Detecting Communities Using Gephi

Part 1) American College football: network of American football

games between Division IA colleges during regular season Fall 2000.
M. Girvan and M. E. J. Newman, Proc. Natl. Acad. Sci. USA 99, 7821-7826 (2002).

- Adjust the resolution of the community detection algorithm and get a sense of what communities are formed at different resolutions. Pick a resolution that makes sense to you. Give a physical interpretation of the communities and justify your choice for the resolution parameter.

The communities in the American College football network should be equal to the NCAA Division 1A conference in Fall 2000. Each college team is part of one conference, and the football teams play every other team in their conference once usually and some games against teams not in their conference.

According to NCAA Division 1A football season of 2000 (https://en.wikipedia.org/wiki/2000_NCAA_Division_I-A_football_season), there were 12 conferences held in college football. So I have a goal to detect about 12 communities.

	Conference	# of Teams	Teams
1	ACC	9	Florida State, Clemson, Georgia Tech, Virginia, North Carolina State, North Carolina, Maryland, Wake Forest, Duke
2	Big Ten	11	Michigan, Northwestern, Purdue, Ohio State, Wisconsin, Minnesota, Penn State, Iowa, Illinois, Michigan State, Indiana
3	Big West	6	Boise State, Utah State, Idaho, New Mexico State, North Texas, Arkansas State
4	Big East	8	Miami (FL), Virginia Tech, Pitt, Syracuse, Boston College, West Virginia, Temple, Rutgers
5	Big 12	12	Nebraska, Kansas State, Iowa State, Colorado, Kansas, Missouri, Oklahoma, Texas, Texas A&M, Texas Tech, Oklahoma State, Baylor
6	Conference USA	9	Louisville, East Carolina, Cincinnati, Southern Mississippi, UAB, Tulane, Memphis, Houston, Army
7	Mountain West	8	Colorado State, Air Force, Nevada-Las Vegas, Brigham Young, New Mexico, Utah, San Diego State, Wyoming
8	Mid-American	13	Marshall, Ohio, Akron, Miami (OH), Bowling Green State, Buffalo, Kent State, Toledo, Western Michigan, Northern Illinois, Ball State, Eastern Michigan, Central Michigan
9	Pacific-10	10	Oregon State, Washington, Oregon, Stanford, Arizona State, UCLA, Arizona, USC, Washington State, California
10	SEC	12	Florida, Georgia, South Carolina, Tennessee, Vanderbilt, Kentucky, Auburn, LSU, Mississippi State, Ole Miss, Arkansas, Alabama
11	Independents	8	Notre Dame, South Florida, UCF, Middle Tennessee State, Connecticut, Louisiana Tech, Louisiana-Lafayette, Louisiana-Monroe, Navy
12	WAC	9	Texas Christian, UTEP, Fresno State, San Jose State, Tulsa, Hawaii, Rice, SMU, Nevada

