Report of Final Paper

MSc in Bioinformatics for Health Sciences UPF

Student: Iñigo Oyarzun Lafuente

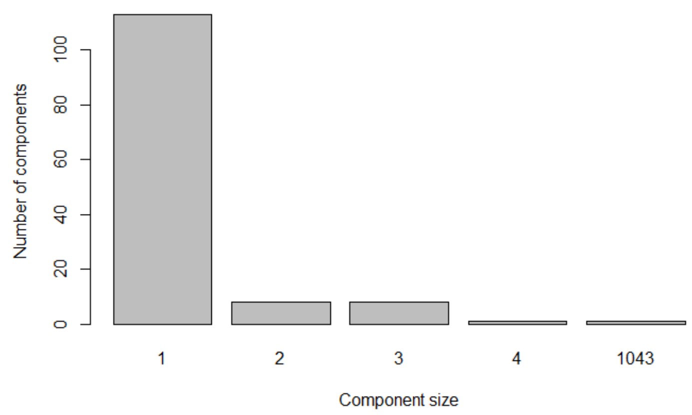
Erdős-Rényi networks

- Erdős-Rényi network N=1200 p=0.001

```
is.connected(erdos0.001)
[1] FALSE
average.path.length(erdos0.001)
[1] 7.479414
```

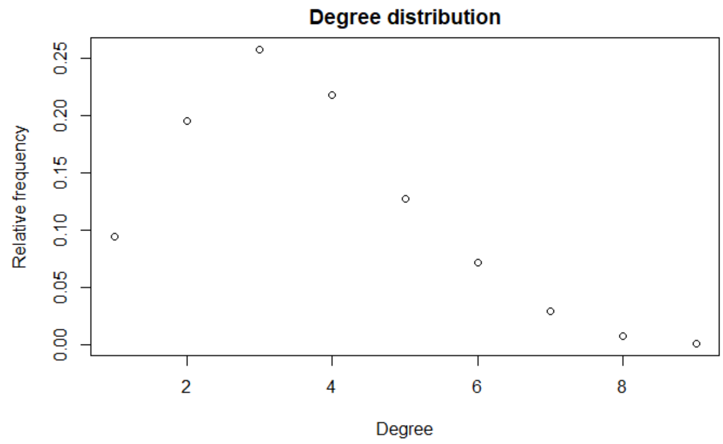
We can see that this network is not connected and the average path length is 7.48. As it is a not connected network, we decided to plot the distribution of the component sizes.

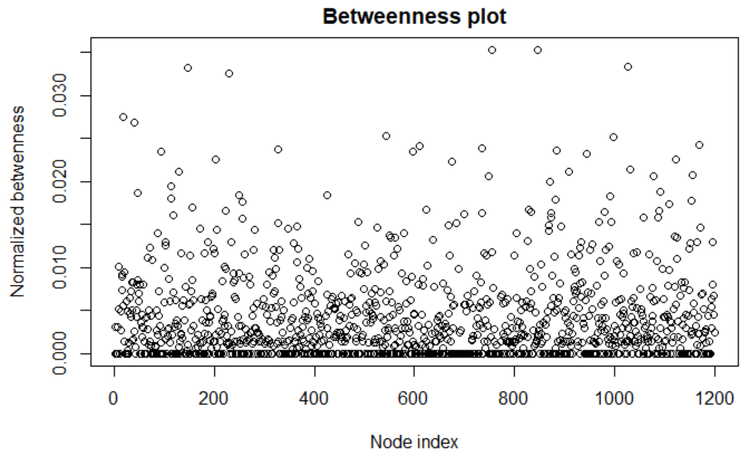
```
barplot(table(components(erdos0.001)$csize))
```



We can observe how most of the components just have one isolated node and we have a huge component that contains 1043 nodes.

```
plot(degree.distribution(erdos0.001))
plot(betweenness(erdos0.001, normalized = T))
```

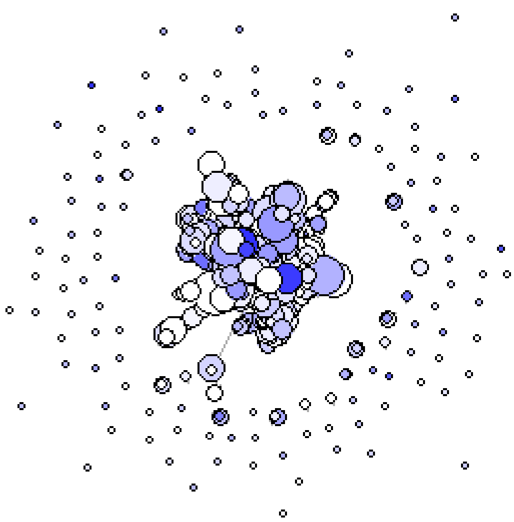




We can see that this network's degree distribution follows a Poisson distribution. The betweenness is randomly distributed along the nodes in the network with many zero values due to the low connectivity of this network.

