INDIVIDUAL PROJECT

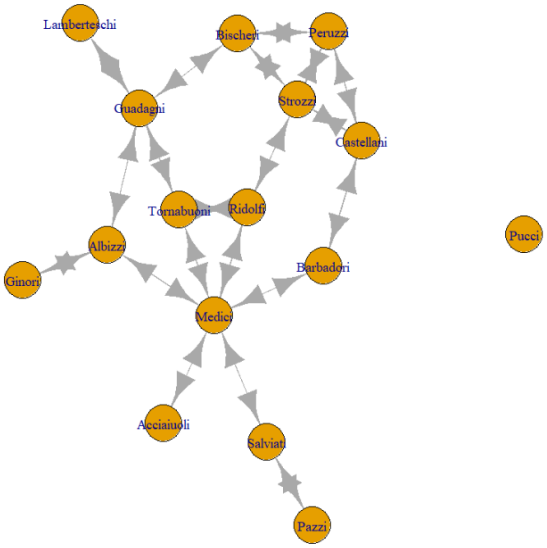
Juan Jose Medina

CENTRALITY MEASURES

What is centrality, centrality is a measure of “importance” for a selected node in a network, but the meaning of importance varies depending on the centrality measure that we select.

DATA

For this project we are going to use the Florentine_M data set, on which we are going to analyze the centrality measure for the following families: Acciaiuoli, Albizzi, Barbadori, Bischeri, Castellani, Ginori, Guadagni, Lamberteschi, Medici, Pazzi, Peruzzi, Pucci, Ridolfi , Salviati, Strozzi, Tornabuoni

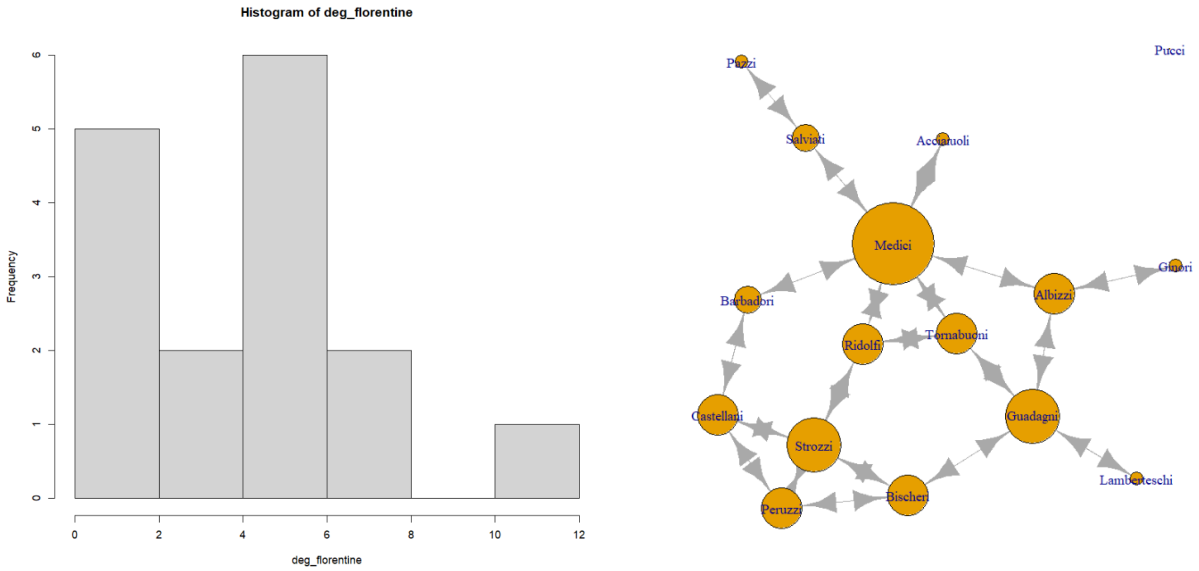


Here we have a visualization on the network that we are going to work with, and we can make initial assumptions just by counting the number of edges that each node has or see that the Pucci family is isolated from the rest of the network, therefore this family is most probable to be on the last place on centrality for multiple measures.

For simplification we assume that the connection between each node is bidirectional, it means that family X is connected to family Y, and family Y is connected to family X.