Modularity

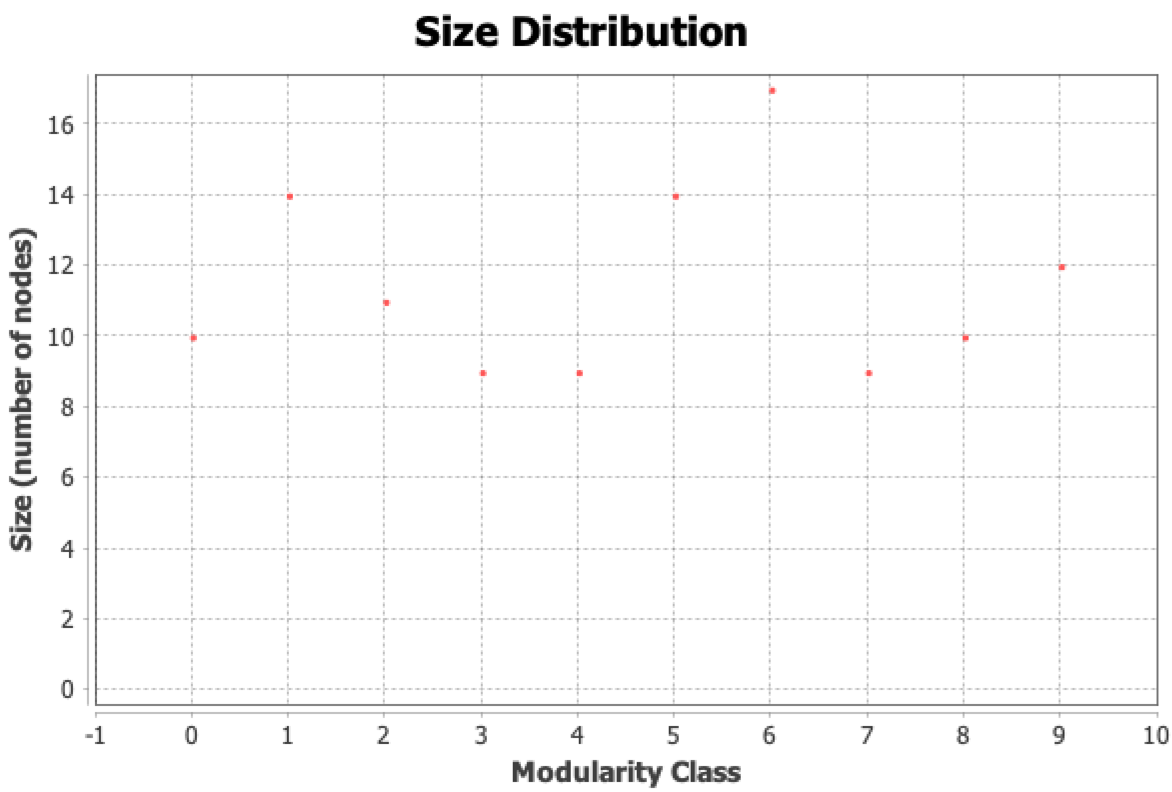
Modularity Report

Parameters:

Randomize: On
Use edge weights: Off
Resolution: 1.0

Results:

Modularity: 0.604
Modularity with resolution: 0.604
Number of Communities: 10



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, *Fast unfolding of communities in large networks*, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona *Laplacian Dynamics and Multiscale Modular Structure in Networks 2009*

