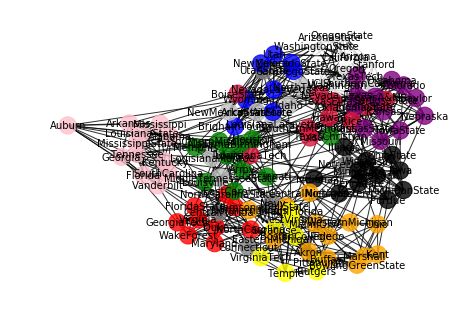
The Analysis of the NCAA Football Conference Structure

1. **Identify and load a network dataset that has some categorical information available for each node.**

For our project, we choose the network of games of the American College Football during the Fall 2000. The nodes are teams and the edges are the games between the teams. Each node was assigned the ID which is from 0 to 114, and a conference ID which is from 0 to 11. The Conference ID is based on which conference the colleges belonged to.

As we are doing ‘nx.info()’, we found that there were 115 teams, 613 games in total and the average number of game is 10.7. This network is also not a directed network, which means from the network, we may not be able to figure who wined or lost. For such a big network, when we graph it, it was a mess. So, we decide to put different colors for different conferences.

The graph is shown below.



Still, it is not a pleasant map either. We break teams into their conferences. (We have the complete codes for all the conference but for here, we just put Atlantic Coast Conference and Big 12 as our example.

1. Atlantic Coast Conference

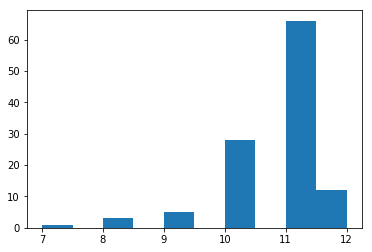
There are 9 teams in this conference: Clemson, Duke, Florida State, Georgia Tech, Maryland, North Carolina, North Carolina State, Virginia, and Wake Forest. Each team played 8 games.

1. Big 12 Conference

There are 12 teams in this conference: Baylor, Colorado, Iowa State, Kansas, Kansas State, Missouri, Nebraska, Oklahoma, Oklahoma State, Texas, Texas A&M, Texas Tech.

Here is some interesting fact, we may expect they will have 11 games for each team which is same as Atlantic Coast Conference. However, it turns out each team only played 8 games. When we look at the rules, for big conference group such as Big 12, they can’t compete with all of the other games in their conference.

Also, The minimum number of games that teams play is 7 and the maximum number of games that teams play is 12. there are 12 teams who are playing 12 games.



1. **For each of the nodes in the dataset, calculate degree centrality and eigenvector centrality.**

For this section, we have chosen not to analyze the degree of each node, since NCAA imposes rules restricting the number of games teams can play, either 11 or 12.

The code in the file will show the degree centrality and eigenvector centrality. We will leave the analysis for the next part.

1. **Compare your centrality measures across your categorical groups.**

* We set 1 for each edge’s weight so that we can calculate the degree centrality for each node and sort them from highest to lowest. From here, we can see that the first 10 teams that play 12 games also have the highest rank of **degree centrality**. (Of course, their degree centralities are very similar.) It also proves that the node has more neighbors to connect, its degree will be higher. In this case, I think teams who play 12 games not only play within their own conference, they may also play with other teams outside their conferences. That’s the reason why they can have higher degree centrality because they not only connect with teams in their own conferences, but they also can connect with teams outside their conferences.
* For the **closeness**, the more central the node is, the closer it is to all other nodes. For our case, the higher closeness may show that the game may have more games with other teams in other conference group. For example, Louisiana Tech has the highest the closeness centrality. In our networks of conference, we find out Louisiana Tech is in the conference of Western Athletic, and it only had 2 games with Hawaii and Tulsa in the Western Athletic. We know that each team had 7 games to 12 games so Louisiana Tech will have other 5 to 10 games with teams from other conference. That’s the reason why Louisiana Tech has the highest closeness centrality.
* For the **betweenness**, nodes with high betweenness may have considerable influence within a network. In the football game, it means that the team play most of games with it neighbor within the same conference. So the team such as Louisiana Tech may not have high betweenness since it only played 2 games within its conference.
* Unlike the degree centrality, the **eigenvector centrality** is that a node is important if it is linked to by other important nodes. It means that a node with high degree centrality does not meant that it also has high eigenvector centrality. For example, BrighamYoung has the highest degree centrality, but it does not have the highest eigenvector centrality accordingly. In this case, it means that Brigham Young may have games with other teams that do not have high eigenvector centrality as well. Conversely, Nevada had the highest eigenvector centrality because it connected with other high eigenvector centrality teams and it also played the most number of game as well.