

1. Zachary's Karate Club

Calculate relevant centralities for the nodes in the Karate Club network. Partition the network into clusters*. Visualize the network so that the important nodes are highlighted#. Do you find two partitions centered on the administrator (node 34) and the instructor (1)? Node 9 was the only one that Zachary's analysis failed to properly place. Does your analysis say that node 9 should stay with the administrator or the instructor? The original article notes that number 9 was close to receiving his/her black belt and suggests that this might have influenced his/her decision.

* See *kernighan_lin_bisection* or *girvan_newman*

For examples, see https://networkx.github.io/documentation/latest/auto_examples/index.html

You might try using *draw_kamada_kawai* or *spring_layout*

.....

Figure 1 presents the Karate Club network. The size of the nodes corresponds to the degree of the node. Hence, the bigger the node the more important and influential the node is. As shown, nodes 34 (degree=17) and node 1 (degree=16) are shown to be more important than the rest. Node 33 (degree=12) comes in third place.

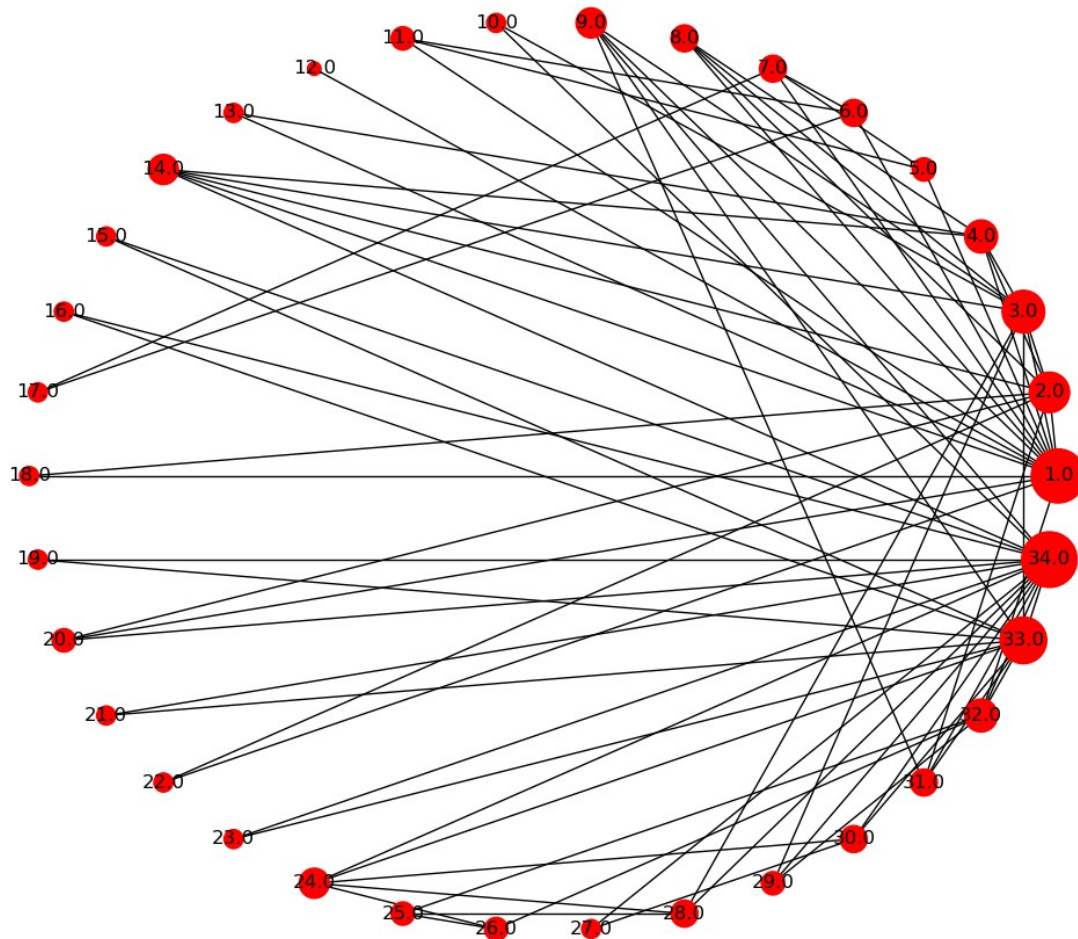


Figure 1. Graphical Representation of the Karate Club network.