DEGREE

The degree centrality computes the importance score based on the number of edges that each node has, it tells us how many connection each family has, this can be seen as the popularity of each family



Looking at the distribution of degrees on our network we can see that most of the families have a degree of 5 to 6 and from 0 to 2 while only one has from 10 to 12 links, we need to remember that the edges are bidirectional therefore if a node has only one connection, it will have 2 degrees

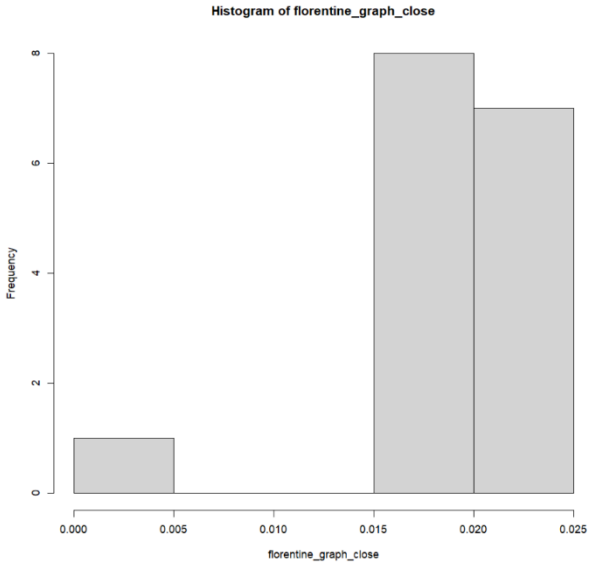
After calculating the degrees for each family we can plot the network giving the size of the node to be in accordance with the degree for each family.

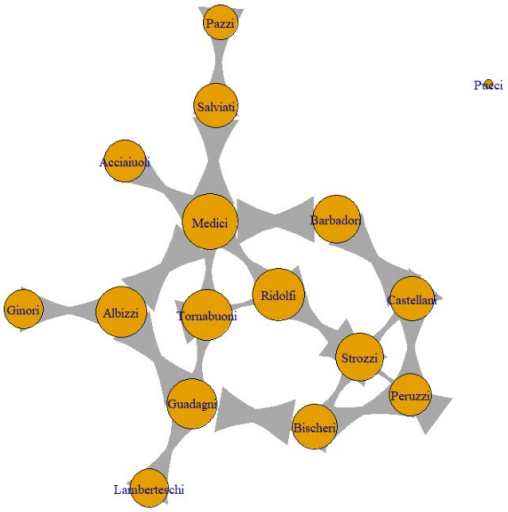
Having the following graph we can see that Medici family is the most popular family having a degree score of 12 with connection to 6 families and the least popular family is the Pucci, having a degree of 0 with not a single connection

```
> names(florentine)[which.max(deg_florentine)]
[1] "Medici"
> names(florentine)[which.min(deg_florentine)]
[1] "Pucci"
```

CLOSENESS

The closeness measure is based on the closeness of one note to the rest of the nodes, this is calculated by measuring the shortest paths and then assigns a score based on the sum of those paths. This can be used to know which family best positioned to have a connection to the most of families.





After calculating the closeness values, we can see that most of the families have high closeness scores and only one that has a low score.

After plotting our network based on the closeness measure we can see that there is almost no difference on families, that most of them are well positioned and have a good influence over the rest of the families.

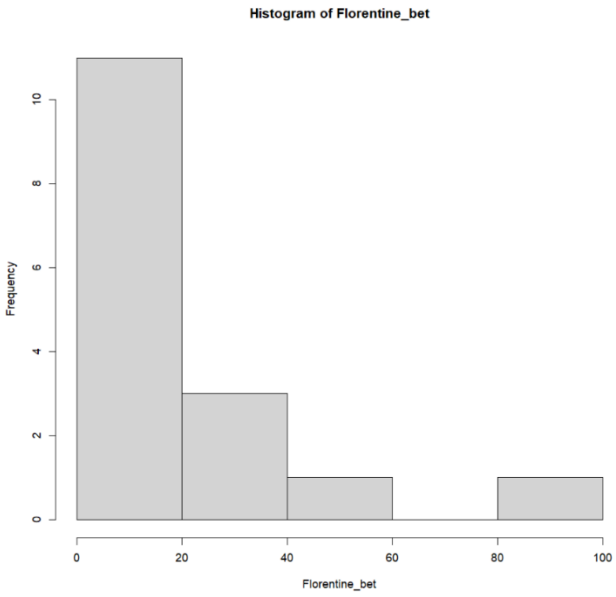
Nevertheless we can see that the Medici family has the highest closeness score between al the families and the Pucci family has the lowest

