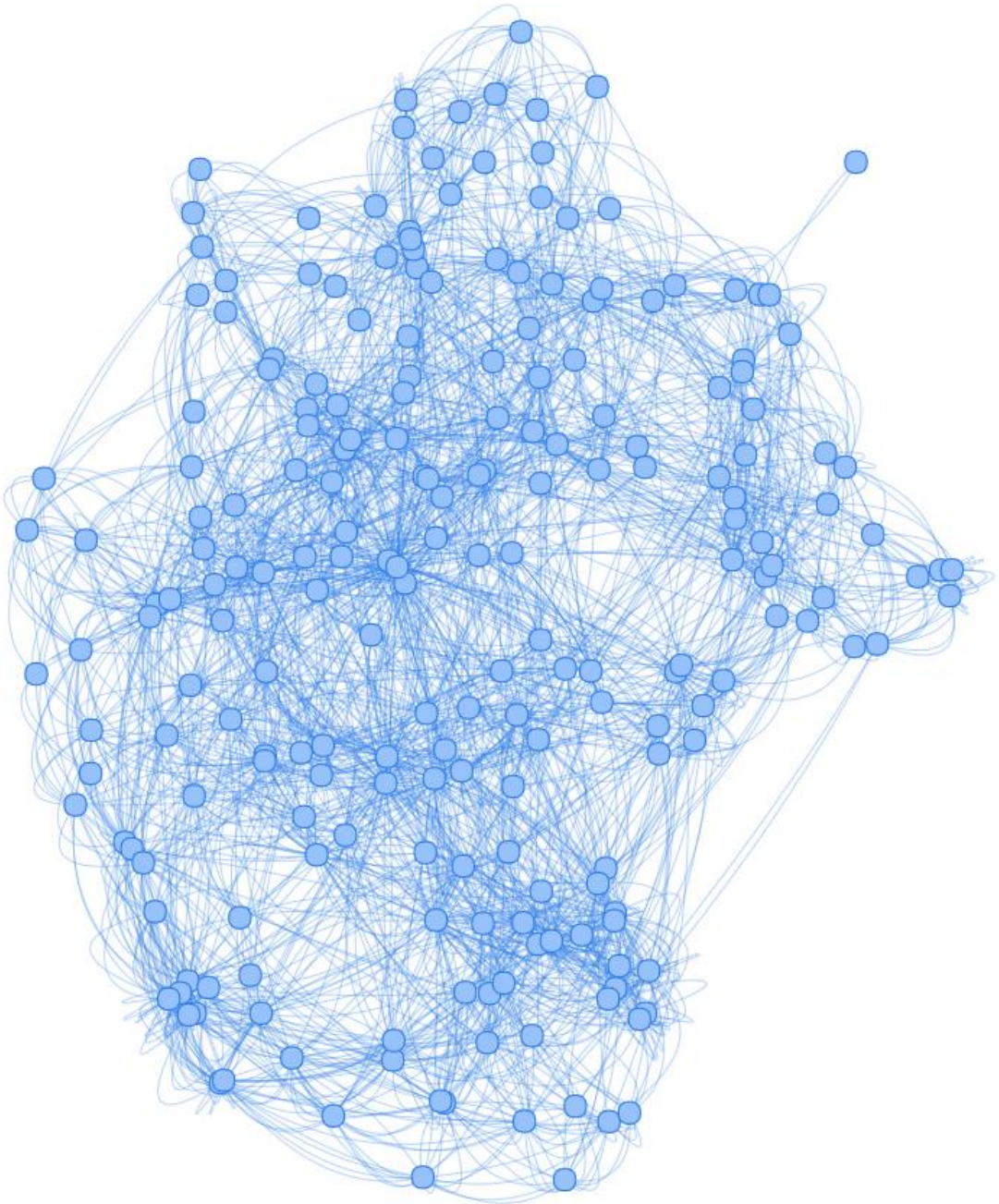
A1: Network Analysis

Elsa Scola Martín

April 2, 2020



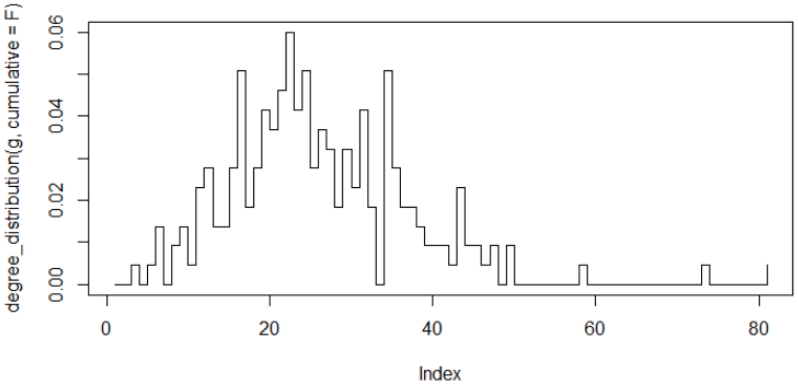
1 Intro

This deliverable aims to present the main properties, as well as the analysis and visualization of the Residence Hall network. In particular, this network comprises of friendship ratings between 217 residents living at a residence hall located on the Australian National University campus.

2 Analysis of the network

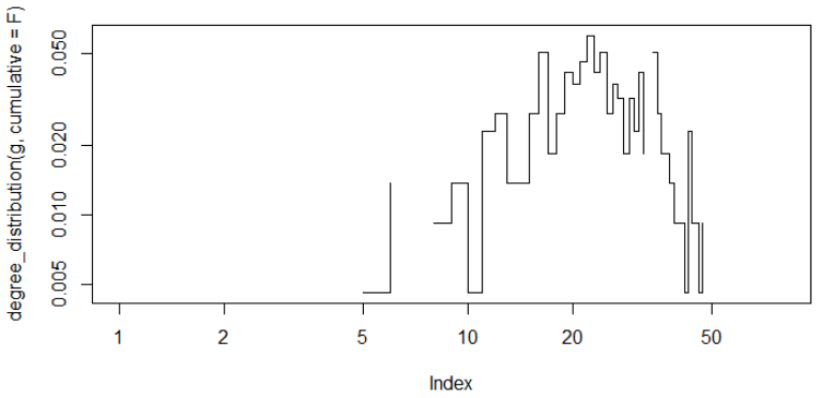
What are the network features?
The Residence Hall network is initially provided as a directed weighted adjacency matrix. This network is composed of the following features: there are 217 vertices, each one represents a resident, and 2672 edges, that represent the friendships among the residents. It is a directed graph in which the weights represent the level of friendhip, that is, the greater the weight the closer is this friend for the resident. The weights range from 1 to 5.

Plot the degree distribution in linear-linear scale and in log-log scale. What is the average/maximum/minimum degree?
Degree distribution in **linear-linear scale**:



Average degree: 24.63
Maximum degree: 80
Minimum degree: 2