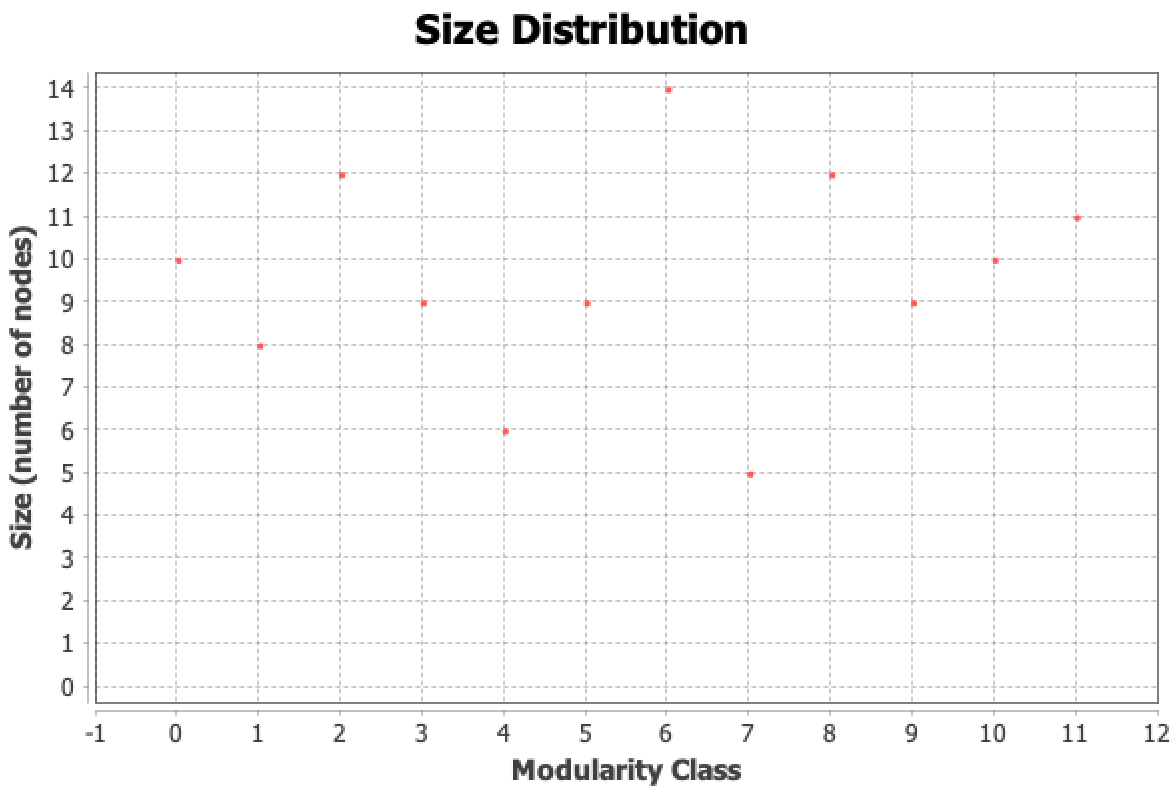
Modularity Report

Parameters:

Randomize: On
Use edge weights: Off
Resolution: 0.5

Results:

Modularity: 0.601
Modularity with resolution: 0.255
Number of Communities: 12



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks 2009
*The first run of **Modularity** with Resolution = 1.0 detected 10 communities. This was a good initial partitioning given the goal is to find 12 communities. Next I reduced Resolution to detect more communities.*

The second run with Resolution = 0.5 detected 12 communities. I reached the goal.

- *Calculate the betweenness centrality for the nodes in the network and create a visualization which shows both the betweenness centrality and community membership (save this figure). How are betweenness centrality and community structure related?*

Layout:

- *Force Atlas 2*
- *Tolerance: 1.0*
- *Approximation: 1.2*
- *Scaling: 7.5*
- *Gravity: 0.01*
- *Prevent Overlap: ON*

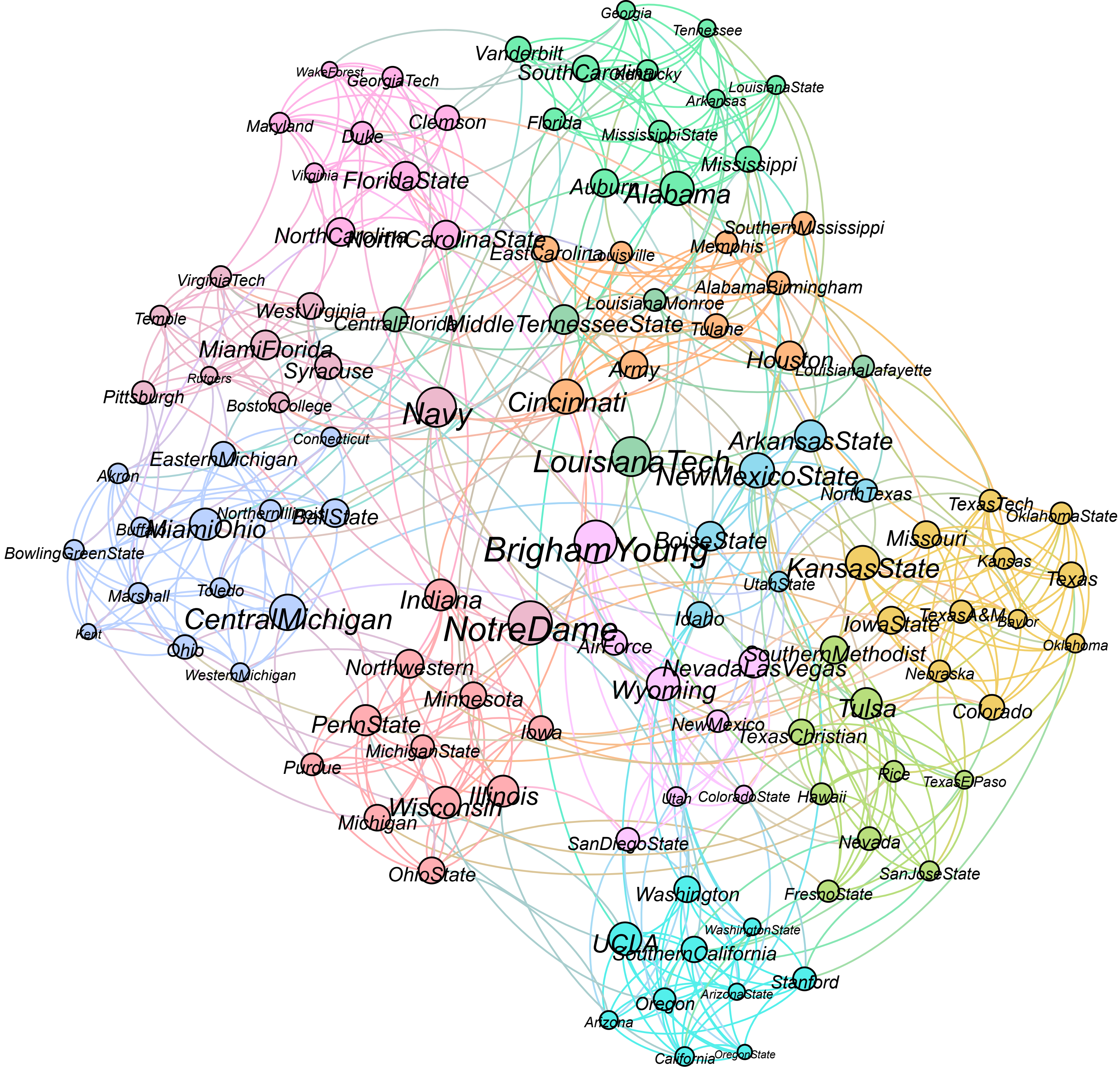
Appearance:

- ***Node Color***
- *Partition: Modularity Class*
- *Custom Palette*
- *Number of Colors: 12*
- *Presets: Fancy (light background)*
- ***Node Size***
- *Ranking: Betweenness Centrality*
- *min size: 4.0*
- *max size: 12.0*

In [1]:

```
from IPython.core.display import SVG
SVG(filename='football.svg')
```

Out[1]:



The nodes with a high Betweenness Centrality measure seem to have the greatest number of connections to nodes in other communities. In this graph above, it is showing the football teams which played the greatest number of out of conference games compared to the other teams in the same conference.

Part 2) Zachary's Karate Club *: social network of friendships between 34 members of a karate club at a US university in the 1970s. Please cite W. W. Zachary, An information flow model for conflict and fission in small groups, Journal of Anthropological Research 33, 452-473 (1977)*

- *Find an appropriate resolution for detecting meaningful communities in the network and adjust the node layout accordingly. Does the network have a distinct community structure? If so, what do these communities represent and how do you think they were formed? If not, why do think that structure is lacking?*

I found more information about getting a better idea for the appropriate resolution for detecting the right number of communities in this network from the ***Network Science*** *book by Barabasi as this network of the Karate Club was also written about there. This network has 34 nodes and 78 connections. Communities presented by Barabasi are:*

Community	Color	Number of Nodes	Membership
1	White	12	9, 10, 15, 16, 19, 21, 23, 27, 30, 31, 33, 34
2	Purple	6	24, 25, 26, 28, 29, 32
3	Green	11	1, 2, 3, 4, 8, 12, 13, 14, 18, 20, 22
4	Orange	5	5, 6, 7, 11, 17

