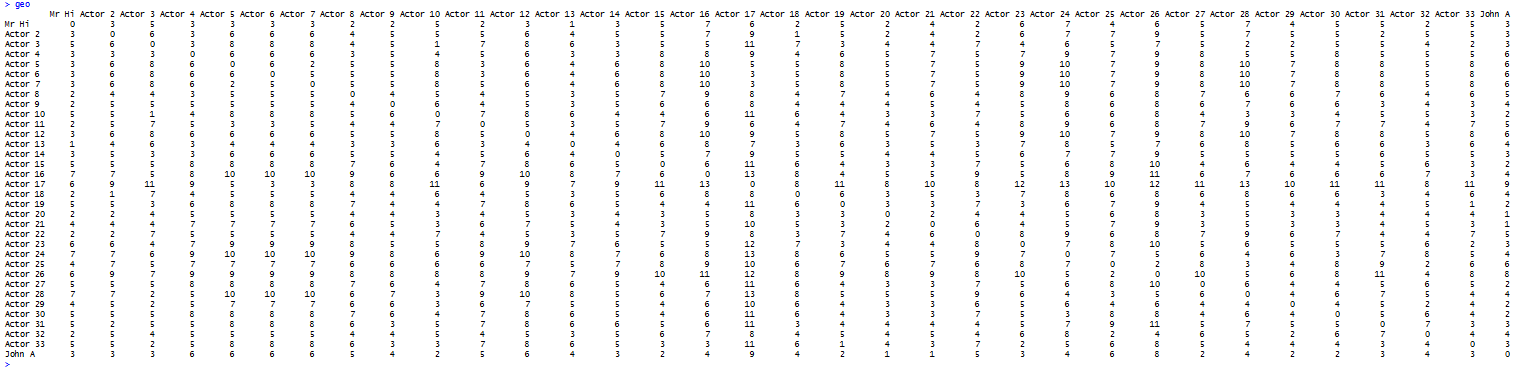
**1e. Is it meaningful to calculate co-citation for this network? Why or why not?**

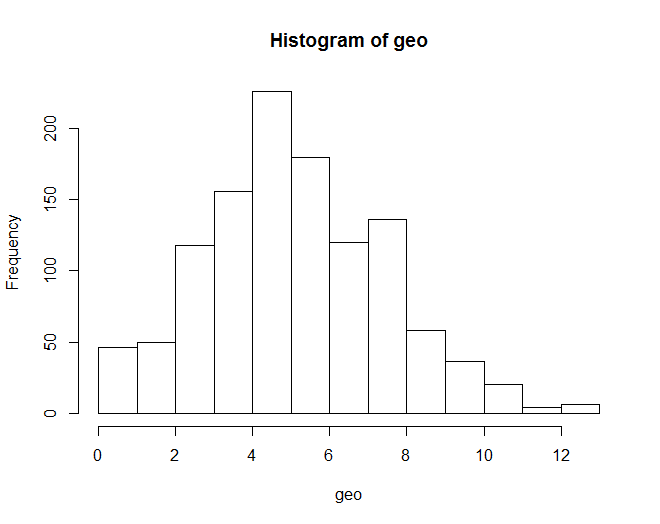
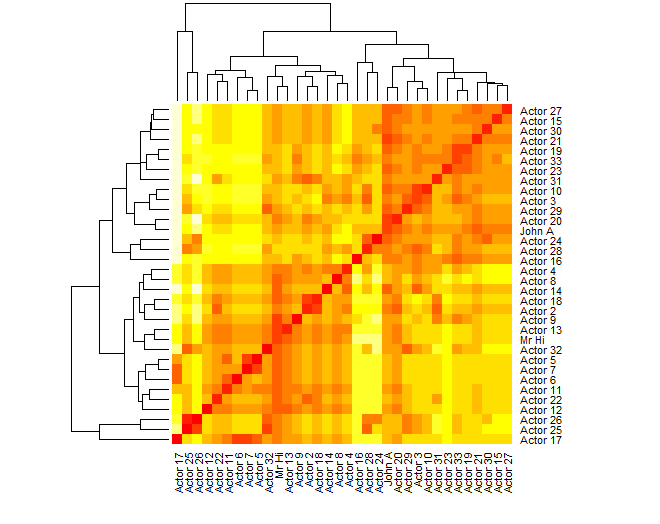
Yes, there are total number of 34 correspondents that we are examining to see the relationship of each. The get.adjacency method creates an undirected matrix (symmetrical) of relationships between 1 and other nodes. This is just the standard representation of the karate adjacencies which doesn't prove much. The cocitation calculates cocitation counts if there is another vertex citing the two vertices, Mr. Hi and John A. Both adjacency and cocitation methods can be visually represented by using the heatmap which is very useful seeing the relationships.

