Report

Zhiyi Jin

I choose three networks from the “networkdata” package. They are Brunson Corporate Leadership network, Radoslaw Email network, and Residence Hall Friendship network. Brunson Corporate Leadership network is undirected and has 44 nodes, while Radoslaw Email and Residence Hall Friendship network are directed with 167 and 217 nodes respectively. Thus, they have some variations in terms of the size and direction for further analysis.

Exercise 1: Descriptive Network Statistics

Table 1: Descriptive Network Statistics

Network	Density	Average Degree	Average Indegree	Average Outdegree	SD Degree	SD Indegree	SD Outdegree	Reciprocity	Transitivity	Average Shortest Path Length	Connected	Isolates
Brunson Corporate Leadership	0.10	9.00	•	•	5.45	•	•	•	0.00	2.76	TRUE	0
Radoslaw Email Network	0.21	69.26	34.63	34.63	57.18	26.87	31.51	0.88	0.54	1.84	TRUE	0
Residence Hall Friendship	0.06	24.63	12.31	12.31	11.16	6.31	6.76	0.62	0.30	2.76	TRUE	0

Note that the Radoslaw Email Network has a high level of density, average degree, and reciprocity. The average shortest path is 1.84. There are on average one step along the shortest paths for all possible pairs of network nodes. The transitivity is 0.54, thus over 50% of connected triples close to form triangles. It implies high efficiency of information flow in the network of email connection. And this network is completely connected without any isolates. However, there does exist one strongly connected component within the network including 126 nodes(over 126/167 = 75% of the nodes) in the network. Residence Hall Friendship network has a similar characteristic. The network seems to be weakly connected but not strongly. 214 nodes strongly connected in a large component. Brunson Corporate Leadership network does not have this feature though.

```
is.connected(radoslaw_email_m, connected = "strong")
```

```
## [1] FALSE
```

```
scc_2 <- clusters(radoslaw_email, mode = c("strong"))
table(scc_2$csize)
```

```
##
##   1 126
##  41   1
```

```
is.connected(hall_m, connected = "strong")
```

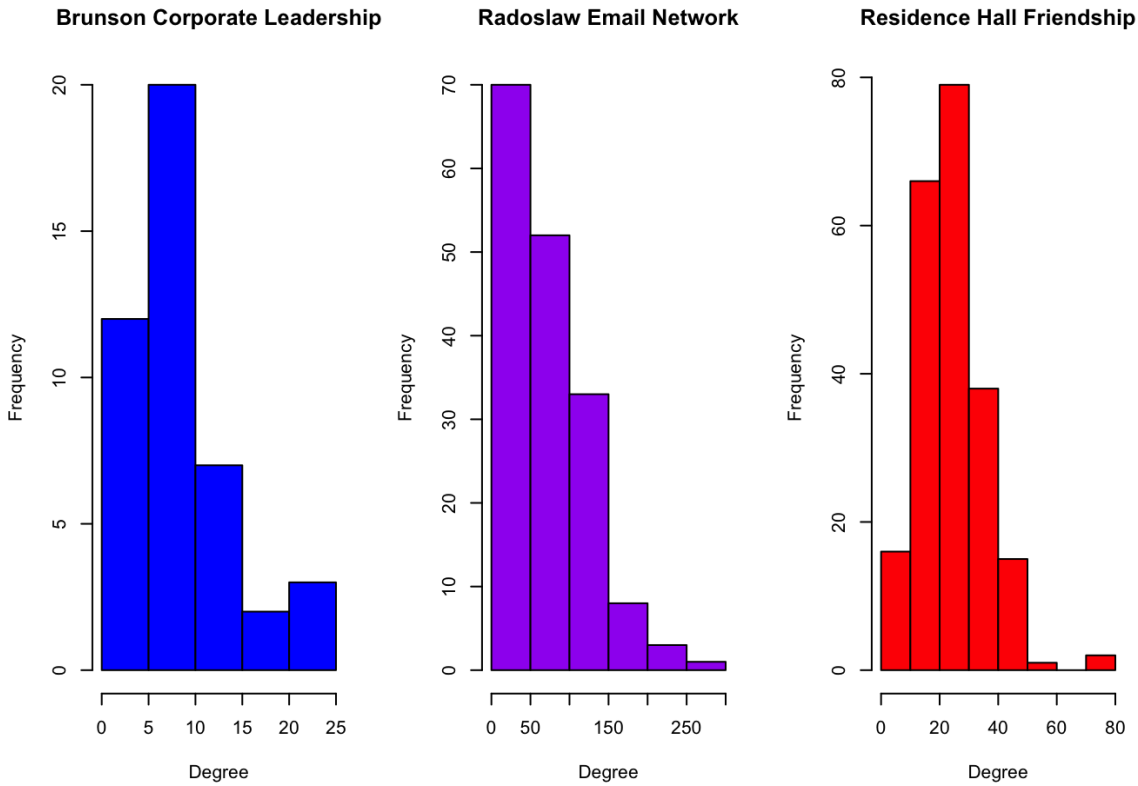
```
## [1] FALSE
```

```
scc_3 <- clusters(hall, mode = c("strong"))
table(scc_3$csize)
```

```
##
##   1 214
##   3   1
```