Type: Graph

Number of edges: 78

Number of nodes: 34

Average degree: 4.5882

Network diameter = 5

Network density = 0.13903743315508021

Network average clustering coefficient = 0.5706384782076823

Network average shortest path length= 2.408199643493761

The following presents the relevant centralities for the nodes in the Karate Club network. Nodes are sorted by each centrality metric in descending order. The administrator (node 34) of the Karate Club is the highest in degree, eigenvector and page rank centrality metrics, second in betweenness and third in closeness centrality metrics. The instructor (node 1) ranks highest in betweenness and closeness centrality metrics, and comes second in degree, eigenvector and page rank centrality metrics. The administrator and the instructor are influential and important in the club in different ways, depending on the centrality metric being measured. See lists below:

Nodes sorted by degree centrality:

```
('34.0', 0.5151515151515151)
('1.0', 0.48484848484848486)
('33.0', 0.36363636363636365)
('3.0', 0.30303030303030304)
('2.0', 0.2727272727272727)
('4.0', 0.18181818181818182)
('32.0', 0.18181818181818182)
('9.0', 0.15151515151515152)
('14.0', 0.15151515151515152)
('24.0', 0.15151515151515152)
('6.0', 0.12121212121212122)
('7.0', 0.12121212121212122)
('8.0', 0.12121212121212122)
('28.0', 0.12121212121212122)
('30.0', 0.12121212121212122)
('31.0', 0.12121212121212122)
('5.0', 0.09090909090909091)
('11.0', 0.09090909090909091)
('20.0', 0.09090909090909091)
('25.0', 0.09090909090909091)
('26.0', 0.09090909090909091)
('29.0', 0.09090909090909091)
('10.0', 0.06060606060606061)
('13.0', 0.06060606060606061)
('15.0', 0.06060606060606061)
('16.0', 0.06060606060606061)
('17.0', 0.06060606060606061)
('18.0', 0.06060606060606061)
('19.0', 0.06060606060606061)
('21.0', 0.06060606060606061)
('22.0', 0.06060606060606061)
('23.0', 0.06060606060606061)
('27.0', 0.06060606060606061)
('12.0', 0.030303030303030304)
```

Nodes sorted by betweenness centrality:

```
('1.0', 0.43763528138528146)
('34.0', 0.30407497594997596)
('33.0', 0.145247113997114)
('3.0', 0.14365680615680618)
('32.0', 0.13827561327561325)
('9.0', 0.05592682780182781)
('2.0', 0.053936688311688304)
('14.0', 0.04586339586339586)
('20.0', 0.03247504810004811)
('6.0', 0.02998737373737374)
('7.0', 0.02998737373737374)
('28.0', 0.02233345358345358)
('24.0', 0.017613636363636363)
('31.0', 0.014411976911976909)
('4.0', 0.011909271284271283)
```

```
( '26.0', 0.0038404882154882154)
( '30.0', 0.0029220779220779218)
( '25.0', 0.0022095959595959595)
( '29.0', 0.0017947330447330447)
( '10.0', 0.0008477633477633478)
( '5.0', 0.0006313131313131313)
( '11.0', 0.0006313131313131313)
( '8.0', 0.0)
( '12.0', 0.0)
( '13.0', 0.0)
( '15.0', 0.0)
( '16.0', 0.0)
( '17.0', 0.0)
( '18.0', 0.0)
( '19.0', 0.0)
( '21.0', 0.0)
( '22.0', 0.0)
( '23.0', 0.0)
( '27.0', 0.0)
```

Nodes sorted by closeness centrality:

```
( '1.0', 0.5689655172413793)
( '3.0', 0.559322033898305)
( '34.0', 0.55)
( '32.0', 0.5409836065573771)
( '9.0', 0.515625)
( '14.0', 0.515625)
( '33.0', 0.515625)
( '20.0', 0.5)
( '2.0', 0.4852941176470588)
( '4.0', 0.4647887323943662)
( '28.0', 0.4583333333333333)
( '31.0', 0.4583333333333333)
( '29.0', 0.4520547945205479)
( '8.0', 0.44)
( '10.0', 0.4342105263157895)
( '24.0', 0.39285714285714285)
( '6.0', 0.38372093023255816)
( '7.0', 0.38372093023255816)
( '30.0', 0.38372093023255816)
( '5.0', 0.3793103448275862)
( '11.0', 0.3793103448275862)
( '18.0', 0.375)
( '22.0', 0.375)
( '25.0', 0.375)
( '26.0', 0.375)
( '13.0', 0.3707865168539326)
( '15.0', 0.3707865168539326)
( '16.0', 0.3707865168539326)
( '19.0', 0.3707865168539326)
( '21.0', 0.3707865168539326)
( '23.0', 0.3707865168539326)
( '12.0', 0.36666666666666664)
( '27.0', 0.3626373626373626)
( '17.0', 0.28448275862068967)
```