Degree distribution in log-log scale:



Average degree: 3.0927
Maximum degree: 4.3820
Minimum degree: 0.6931

As it can be seen, the majority of degrees lie between 17 and 31, moreover we can notice the distribution is right skewed, therefore it is not normally distributed. However, the log transformation makes it a bit more gaussian. We can also notice that the most popular student at the residence is the node *V156* with a degree of 80, which means 80 residents considered him as a friend.

What is the assortativity (Pearson) coefficient in the network?

In some networks, nodes with high degree are more often connected with other nodes of high degree and low degree with low degree. This property is called assortativity and it's defined by ρ as the Pearson correlation coefficient between the degrees of connected nodes.
In this case the Pearson coefficient ρ is 0.0805741, as it is almost zero, we can conclude that the residents don't have a particular tendency of being friends with residents similar to them (similar, as in social characteristics).

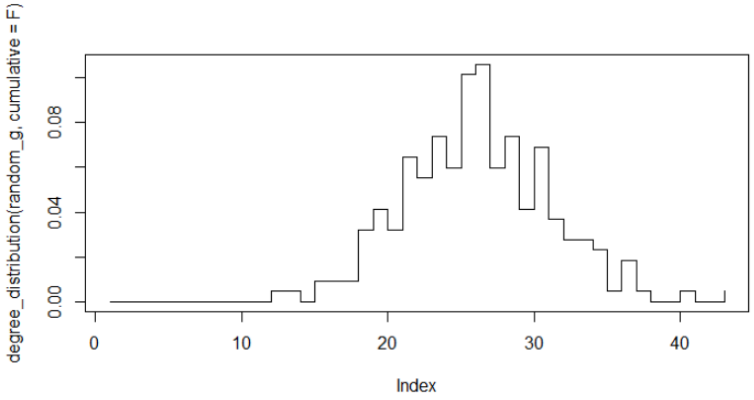
What is the clustering coefficient in the network? Specify if you calculated the global or the average local clustering coefficient, or both.