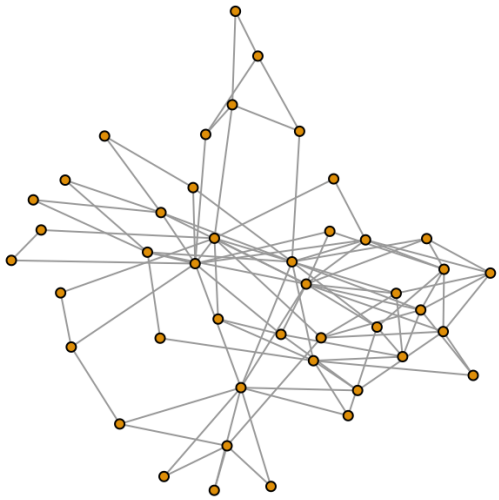
Exercise 2: Degree distribution

It can be seen from the following graphs that in the Brunson Corporate Leadership network there are 3 most highly connected nodes, while most nodes are connected to 5-10 nodes. In the Radoslaw Email Network, most nodes are connected to at least 50 nodes. This may not be a surprise considering its high average degree of 69 nodes. And this network also has a large standard deviation of average in degree in Table 1, and this variation is even more clear in the figure. Note that in the Residence Hall Friendship Network, few nodes are either extremely highly connected with 80 nodes or poorly connected to even less than 20 nodes, while most of them are connected to around 30 nodes, which approximate the shape of normal distribution.

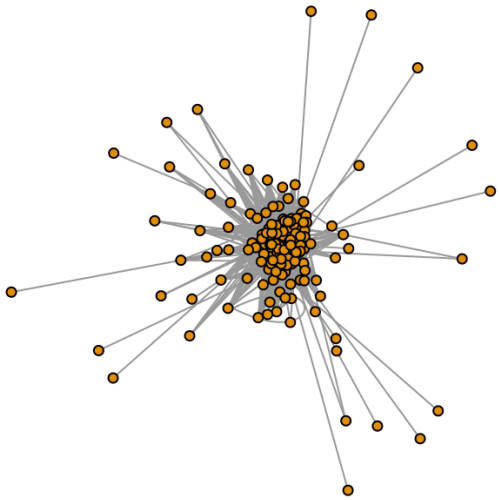


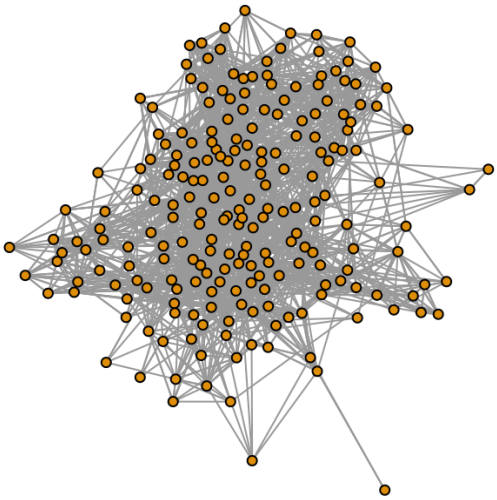
Exercise 3: Visualization of the networks

Brunson Corporate Leadership



Radoslaw Email Network





Exercise 4: Community detection

(1) Brunson Corporate Leadership network

```
# community detection for -brunson_corporate_leadership-
b <- fastgreedy.community(brunson_corporate_leadership)
length(b)
```

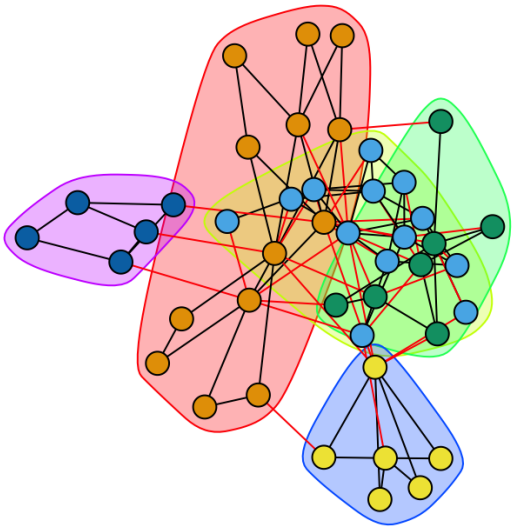
```
## [1] 5
```

```
sizes(b)
```

```
## Community sizes
##  1  2  3  4  5
## 13 13  7  6  5
```

I implemented a fast-greedy approach to the Brunson Corporate Leadership network. The result of this method has declared there to be five communities in the network. There are two largest communities containing 13 nodes. Nodes within these two communities are indicated as yellow and light blue in the following plot.