Modularity

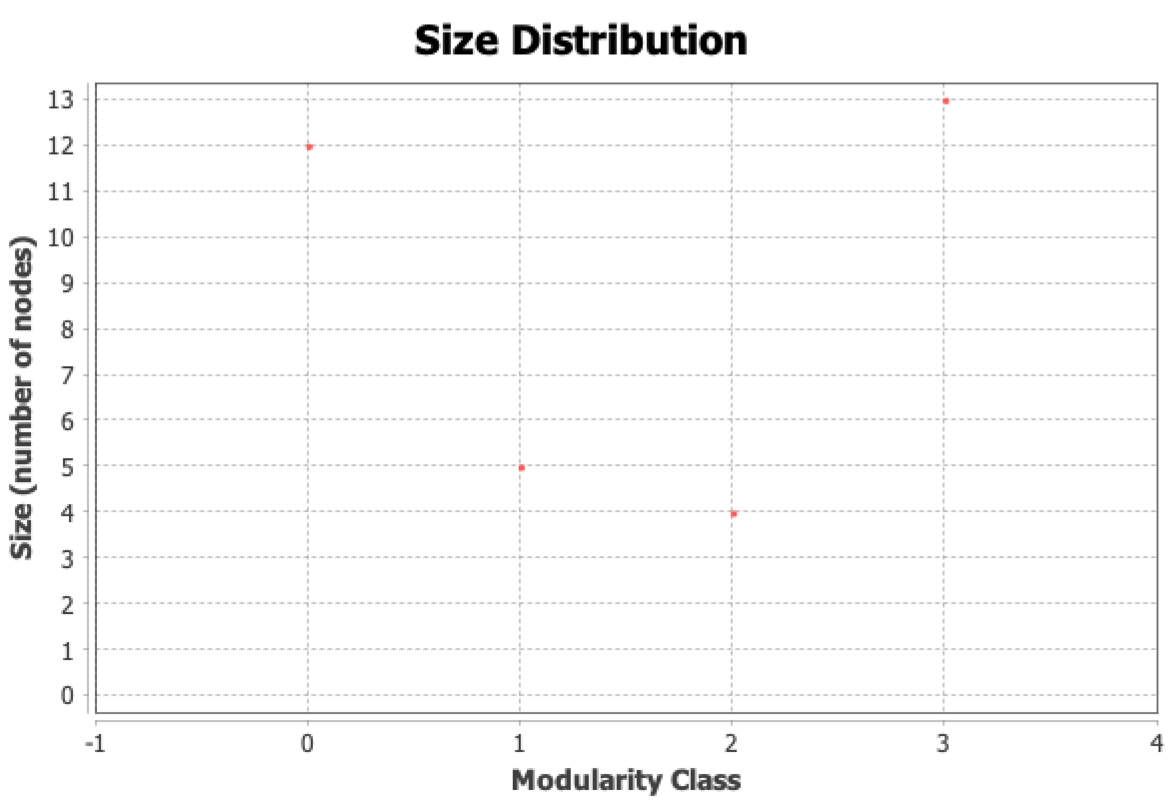
Modularity Report

Parameters:

Randomize: On
Use edge weights: Off
Resolution: 1.0

Results:

Modularity: 0.416
Modularity with resolution: 0.416
Number of Communities: 4



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks 2009

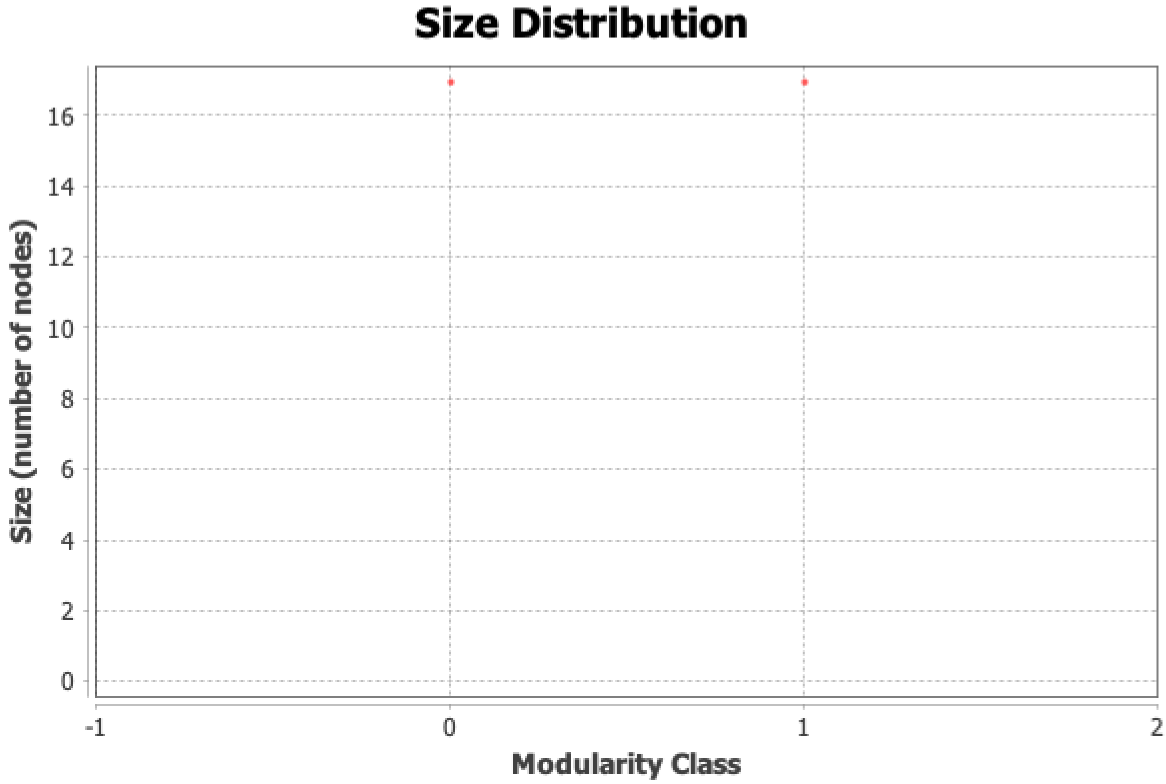
Modularity Report

Parameters:

Randomize: On
Use edge weights: Off
Resolution: 2.0

Results:

Modularity: 0.372
Modularity with resolution: 1.244
Number of Communities: 2



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks 2009
Selected Result

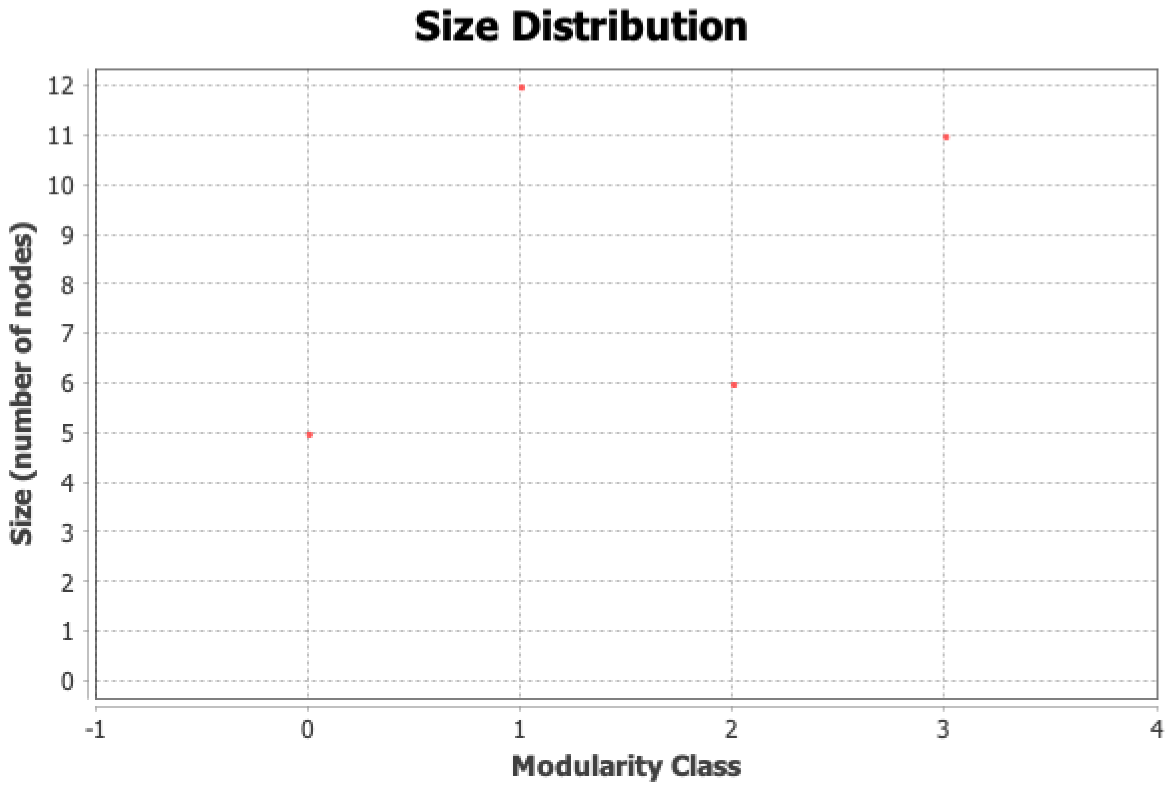
Modularity Report

Parameters:

Randomize: On
Use edge weights: Off
Resolution: 0.9

Results:

Modularity: 0.419
Modularity with resolution: 0.346
Number of Communities: 4



Algorithm:

Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, Etienne Lefebvre, Fast unfolding of communities in large networks, in Journal of Statistical Mechanics: Theory and Experiment 2008 (10), P1000

Resolution:

R. Lambiotte, J.-C. Delvenne, M. Barahona Laplacian Dynamics and Multiscale Modular Structure in Networks 2009
*The first attempt on **Modularity** with Resolution = 1.0 detected the goal of 4 communities but their sizes were not what I wanted according to the Barabasi reference. With Resolution = 2.0 there were 2 communities. Decreasing Resolution to 0.9 detected four communities again and also with the targeted sizes.*