Network visualization coloured by betweenness and node size depending on the degree:

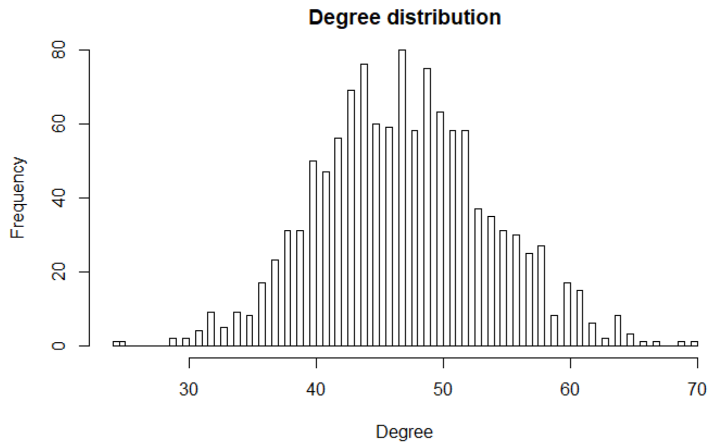
```
plot(erdos0.001,vertex.label=NA,vertex.size=degree(erdos0.001)*2+3,
vertex.color= cols[betweenness(erdos0.001,normalized = TRUE)*1000])
```

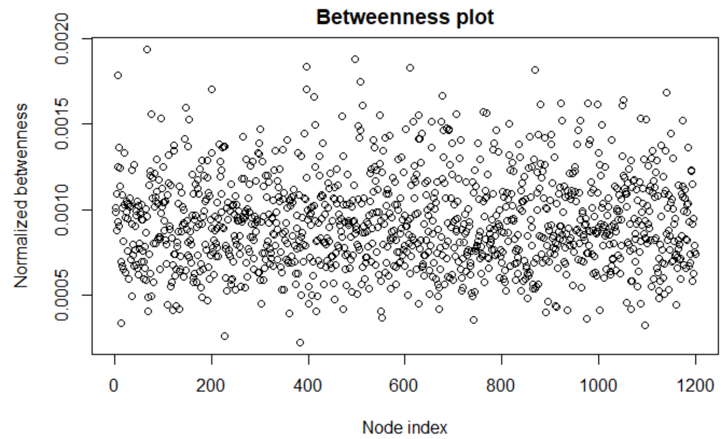


- Erdős-Rényi network N=1200 p=0.02

```
is.connected(erdos0.02)

[1] TRUE

average.path.length(erdos0.02)

[1] 2.109817
```

This network is connected so it only has one component with 1200 nodes. The average path length is 2.11 which is a low value for a big network.

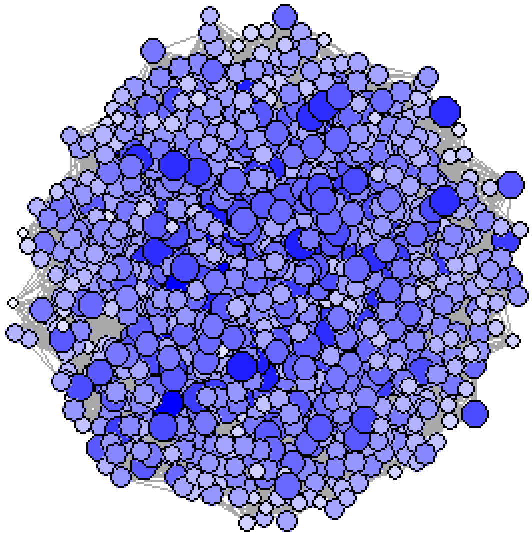




The degree distribution is bell-shaped following a normal distribution. The betweenness is randomly along the nodes along the nodes.