Brunson Corporate Leadership(Fast-Greedy)



(2) Radoslaw Email Network

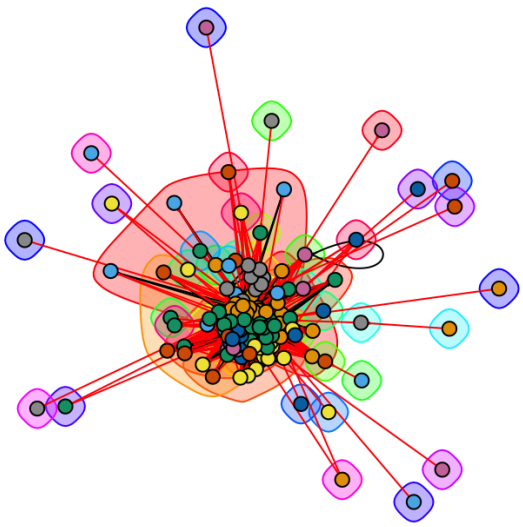
```
r <- walktrap.community(radoslaw_email)
length(r)
```

```
## [1] 47
```

```
sizes(r)
```

```
## Community sizes
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
## 19  7 38 17 12  8  3  9 16  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46 47
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
```

For the Radoslaw Email Network, as a directed network, I used the walk trap method. The result indicates 47 communities. However, 38 of them are isolated nodes. The largest community has 38 nodes.



(3) Residence Hall Friendship network

```
h <- walktrap.community(hall)
length(h)
```

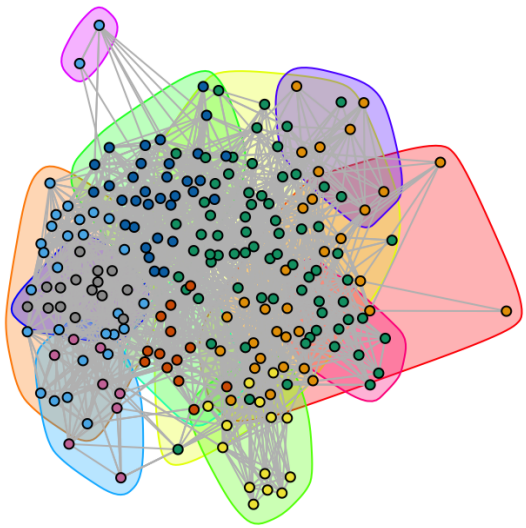
```
## [1] 11
```

```
sizes(h)
```

```
## Community sizes
##  1  2  3  4  5  6  7  8  9 10 11
## 25 27 59 14 33 13  8 15  8  2 13
```

A similar approach was applied for the Residence Hall Friendship network. 11 communities are detected. The largest one contains 59 nodes represented by green dots. The smallest one includes only two nodes.

Residence Hall Friendship(Walktrap)



Exercise 5: Symmetrize the data

```
# Symmetrize the data
n2 <- as.undirected(radoslaw_email)
n3 <- as.undirected(hall)

r1 <- walktrap.community(n2)
length(r1)
```

```
## [1] 46
```

```
sizes(r1)
```

```
## Community sizes
##  1  2  3  4  5  6  7  8  9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26
##  8  9 38 17 12 26  8  9  3  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
## 27 28 29 30 31 32 33 34 35 36 37 38 39 40 41 42 43 44 45 46
##  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1  1
```

```
h1 <- walktrap.community(n3)
length(h1)
```