Adjacency heatmapCo-citation heatmap

**1f. Calculate the geodesic distance between all pairs of nodes. Construct a heatmap showing geodesic distances between node pairs. Also construct a histogram of geodesic distances. What is the average geodesic distance in the network?**







**1g. Calculate**

**i) degree centrality,**

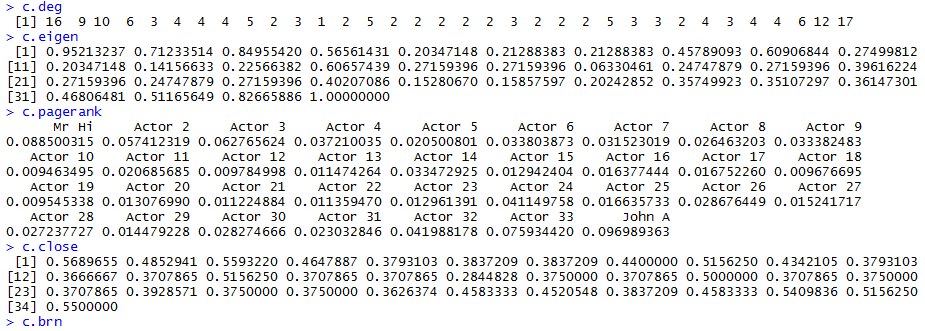
**ii) eigenvector centrality,**

**iii) pagerank,**

**iv) closeness centrality, and**

**v) betweenness centrality for all nodes in the network.**

**Construct four plots, with degree centrality in the x-axis, and the other four centrality measures in the y-axis.**



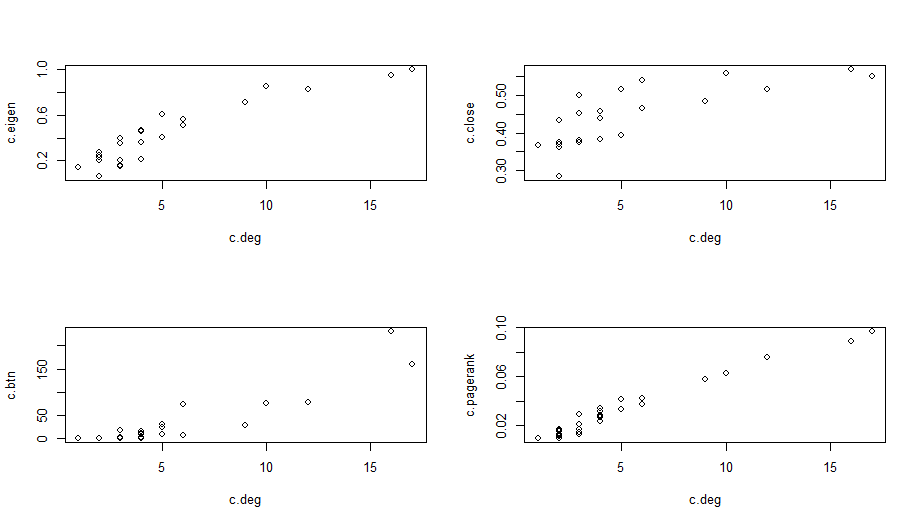


**What is your interpretation of these plots?**

Definition: Centralization is a method for creating a graph level centralization measure from the centrality scores of the vertices.

Plot for the Eigenvector centrality shows increasing trend as both c.eigen and c.deg increases. Each vertex is proportional to the sum of the scores of its neighbors. So the central vertices are those with larger central neighbors with degrees depended. Closeness centrality is more applicable in this case since at low degree, there are higher number of closeness centralities with increasing plot. This implies that the many vertices at low degree have short average distance linking to the other vertices. Betweenness centrality by definition, it measures the extent to which a vertex lies on paths between other vertices.

Vertices with higher betweenness centrality have more influence on the network, but in this case, overall betweenness centrality is very low. Pagerank centrality is very much like the Eigenvector centrality plot where it depends heavily on the degree of vertex. In this plot however, centrality points derived from neighbors and their out degree aren't so high, so the graph shows more fixed trend overall.



**1h. For each centrality measure, report the five most important vertices according to that measure. Using the vertex attribute name, identify these important vertices by name. Overall, which two vertices do you think are most important in the network?**

According to the output values from the centrality measure, Actor 3 and Actor 33 are most important in the network.