Nodes sorted by eigenvector centrality:

```
( '34.0', 0.373371213013235)
( '1.0', 0.3554834941851943)
( '3.0', 0.31718938996844476)
( '33.0', 0.3086510477336959)
( '2.0', 0.2659538704545025)
( '9.0', 0.2274050914716605)
( '14.0', 0.22646969838808148)
```

```
('4.0', 0.2111740783205706)
('32.0', 0.19103626979791702)
('31.0', 0.17476027834493085)
('8.0', 0.17095511498035434)
('24.0', 0.15012328691726787)
('20.0', 0.14791134007618667)
('30.0', 0.13496528673866567)
('28.0', 0.13347932684333308)
('29.0', 0.13107925627221215)
('10.0', 0.10267519030637758)
('15.0', 0.10140627846270832)
('16.0', 0.10140627846270832)
('19.0', 0.10140627846270832)
('21.0', 0.10140627846270832)
('23.0', 0.10140627846270832)
('18.0', 0.09239675666845953)
('22.0', 0.09239675666845953)
('13.0', 0.08425192086558088)
('6.0', 0.07948057788594247)
('7.0', 0.07948057788594247)
('5.0', 0.07596645881657382)
('11.0', 0.07596645881657381)
('27.0', 0.07558192219009324)
('26.0', 0.05920820250279008)
('25.0', 0.05705373563802805)
('12.0', 0.05285416945233648)
('17.0', 0.023634794260596875)
```

Nodes sorted by page rank:

```
('34.0', 0.1009179167487121)
('1.0', 0.09700181758983709)
('33.0', 0.07169213006588289)
('3.0', 0.057078423047636745)
('2.0', 0.05287839103742701)
('32.0', 0.03715663592267942)
('4.0', 0.03586064322306479)
('24.0', 0.03152091531163228)
('9.0', 0.029765339186167028)
('14.0', 0.029536314977202986)
('6.0', 0.02911334166344221)
('7.0', 0.02911334166344221)
('30.0', 0.02628726283711208)
('28.0', 0.025638803528350497)
('31.0', 0.02458933653429248)
('8.0', 0.024490758039509182)
('5.0', 0.021979406974834498)
('11.0', 0.021979406974834498)
('25.0', 0.021075455001162945)
('26.0', 0.021005628174745786)
('20.0', 0.019604416711937293)
('29.0', 0.01957296050943854)
('17.0', 0.016785378110253487)
('27.0', 0.015043395360629753)
('13.0', 0.014645186487916191)
('18.0', 0.014558859774243493)
('22.0', 0.014558859774243493)
('15.0', 0.014535161524273825)
('16.0', 0.014535161524273825)
('19.0', 0.014535161524273825)
('21.0', 0.014535161524273825)
('23.0', 0.014535161524273825)
('10.0', 0.014308950284462801)
('12.0', 0.009564916863537148)
```

The Karate Club network was partitioned into two clusters using two methods: the Kernighan_lin_bisection¹ and the *girvan_newman* algorithms.² Both algorithms resulted in two clusters with the administrator (node 34) and the instructor (node 1) as the center of each. In both cases, node 9 stays in the cluster with administrator (node 34).

Figures 2 and 3 below presents the clusters resulted from Kernighan_lin_bisection algorithm. The algorithm bisects the nodes into two disjoint subsets of the same size using the minimum edge cut approach. Ten edges were cut as a result. Figures 4 and 5 below presents the partitions resulted from Girvan Newman algorithm. With this algorithm, the cluster with the administrator (node 34) is slighter bigger than the one with the instructor (node 1). Girvan Newman algorithm removes one edge at a time in the descending order of betweenness centrality. Ten edges from the most betweenness were removed.