The global clustering coefficient is 0.3036058, this number represents the ratio of the triangles and the connected triples in the graph. For directed graph the direction of the edges is ignored.

What is the average shortest path length of the giant component?

The average shortest path is 2.764981, this number shows that on average the shortest paths to arrive from one node to another are of around three nodes.

Create an ER random graph with the same number of nodes and links and compare its statistics with those of your chosen network.



The plots obtained as for the degree of the randomly generated network are more normally distributed, as expected. The assortativity degree is very close to zero (0.04), which reinforces our previous ideas of how the residents where friends of people regardless of their social characteristics. The clustering coefficient (0.11) is even lower than the one obatined on the chosen network, as it is random and there aren't any particular concentrations of friendships. However, 2.41 the average shortest path it's pretty close to the one observed in the chosen network.

Create a configuration model network with the same degree distribution of your chosen network and compare its statistics with those of your network.

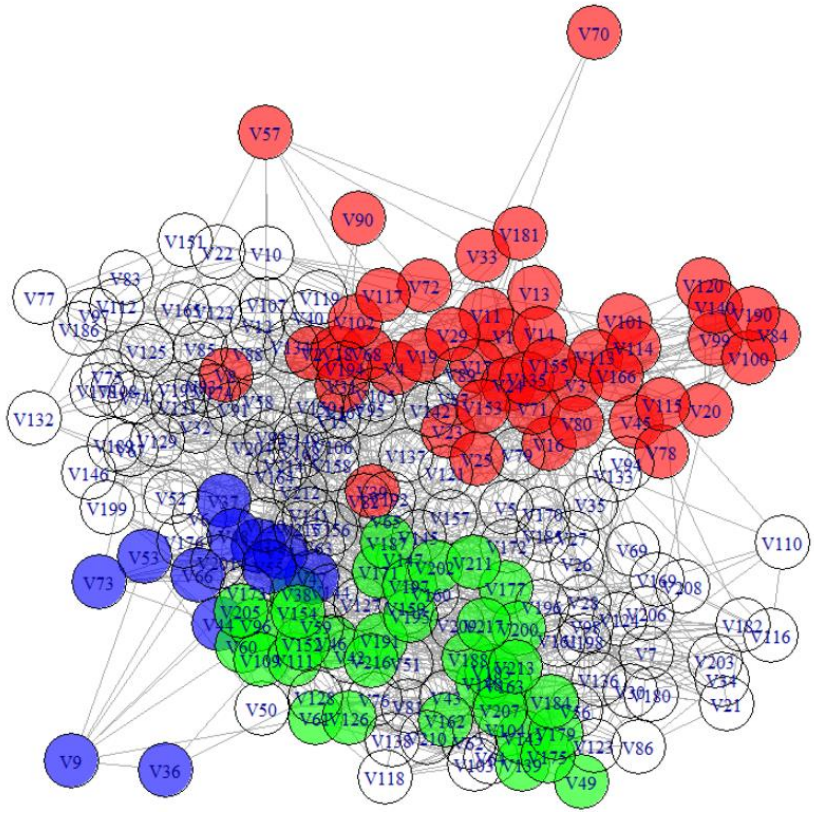
Expectedly, as the degree was fixed in advance, the metrics that describe the degree (mean, max, plots…) are the same as with the chosen data. However, there are some noticeable changes in the statistics: in this case the assortativity is negative (-0.017), but still very close to zero, which reinforces the ideas discussed in the previous section. The clustering coefficient (0.13) is also small as in the randomly generated network, and the

average shortest path distances more from the one of the chosen network with a value of 2.01.

Using the Louvain method, does your network show a community structure?