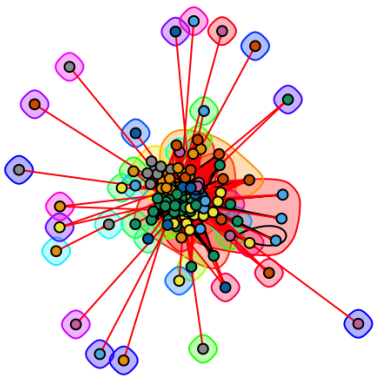
```
## [1] 7
```

```
sizes(h1)
```

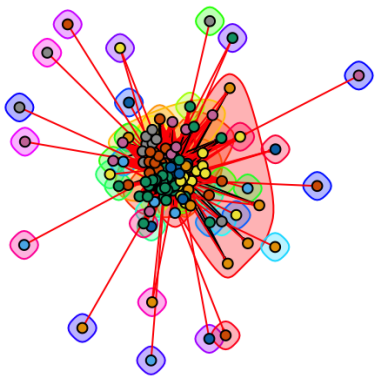
```
## Community sizes
##  1  2  3  4  5  6  7
## 70 53 30 35 14 13  2
```

I symmetrized the Radoslaw Email Network and Residence Hall Friendship network and applied the same community detection algorithm. Note that the number of communities decreases from 47 to 46 in the Radoslaw Email Network and from 11 to 9 in the Residence Hall Friendship network. The community size in the Radoslaw Email Network does not change very much. But the size of the largest community in the Residence Hall Friendship network changes to 70 in Residence Hall Friendship network. Overall, the difference does not seem to be very significant.

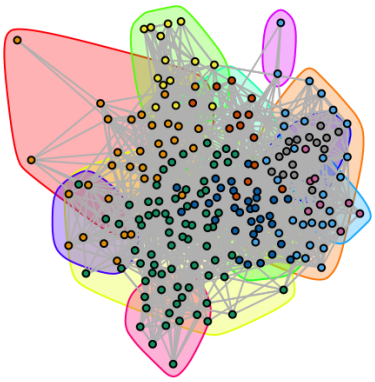
Radoslaw Email Network



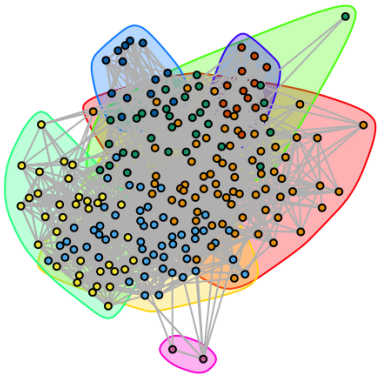
Symmetrized



Residence Hall Friendship

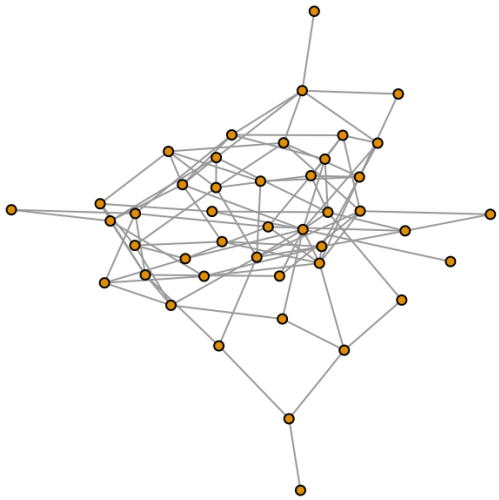


Symmetrized

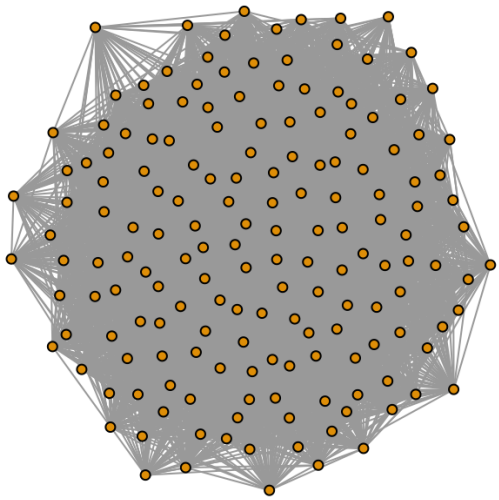


Exercise 6: Random graph “counterpart”

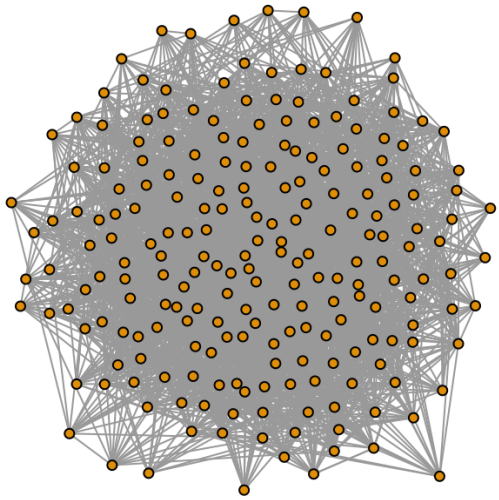
Brunson Corporate Leadership(counterpart)