The original article shows that node 9 has a stronger (higher weight) relationship with node 34 (the administrator/John A.) than with node 1 (the instructor/Mr. Hi). See Figure 3 “Quantified Matrix of Relative Strengths of the Relationships in the Karate Club: The Matrix C” on page 462. My analysis using either Kernighan_lin_bisection or the *girvan_newman* algorithms on the provided, unweighted network data shows that node 9 stays with node 34. This agrees with run result from the *NETFLOW* presented in the original article (page 465). The NETFLOW results show that node 9 sides with the Sink (the administrator/node 34/John A.). However according to the original report, node 9 joined the instructor’s (Mr. Hi’s/node 1’s) club after the Karate Club split. So, all three algorithms failed to correctly predict node 9’s decision. The original article notes that node 9 was close to receiving his/her black belt and suggests that this might have influenced his/her decision. After karate club split, it makes sense that node 9 chose to join Mr. Hi’s private karate studio so that he could continue his practice with the same instructor till he got his black belt. However, this decision factor was not part of the network data and hence would not have an effect in the result of the partition by any of the three algorithms.

¹ The input to the algorithm is an undirected graph $G = (V, E)$ with vertex set V , edge set E , and (optionally) numerical weights on the edges in E . The goal of the algorithm is to partition V into two disjoint subsets A and B of equal (or nearly equal) size, in a way that minimizes the sum T of the weights of the subset of edges that cross from A to B . If the graph is unweighted, then instead the goal is to minimize the number of crossing edges; this is equivalent to assigning weight one to each edge. (Source: Wikipedia.org; link: https://en.wikipedia.org/wiki/Kernighan%E2%80%93Lin_algorithm)

² The Girvan–Newman algorithm detects communities by progressively removing edges from the original network. The connected components of the remaining network are the communities. Instead of trying to construct a measure that tells us which edges are the most central to communities, the Girvan–Newman algorithm focuses on edges that are most likely “between” communities. (Source: Wikipedia.org; link: https://en.wikipedia.org/wiki/Girvan%E2%80%93Newman_algorithm)

