Analyzing my Facebook Network using Gephi
The graph shown in the first page is a representation of my network of friends taken from Facebook, and rendered using Gephi. Even that it is very hard or not possible at all to get any sense of something from it, it allowed me to better understand the importance of how we lay out graphs. It is not clear at all how to sort it, color it, choose a size for its nodes or edges, or where to position each node with respect to each other. All of those are decisions that we need to make in order to find those hidden highlights in the data.

Description of the analysis work

In order to get my network of friends from Facebook, I used an application called netvizz, that can extract data from different parts of the social network:

- Your personal friend network.

- The network of likes. This is, all the pages that have been liked by your friends.

- Friendship connections or interactions within Facebook groups.

- Networks related to Facebook pages.

I chose the first one, and I got a .gdf file that I imported to Gephi. In my case this graph consisted on 1485 nodes, representing my friends (and acquaintances) and 6.154 edges, that represents all the connections in between them.

After I imported the graph, the first step I took in order to better visualize my network was changing the graph layout. I chose the “Force Atlas” layout, that, as is explained in its paper, makes nodes repulse one to each other while edges attract the nodes they connect. After letting this algorithm run for a few seconds, the graph converges to the state shown in the left figure below.

---

1 The GUESS .gdf format: [http://guess.wikispot.org/The_GUESS_gdf_format](http://guess.wikispot.org/The_GUESS_gdf_format)

The only thing that we can clearly appreciate is that there are two groups almost disconnected one from each other. It is less clear, but still noticeable, that in the biggest group there are some sub groups as well.

In order to make those sub groups more distinguishable, I used a statistics available in Gephi known as Modularity, that makes use of a method that allows us to (as its paper’s title reads) fast unfolding of communities in large networks. In the modularity statistical report it is stated that my network has 13 communities, but 7 of them have only 1 member in it. So in practical terms, there are only 6 communities, ranging from 30 to 110 members. Those 6 groups are color identified in the right image above.

So far, I have laid out my network and color coded the clusters in it. Now I can try to find other meaningful from single-nodes statistics. I will rank the size of each node based on its betweenness centrality. This is a measure equal to the number of shortest paths from all vertices to all others that pass through that node\(^3\). To get that measure, I ran statistics on graph distance, from what I learnt that the radius of my network is 11, and the average path length is almost 3. The graph ranked by betweenness centrality is shown in the next image:

---

It is hard to miss that there are two nodes that stand out as central nodes. This is explained because they act as gates between the two disconnected groups. The group in green is constituted by my fellow Fulbrighters, all of them from different part of the world, that do not know any of my other friends, except for one of them, and she is shown as that big green node. But, as we can see in the small image at the left of this paragraph, she does not know many of my friends on the northern cluster, on the contrary, she only knows one of them, an he is the other gate, the big blue node.
Another single-node question that I wanted to ask to my data, is about connectivity. Who among my friends know more of my friends? In order to do that, I ran the Average Degree statistics, and rank the nodes’ sizes by its degree, so the most connected nodes appear bigger. As an overall network statistic, the average degree is 12.6, but with a high dispersion, ranging from 0 to more than 100, as shown in the figure at the left.

The image below shows the nodes size ranked by degree. It may look similar to the graph obtained when ranking the nodes by betweenness centrality, but it is quite different. Actually, if we recall the analysis of the two biggest central nodes where there was one big node in the green community, but as we explained, it was green not because it was super connected, but because it was centrally connected. Indeed, as expectable, there are no nodes highly connected in the green group, since it is a relatively small cluster.
Finally, I would like to deeply visualize the 5 communities shown in the upper side of the graph. To do that I am just selecting all those nodes and running the statistics and the layout algorithms, following the same procedure using for the complete dataset.

This time, I used both, Force Atlas and Fruchterman-Reingold in order to position the nodes. Both are force-directed graph drawing algorithms, but with different parameters and very different outcomes.

When identifying communities, it seems to me that the Force Atlas is more helpful to visualize the graph, even if it looks more cluttered.
I must say that I did not find interesting highlights in my friends network. This is probably because I mostly saw things that I was already expecting. But also, because the dataset is naturally too personal to find generally interesting findings.

Nevertheless, this work has been very illuminating and useful in order to become more familiar with graphs and how to visualize them, and I can point out things that I found interesting regarding my own learning.

**Highlight 1**

The measure of betweenness centrality shows a measure of the whole network, since the shortest paths should visit all the nodes eventually. On the contrary, the degree or connectivity may throw highly localized results. For instance, the most connected node could be highly connected inside one relatively small group, but this group could be isolated from the rest of the network, where the nodes have a low degree.

**Highlight 2**

Five out of six of the communities distinguishable on my network are highly connected in between. My family, my jobs, my school and my academic groups are very close one to each other. All of them mostly based in Chile.

On the contrary, the network that I have acquired when living in the United States, in great part explained by the Fulbright network, is highly disconnected from my previous contacts, but with a high internal degree.

**The Gephi Tool**

This was my first time interacting with graphs and network visualizations, and I think that it was very enjoyable in large degree due to Gephi. Even that I had several problems with the JVM and Gephi when I started, and that it has several rough edges, Gephi allowed me to easily navigate through the layouts, partitioning, ranking and filtering of my network.

It was very nice to be able to hover or click on a node, and see its connections. Also, it proved to be very useful the ability to select a set of networks, and copy them to another Workspace. That served as a “hard filter”, and it helped me to perform separated analysis for the different groups.
Also, it was able to manage large datasets. I played with a dataset several times larger than my Friends’ (it was the Facebook Likes network, with more than 20,000 nodes), and I was able to do the same things that I did with the network that I ended up using (it took longer, of course).

The feature that I probably missed the most, was the ability to undo actions; it was not supported at all.