Radoslaw Email Network(counterpart)



Radoslaw Email Network(counterpart)



Exercise 7: Small network test

Table 2: Descriptive Network Statistics(Random graph counterpart)

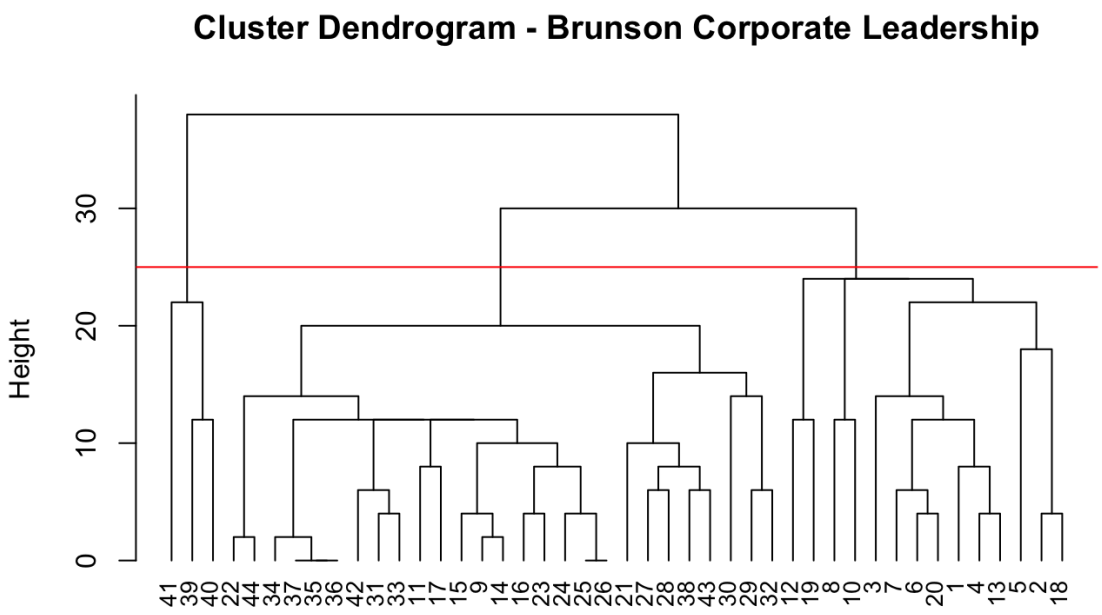
Random Graph	Average Shortest Path Length	Transitivity
Brunson Corporate Leadership	2.63	0.07
Radoslaw Email Network	1.79	0.21
Residence Hall Friendship	2.42	0.06

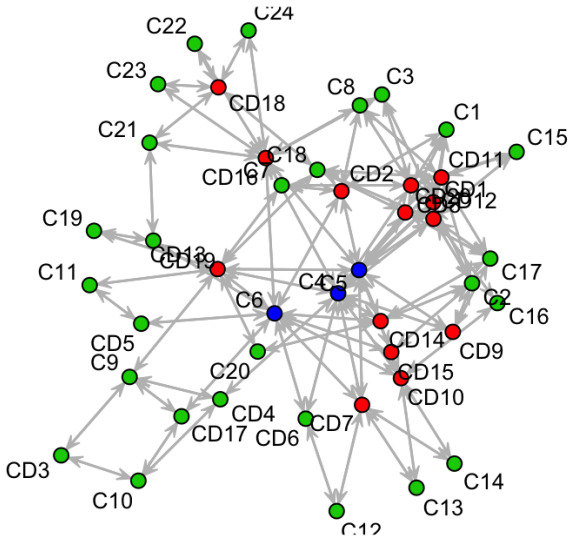
Comparing the statistics in Table 2 to Table 1, the Brunson Corporate Leadership network has a relatively higher shortest path length and lower clustering coefficient than its counterpart of Erdős–Rényi random graph. Thus, it can not be a small-world network, which tends to have a high clustering coefficient and small shortest path length. However, the Radoslaw Email Network has a very similar average shortest path length to its Erdős–Rényi random graph counterpart. In the meantime, the transitivity(1.84 in Table 1) is higher than its counterparts (0.21 in Table 2), which indicates a probability for a small-world network. Residence Hall Friendship network could also a small-world network. Its average shortest path length is 2.76 (see Table 1), which is not very different from 2.41 of its random graph(see Table 2). But the clustering coefficient is much higher than that of its random graph.