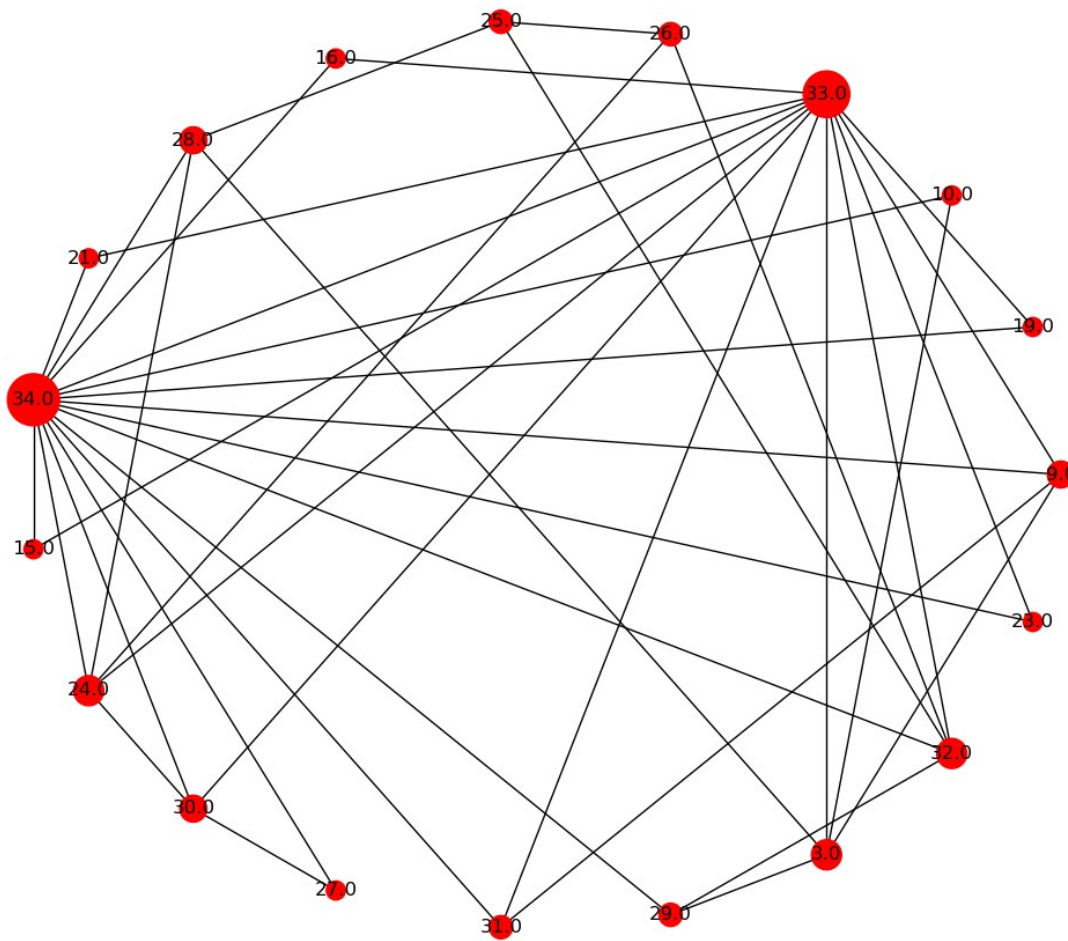


Figure 2. Bisected Cluster Containing Node 34, the Administrator and Node 9.

Using Kernighan_lin_bisection algorithm:

Set 1 = {'24.0', '15.0', '21.0', '30.0', '29.0', '**34.0**', '33.0', '32.0', '31.0', '25.0', '**9.0**', '27.0', '16.0', '26.0', '19.0', '23.0', '28.0'}

Type: SubGraph

Number of nodes: 17

Number of edges: 34

Average degree: 4.0000

Network diameter = 3

Network density = 0.25

Network average clustering coefficient = 0.6911578617460971

Network average shortest path length= 1.875

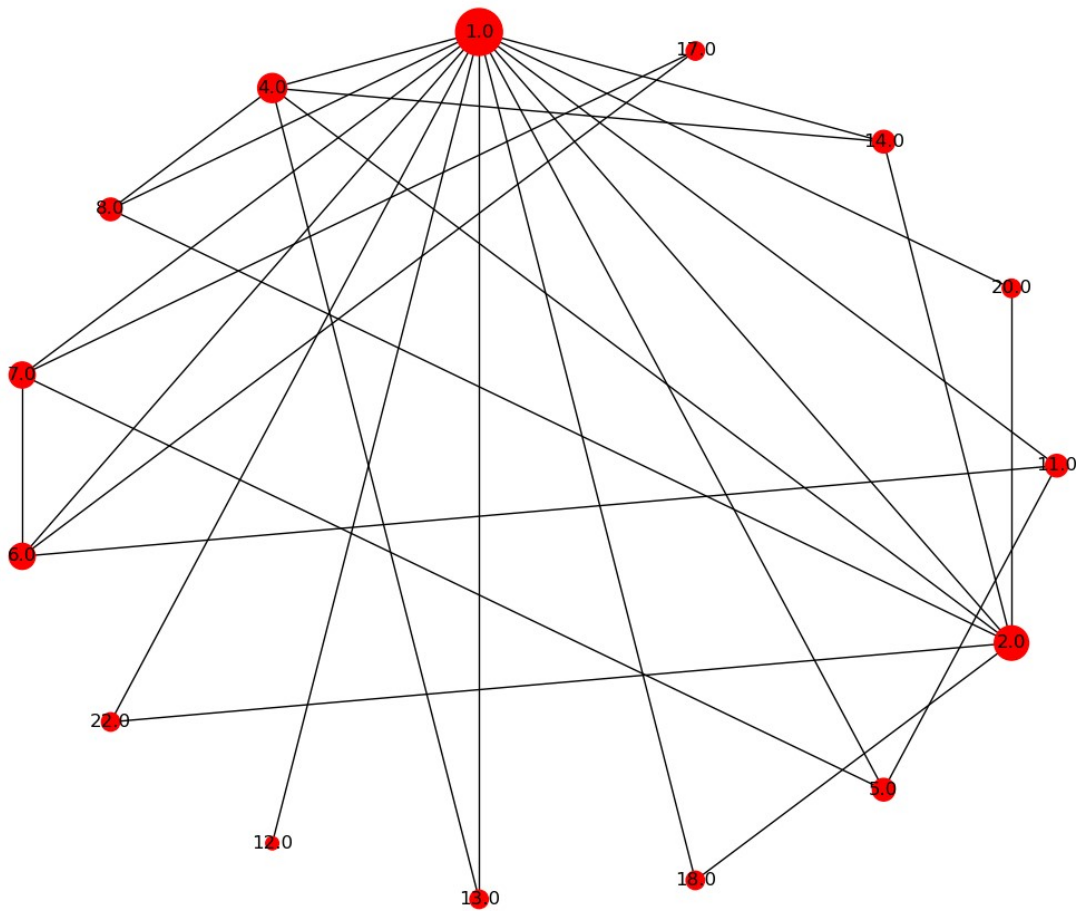


Figure 3. Bisected Cluster Containing Node 1, the Instructor.