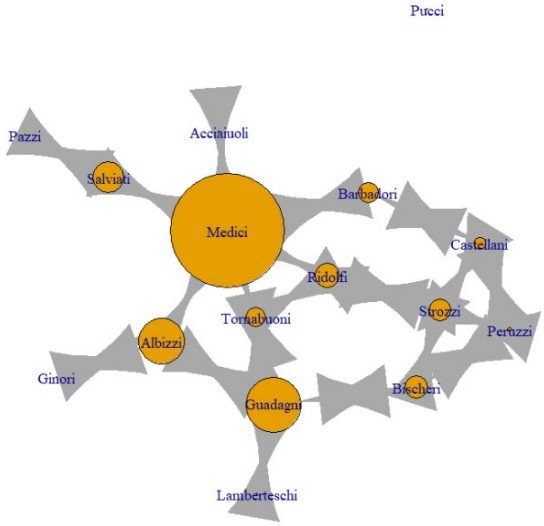
```
> names(florentine)[which.max(florentine_graph_close)]
[1] "Medici"
> names(florentine)[which.min(florentine_graph_close)]
[1] "Pucci"
```

BETWEENESS

This measure takes into account the number of times a node is used as a bridge to get to other nodes taking into account that it must be the shortest path, this can be used to know which family the best intermediary is to get to the rest of the families. The problem with his measure is that it can be miss

interpreted so one has to be careful with that.





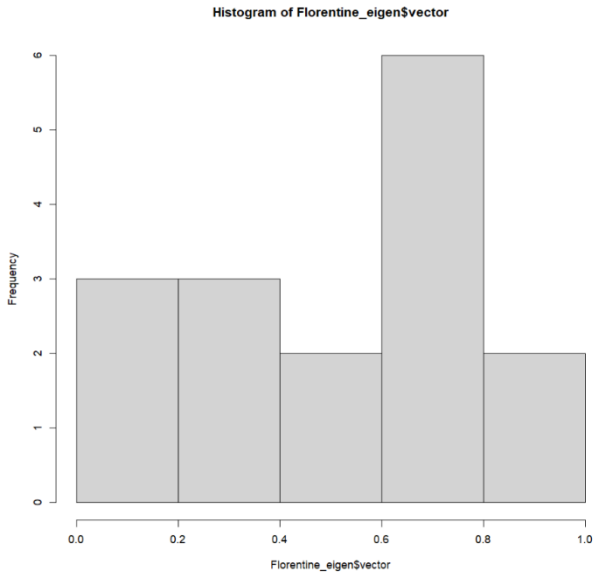
After calculating the betweenness we see that there a lo nodes that have a low betweenes score, this is due that most of the families are on the periphery of the network forming a circle around a node that has the highest value.

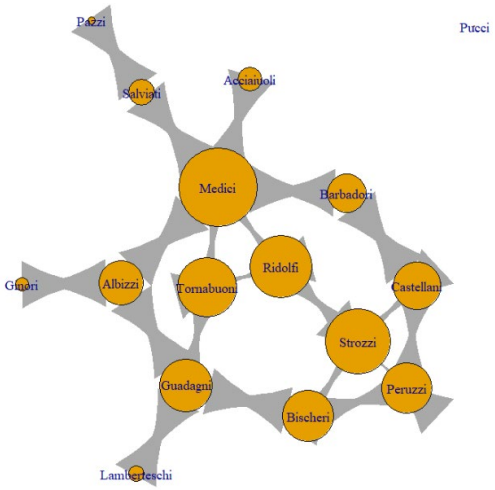
As the graph shows the Medici is the best intermediate, since most of the shortest paths pass throung that familiity and we have not one family that has no score but rather we have multiple, that are only linked to the network by one family, which are the Acciaiuolu, ginori, Lamberteschi, Pazzi and Pucci.

```
> names(florentine)[which.max(Florentine_bet)]
[1] "Medici"
> names(florentine)[which.min(Florentine_bet)]
[1] "Acciaiuoli"
```

EIIGENVECTOR

The eigenvector works similar to the degree measure, it takes into account the number of connections that a node has but also adds the number of connections that the connected nodes have. This measure tell us how much influence a node have over the whole network rather than only at those around the specific node





We can see that after calculating the Eigenvector measure that most of the families have a good Eigenvector score having a almost flat distribution for the rest of the families.