As the initially chosen graph is directed and the *cluster_louvain* function only works for undirected, we first need to convert it to undirected, for which I use the *as.undirected* function with the *collapse* mode, so we don't obtain repeated edges among the same node pair.

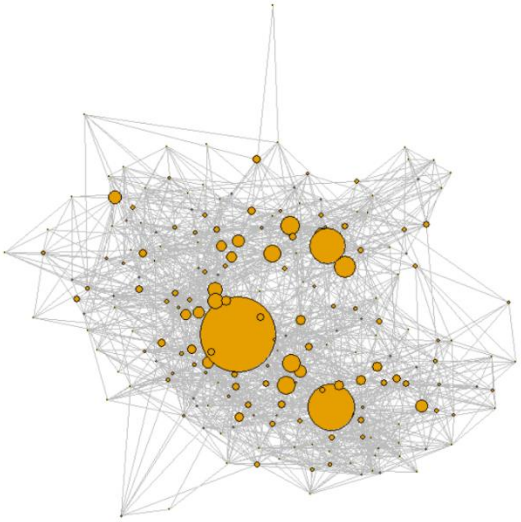
The result is the following:



As it can be seen, there could be some communities in this network, nevertheless, they are very light, which makes it look like a big community. This supports the idea commented above, that indicates the absence of an elevated assortativity degree.

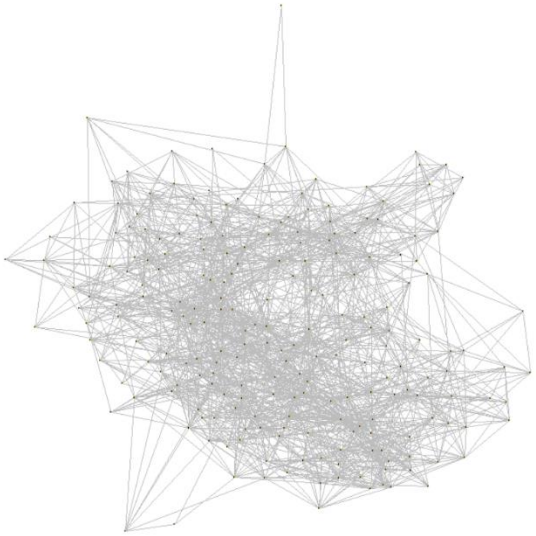
Visualize betweenness and closeness centrality measures of the network.

The betweenness centrality of the nodes in our graph, which is the number of shortest paths through the network of which a node is a part.

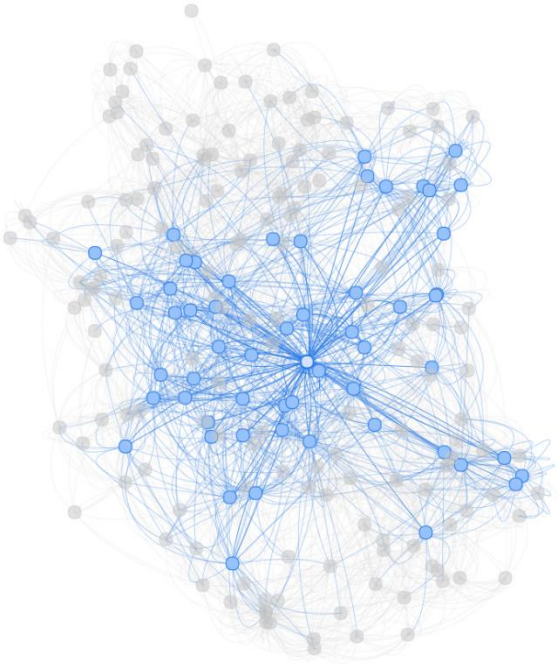


The biggest nodes are the ones act as “bridges” between the nodes of the network, that means, that these are most influential residents.

On the other hand, closeness centrality represents how close any one node is on average to everyone else.



This measure lets us find the residents that are best placed to influence the entire network most quickly, therefore, if in the residence there were to be any announcement to make, the individuals who would broadcast it better would be the ones with higher closeness centrality.



We can also get the in-degree centrality of the most popular resident, which I mentioned above.

Visualize the network showing, if any, its communities.
As it can be seen, this network does not show any particular communities, which makes sense after having a look at the statistics shown above.