Using Kernighan_lin_bisection algorithm:

Set 2 = {'13.0', '5.0', '7.0', '8.0', '11.0', '3.0', '17.0', '22.0', '18.0', '10.0', '2.0', '12.0', '20.0', '6.0', '4.0', '14.0', '**1.0**'}

Type: SubGraph

Number of nodes: 17

Number of edges: 34

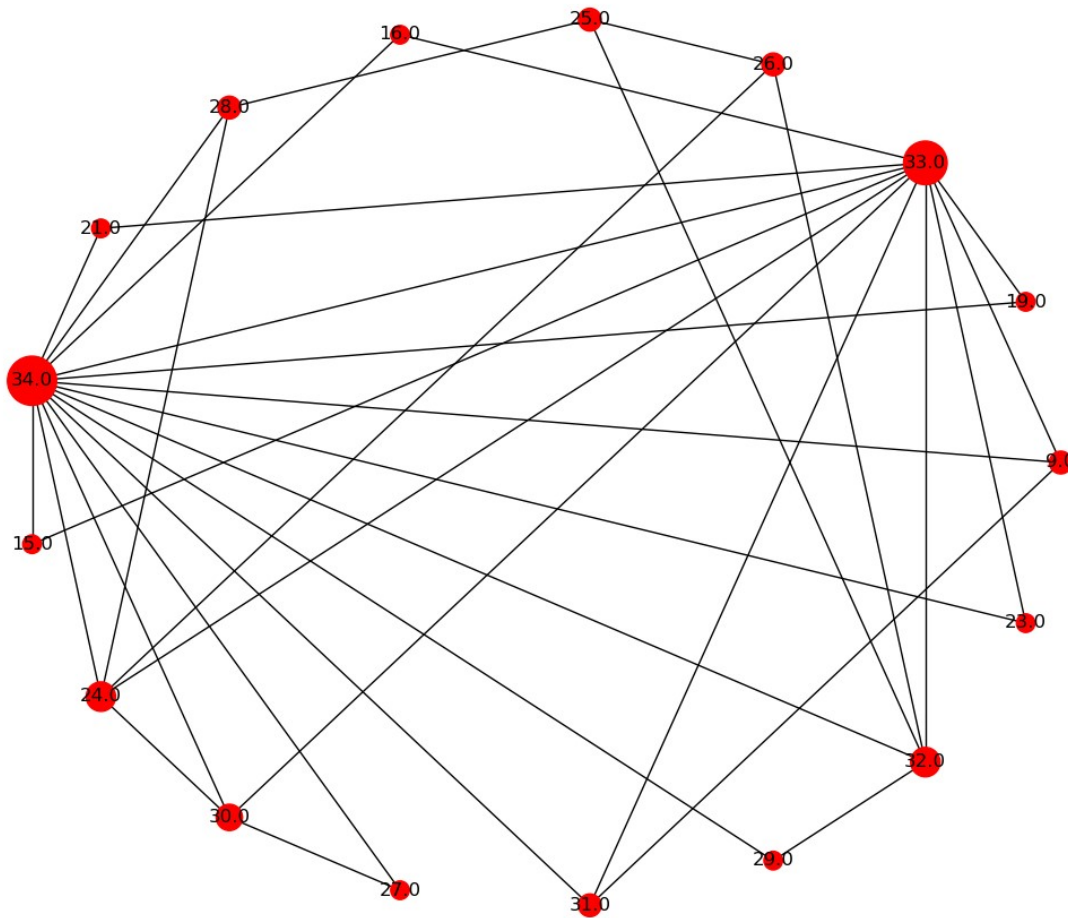
Average degree: 4.0000

Network diameter = 4

Network density = 0.25

Network average clustering coefficient = 0.6597285067873303

Network average shortest path length= 1.9044117647058822



*Figure 4. Girvan Newman Algorithm:
Cluster Containing Node 34, the Administrator and Node 9*