Exercise 8: Structural equivalence blockmodels

(1) Brunson Corporate Leadership Network

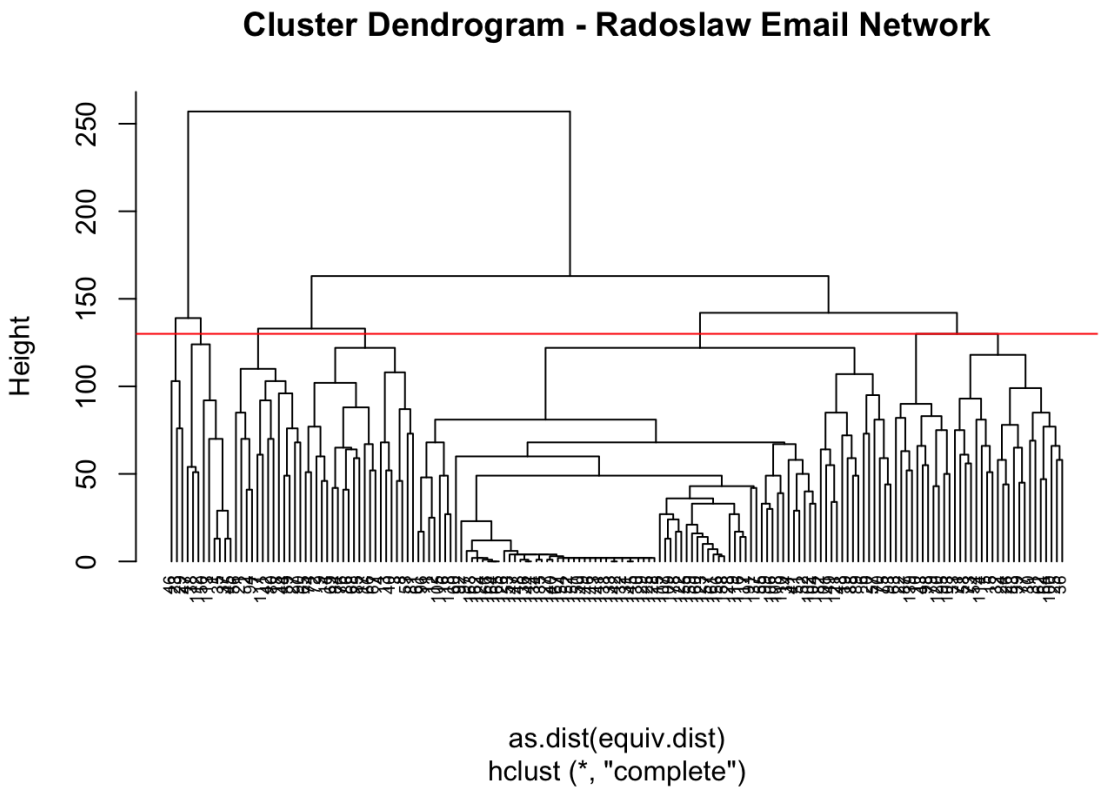
I partition the nodes in the Brunson Corporate Leadership network into 3 positions based on their structural equivalence, namely the identical sets of ties one node has to other nodes, in which an indirect measure of equivalence was adopted. It can be seen from the plot that nodes C4, C5, C6(in blue) are “central” and connected in the middle. The remaining red nodes hanging on to these, other green nodes are relatively poorly connected and peripheral.

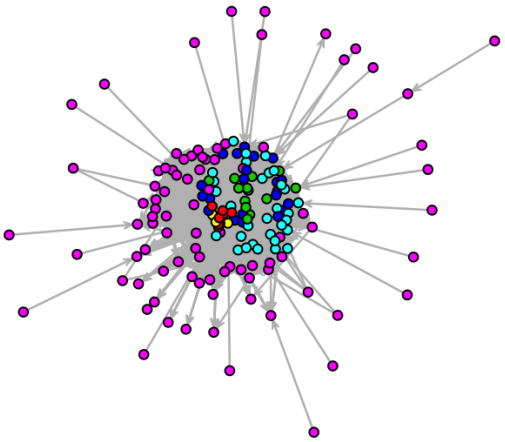

Cluster Dendrogram - Brunson Corporate Leadership
Height
0
10
20
30
41 39 40 22 44 34 37 35 36 42 31 33 11 17 15 9 14 16 23 24 25 26 21 27 28 38 43 30 29 32 12 19 8 10 3 7 6 20 1 4 13 5 2 18
as.dist(equiv.dist)
hclust (*, "complete")


C24
C22
C23
CD18
C8
C3
C1
C15
C21
C18
CD11
CD16
C7
CD2
CD1
CD12
C19
CD13
CD19
C11
C4
C5
C17
C6
C2
C16
CD5
CD14
CD9
C9
C20
CD15
CD10
CD4
CD7
CD17
CD6
CD3
C14
C10
C13
C12