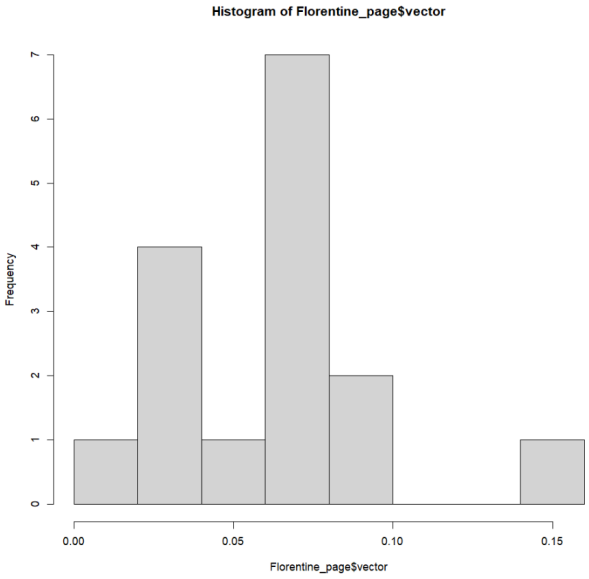
As we see on the graph most of the families have a good influence over the whole network but those that are only connected to the network by one family have a lower eigen Score, such as the Lamberteschi and Pazzi. Since the Medici has the most edged it has the best score and the Pucci the worst one since it doesn't have a single connection.

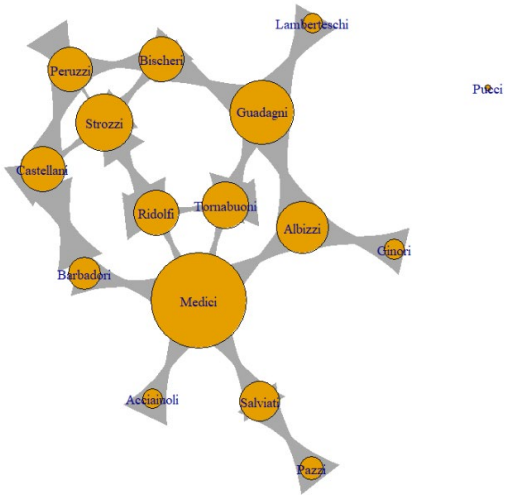
```
> names(florentine)[which.max(Florentine_eigen$vector)]
[1] "Medici"
> names(florentine)[which.min(Florentine_eigen$vector)]
[1] "Pucci"
```

GOOGLE PAGE RANK

The google page rank is a variation of the Eigenvector, it also calculates its score based on the number of edged and the edged that the connected nodes have, the difference is that this also takes into account the direction and weigh, to nodes are only able to pass influence in one direction and different amounts.

This is used to discover which nodes have influence that are beyond their direct connections, as an example this can be used to determine who has more authority on a network.





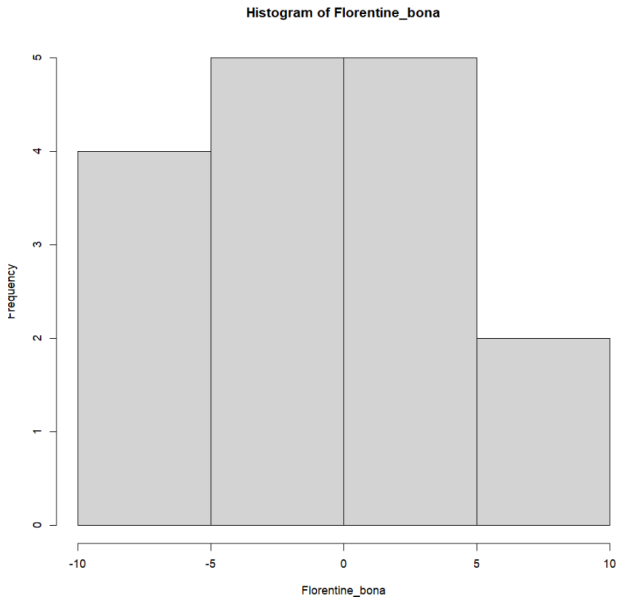
After calculating the google page score, we can see that most of the families have a medium score and only one has the higher score in the network