Nodes were assigned to the following communities:

1. ***Orange** (5): 5, 6, 7, 11, 17*

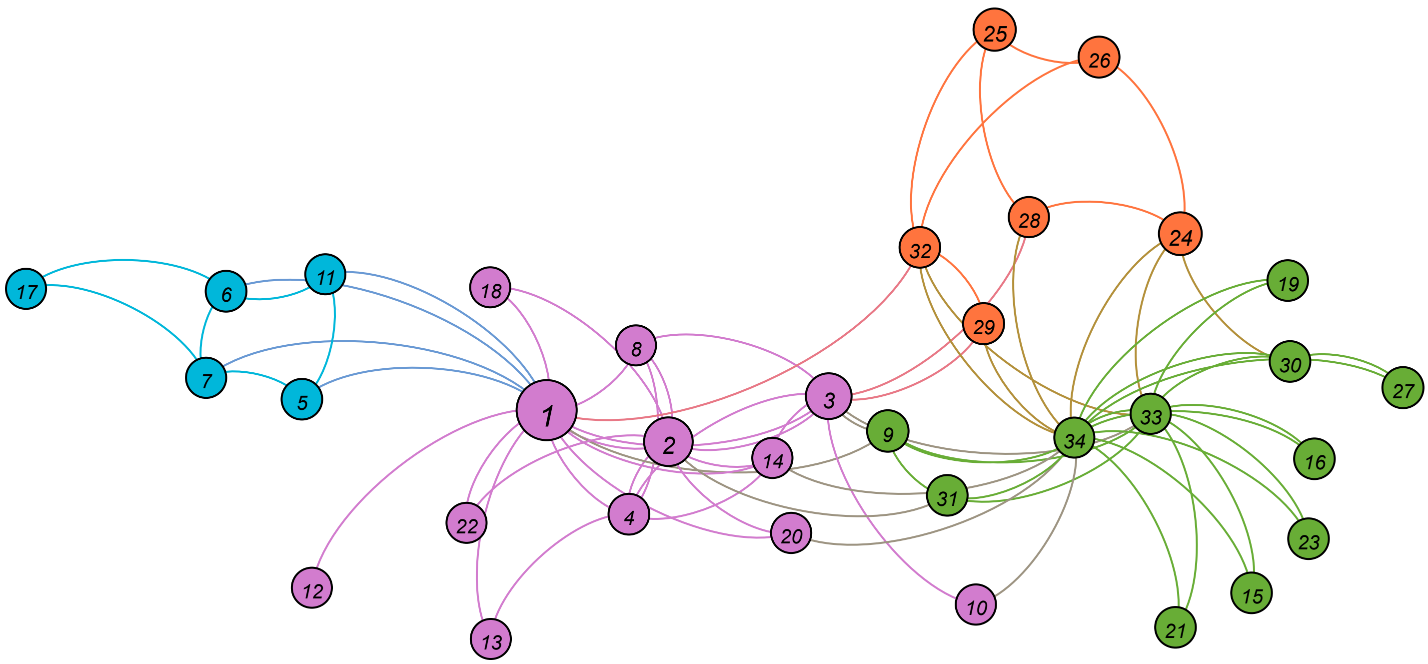
2. **Green** (12): 1, 2, 3, 4, 8, 10, 12, 13, 14, 18, 20, 22
3. **Purple** (6): 24, 25, 26, 28, 29, 32
4. **White** (11): 9, 15, 16, 19, 21, 23, 27, 30, 31, 33, 34

- *Pick another centrality measure and create (and save) a visualization that captures the community membership along with this centrality measure. Explain what relationships (or lack thereof) you see between the centrality measure you chose and community structure. Can you relate your explanations back to the real-world elements that make-up the network?*

In [2]:

```python
from IPython.core.display import SVG
SVG(filename='karate.svg')
```

Out[2]:



Appearance:

- ***Node Color***
- *Partition: Modularity Class*
- *4 colors (default)*
- ***Node Size***
- *Ranking: Eigenvector Centrality*
- *min size: 10.0*
- *max size: 15.0*

Layout:

- *Yifan Hu*
- *default settings*
- *Noverlap*
- *default settings*

I used Eigenvenctor Centrality because it signifies the overall center of the communities most effectively. Hubs can be clearly seen and how they impact the whole community specifically throug the nodes 33, 34, 1, and 2.