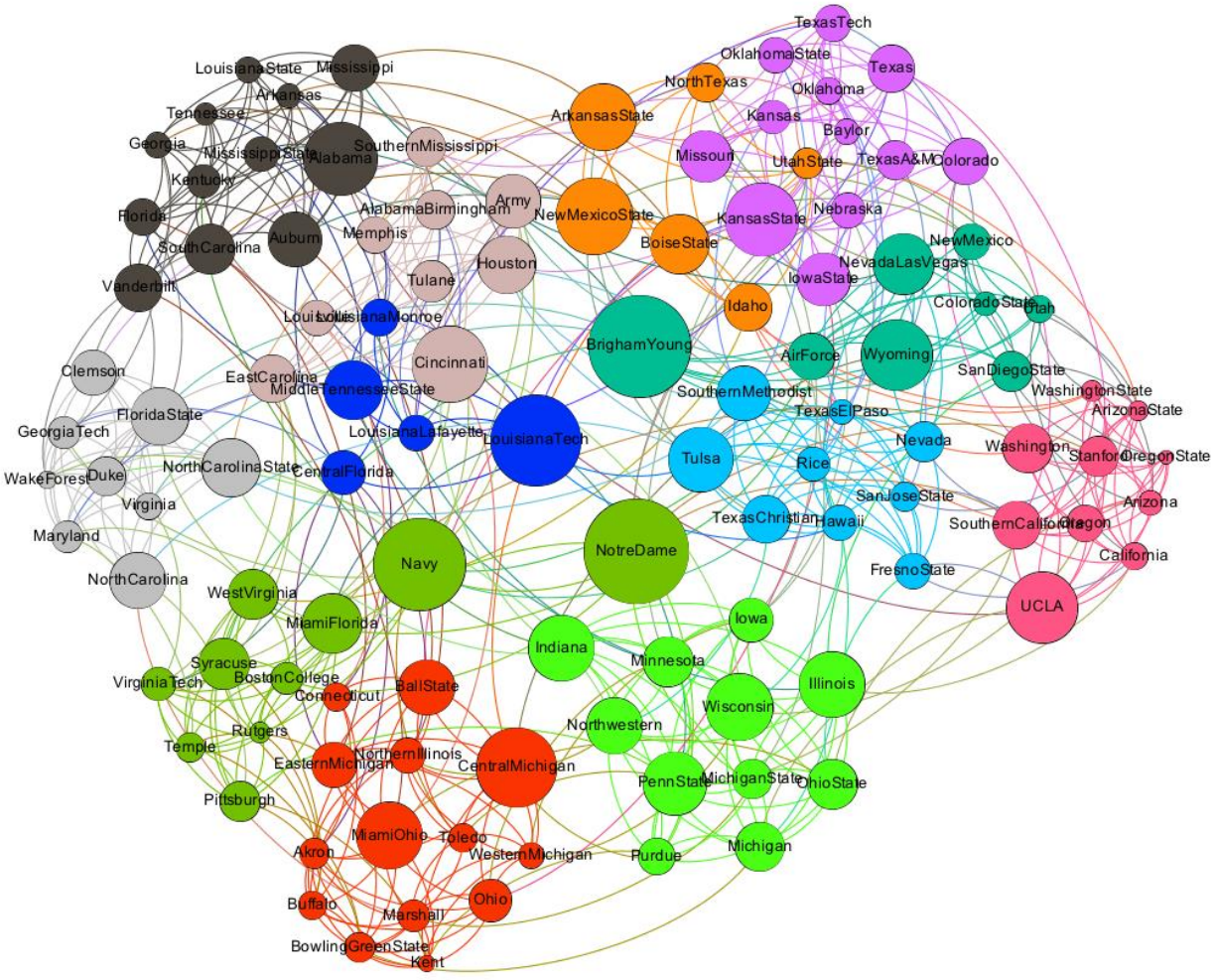
# Praxis: Detecting and visualizing communities

## Min Chen

- The first part of the assignment is done by using the downloaded the American college football
  1. I chose a resolution of 0.8. I started with the default value 1, and try to see how many communities are generated. According to the nature of the graph, the football games contain 12 conferences. It is natural to project that the network should have 12 communities, each of which stands for a conference. When I decrease the resolution to 0.8, it has 12 communities. Even if I further decrease it to 0.3, there are still 12 communities generated. I just chose the largest resolution that I used to get the “correct” community number.



2. This is the visualization of network that I prefer. Here, 12 communities are indicated by different colors of the nodes and sizes of the nodes are proportional to the betweenness centrality calculated. The relationship of betweenness centrality and communities I found is as follows: for each community, the node with biggest betweenness centrality is the one that "connects" the community to other communities, i.e. the node which has most links to other communities. This makes sense according to the definition of betweenness centrality, the node with higher value means that more shortest path is passing through this node, which serves as a bridge between communities. Accordingly, the node "inside" the community, i.e. has a lot links within community but not to other communities, generally has low betweenness centrality.

- The second graph I am looking at is the weighted network of coauthorships between scientists posting preprints on the High-Energy Theory E-Print Archive between Jan 1, 1995 and December 31, 1999. M. E. J. Newman, Proc. Natl. Acad. Sci. USA 98, 404-409 (2001).
    1. Choosing the resolution is kind of tricky for this graph, which has 8361 nodes and 15751 edges. Resolution value 5, 10, 20 gives number of communities 1342, 1335, and 1332 respectively. There is no big difference and the number of communities is around 1335. Therefore, I chose the value 10.

       However there are only several huge communities in the graph:
        - The pink one, containing 38.68% of the nodes
        - The blue one, containing 14.81% of the nodes
        - The green one, containing 12.73% of the nodes.

       These communities represents researchers who have interest in similar fields and each popular field may create a community. Therefore, the majority of the researchers focus on maybe 3 major fields of the high-energy theory.

    2. In this graph, I also visualize another centrality measure, the eigenvector centrality, which is proportional to the size of the nodes. I really do not find a general relationship between this centrality measure and formation of communities. I guess eigenvector centrality considers the degree of the neighbors but does not differentiate whether the neighbor is within the same community or not.

       However, there is one interesting finding is that on the top part of the graph, there is a community of color blue-green, and all nodes in this communities has a large value of eigenvector centrality, even bigger than most of the nodes that in the middle of the network within the largest pink community. My intuition is that this community, although not super big, but is highly clustered. Each node in this community has a lot of connections among each other. This looks like a small group of scientist that working on one field and publish a lot paper within that field and do not co-author with other scientist a lot.