Network visualization coloured by betweenness and node size depending on the degree:

```
plot(erdos0.02,vertex.label=NA,vertex.size=degree(erdos0.02)*0.2,
vertex.color= cols[betweenness(erdos0.02,normalized = TRUE)*10000])
```



The main characteristic of an Erdős-Rényi network is that each of the possible links in the network is added given a certain probability, parameter p. As this p is equal for all the nodes, this model creates random networks.

Analysing these two Erdős-Rényi networks we can see how incrementing the parameter p the degree distribution goes from a Poisson distribution to a normal distribution. Also, as higher is p as lower will be the average path length, which makes perfectly sense in a random network as the nodes will be more connected and the shortest path between them will be lower. So, this show us how the small-world property is reduced in poorly connected random networks.

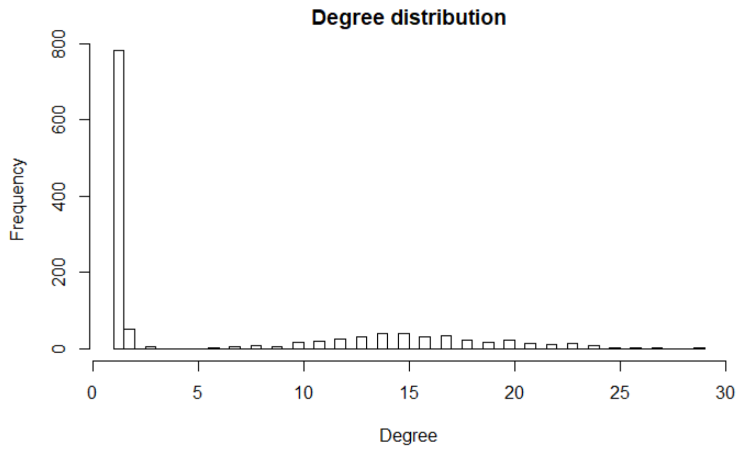
Moreover, we can see that when p is higher there are no zero values for betweenness this is because as the nodes have more links the probability of they to be part of the shortest path way between two other nodes increase.

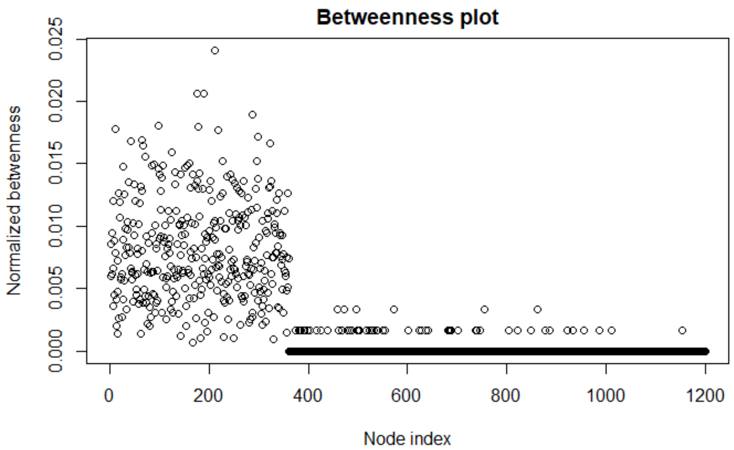
Barabási-Albert networks

- Barabási-Albert N=1200 E=1

```
is.connected(bara1)
[1] TRUE
average.path.length(bara1)
[1] 4.022274
```

The network is connected so it has a single component with 1200 nodes and the average path length is 4.022.

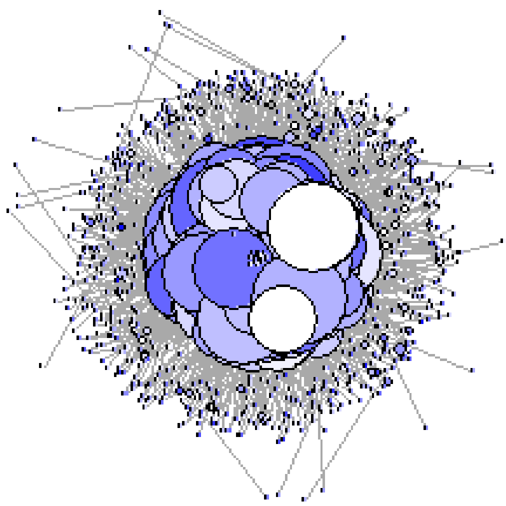




We can observe a bimodal distribution in the degree distribution where most of the nodes have one degree level which are the new arrival nodes (E=1) and then other nodes, most likely the core nodes, have a much higher degree. Something similar happen with betweenness. In the script we settled up the core percentage at 30% of the network size, thus we can see how the core nodes have a much higher betweenness than the new arrival nodes. This show how the core nodes have much higher control on the information of the network.