(2) Radoslaw Email Network

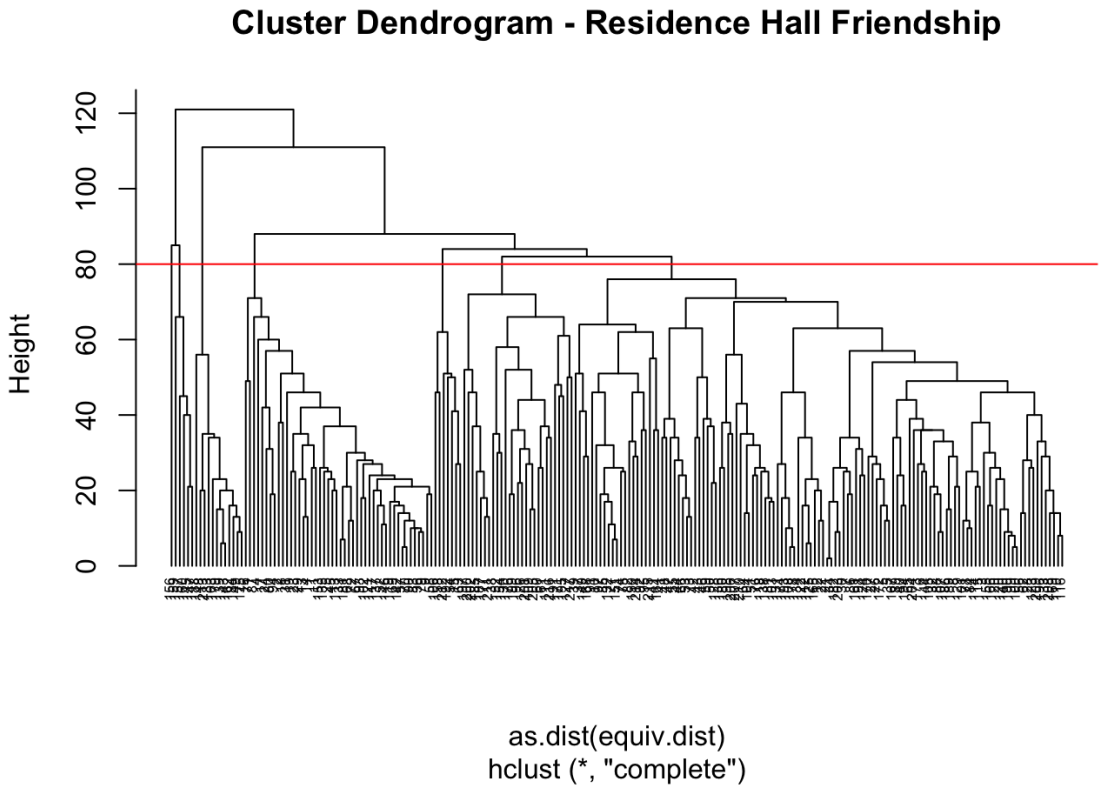
For the Radoslaw Email Network, I partition it into 6 positions. Several small groups(red, green, and yellow) cluster in the middle, while most pink nodes are peripheral and isolated. Most of them only have one outward tie or few inward ties. In general, this network seems very centered on a small group of nodes.