Using girvan_newman algorithm:

Community 2 = {'24.0', '21.0', '32.0', '15.0', '31.0', '30.0', '10.0', '25.0', '9.0', '27.0', '29.0', '34.0', '26.0', '16.0', '3.0', '33.0', '19.0', '23.0', '28.0'}

Type: SubGraph

Number of nodes: 19

Number of edges: 40

Average degree: 4.2105

Network diameter = 3

Network density = 0.23391812865497075

Network average clustering coefficient = 0.5599908863066758

Network average shortest path length= 1.8888888888888888

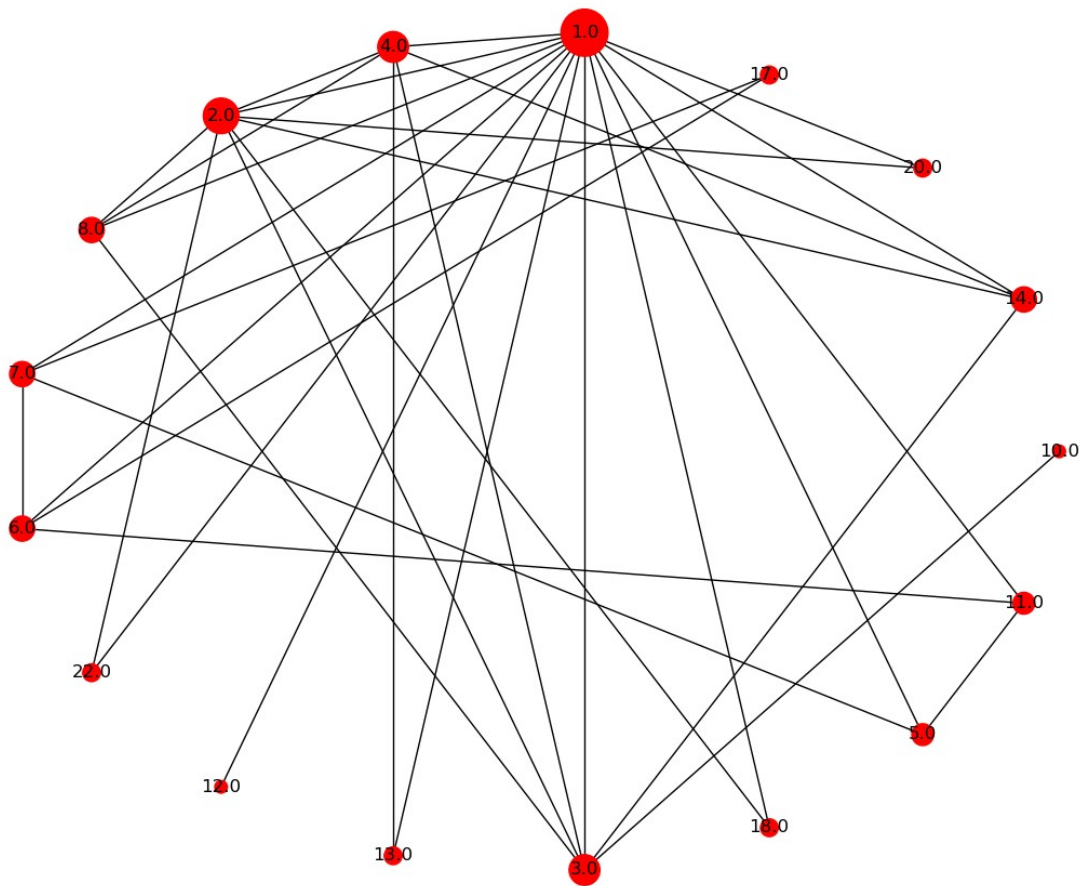


Figure 5. Girvan Newman Algorithm: Cluster Containing Node 1, the Instructor.

Using girvan_newman algorithm:

Community 1 = {'13.0', '22.0', '5.0', '18.0', '7.0', '2.0', '20.0', '12.0', '8.0', '6.0', '4.0', '11.0', '14.0', '1.0', '17.0'}

Type: SubGraph

Number of nodes: 15

Number of edges: 28

Average degree: 3.7333

Network diameter = 3

Network density = 0.26666666666666666

Network average clustering coefficient = 0.6987301587301586

Network average shortest path length= 1.819047619047619