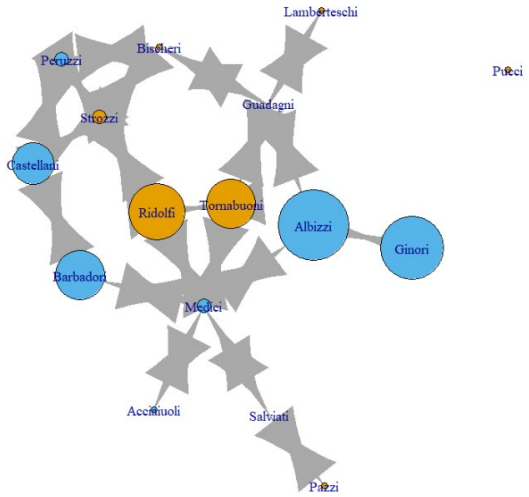
The graph for this centrality measure is like the eigenvector since our network doesn't have weights and is bidirectional, nevertheless the Medici again has the highest value and the Pucci has the lowest.

```
> names(florentine)[which.max(Florentine_page$vector)]
[1] "Medici"
> names(florentine)[which.min(Florentine_page$vector)]
[1] "Pucci"
```

BONACICH ALPHA

This centrality measure takes into account the idea that the more influential your neighbors nodes the more power a specific node has, on this centrality we have positive and negative values , a positive values implies that a node get more importance the more importance their connected nodes are, and a negative value if the relation is the opposite, the node becomes more important as their neighbor nodes get less important.





We can see that we have more nodes that have a negative value meaning that the get more influential the les important their connected nodes are.

We can see that this measure is totally different from what we had, on this measure we can see that the Medici family is less important and other families take its place, since the Medici family has the most edges, but the connected edges don't have much influence its score its diminished and now the Albici Family is the most important with a negative value and Ridolfi is the most important family with a positive value. While Guadagni and Salviati are the least influential families in this model .

Acciaiuoli	Albizzi	Barbadori	Bischeri	Castellani
-1	-10	-7	1	-6
Ginori	Guadagni	Lamberteschi	Medici	Pazzi
-9	0	1	-2	1
Peruzzi	Pucci	Ridolfi	Salviati	Strozzi
-2	1	8	0	2
Tornabuoni				
7				

CORRELATION

	DEGR	CLOSE	BETW	EIGEN	PAGE	BONACICH
DEGR	1.00000000	0.75727783	0.83337627	0.9404647	0.98846265	0.07038153
CLOSE	0.75727783	1.00000000	0.57634871	0.7989789	0.74969920	-0.01143778
BETW	0.83337627	0.57634871	1.00000000	0.6737162	0.89571160	-0.06101905
EIGEN	0.94046471	0.79897890	0.67371624	1.0000000	0.89152411	0.20714536
PAGE	0.98846265	0.74969920	0.89571160	0.8915241	1.00000000	0.03189768
BONACICH	0.07038153	-0.01143778	-0.06101905	0.2071454	0.03189768	1.00000000

On this data set we can see that the Degree measure has many similarities with every measure except with the Bonacich, Closeness has a positive strong correlation with Degree, Eigen and Page.

Only the Bonacich measure it not correlated to any other type of measure, this is supported by analyzing that Medici was the family that had the most higher score on almost all the measures and on Bonacich it had a low value..

Having analyzed different centrality measures we can see that each measure has a different purpose that can be used to know the most important family depending on the context of the problem.\

REFERENCES

https://bookdown.org/markhoff/social_network_analysis/centrality.html

https://cambridge-intelligence.com/keylines-faqs-social-network-analysis/