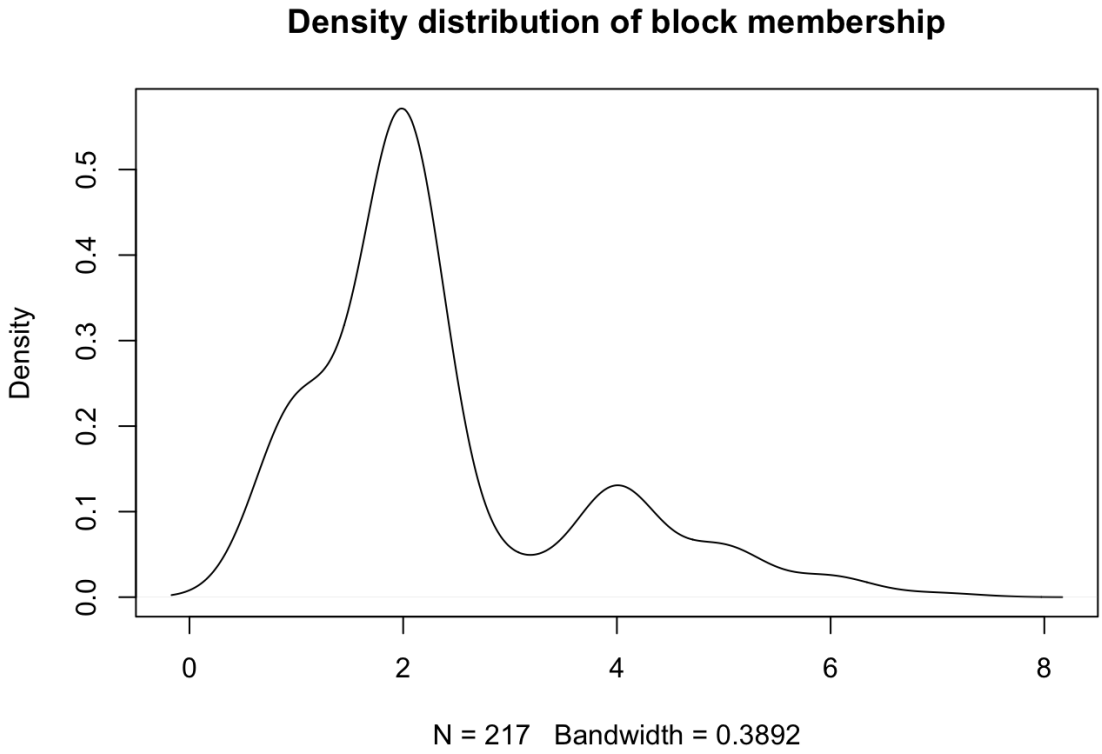


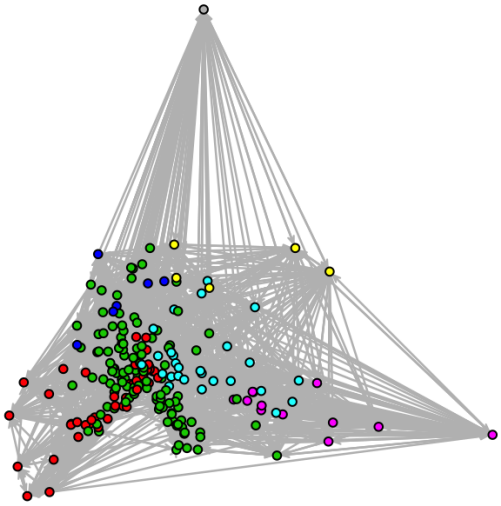


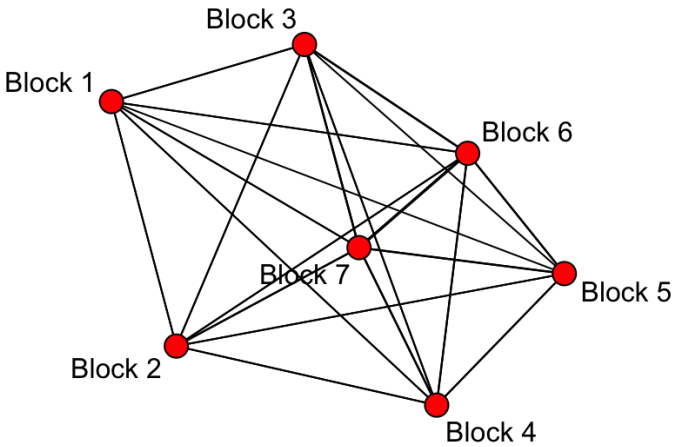
(3) Residence Hall Friendship network

As for the Residence Hall Friendship network, I partition it into 7 positions. Note from the density distribution of block membership that the majority of the nodes fall into block 2, making it the largest group. However, it can also be seen from the reduced version of formed blocks that block 7 including the relatively small number of nodes is centrally connected. Overall, looking at the final plot, the network is very compact with many densely connected nodes and quite a few sparsely connected nodes on the right side.


Cluster Dendrogram - Residence Hall Friendship
Height
0
20
40
60
80
100
120
as.dist(equiv.dist)
hclust (*, "complete")


Density distribution of block membership
Density
0.0
0.1
0.2
0.3
0.4
0.5
0
2
4
6
8
N = 217 Bandwidth = 0.3892




Block 3
Block 1
Block 6
Block 7
Block 5
Block 2
Block 4

Exercise 9

For me, the Radoslaw Email Network is most interesting. In a theoretical sense, the email network data illustrate the communication network between employees of a mid-sized manufacturing company, which could have certain implications on the health of the company by studying the information exchange along with the structural hierarchy. In a network sense, this email network has many communities of small size and isolated nodes that either only send or receive the email. The graph also looks like a star with one large internal group and many leaves. A possible hierarchy distribution of information might exist in this company.