Network visualization coloured by betweenness and node size depending on the degree:

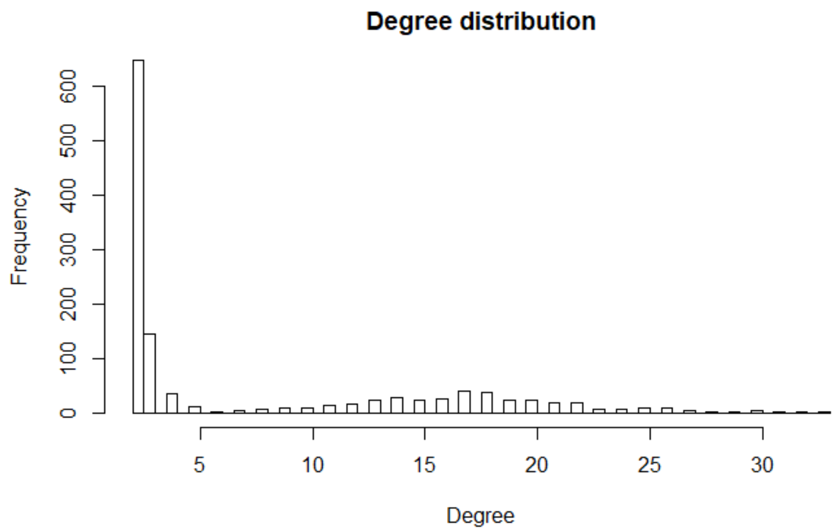
```
plot(bara1, vertex.label=NA, vertex.size=degree(bara1)*2, vertex.color=cols[betweenness(bara1,normalized = TRUE)*1000],layout=layout_with_kk(bara1))
```



- Barabási-Albert N=1200 E=2

```
is.connected(bara2)
[1] TRUE
average.path.length(bara2)
[1] 3.614975
```

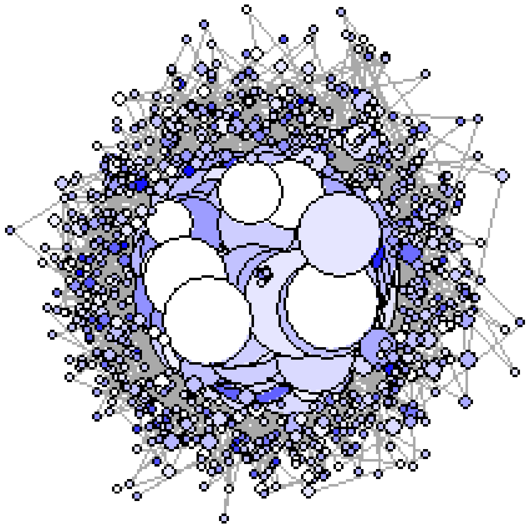
The network is also connected in a single component with 1200 nodes and the average path length is 3.61.



Again, we observe a bimodal distribution in the degree distribution where most of the nodes have two degrees level which are the new arrival nodes (E=2) and then other nodes, most likely the core nodes, have a higher degree. Same with betweenness, we see again how the core nodes have a much higher betweenness than the new arrival nodes.

Network visualization coloured by betweenness and node size depending on the degree:

```
plot(bara2,vertex.label=NA,vertex.size=degree(bara2)*2,        vertex.color=
cols[betweenness(bara2,normalized = TRUE)*1000],layout=layout_with_kk(bara2))
```



Barabási-Albert have a core where the nodes are connected randomly and after that, new nodes are connected E times to the nodes in the network. But, the main characteristic of this model is that these new nodes will connect preferably with the nodes that have a higher degree, thus, the nodes on the core have more possibilities to receive new links.

Analysing our two Barabási-Albert networks we can clearly differentiate two clusters, the core nodes and the new arrivals nodes. We see that when the parameter E is increased the average path length tends to go down, which makes sense because we will have a more connected network. Still, when E=1, the average path length is 4.022, which is low, thanks to the core nodes that have high values for betweenness centrality. Moreover, even though in the second Barabási-Albert network the new arrival nodes are more connected (E=2), the decrease on the average path length is low due precisely to the fact that both networks share a similar core.