# Facebook Network Analisys

Jorge A. Aque G.

Universidad Politecnica de Yucatan

Social Network Analysis
Merida Yucatan
st1809007

**Keywords— Social Network, Facebook, social network analysis, graph, metrics.**

***Abstract***

*During the Project we will analyze the relationship between different users of the social network Facebook. It is worth mentioning that the positions and / or names of the nodes have been made anonymous for their own privacy.*

*During the analysis process it is intended to understand the limitations of the graph, it is that throughout the project the information obtained is the most accurate possible.*

*As an essential part, the content of the analysis of this graph is intended to be the main characteristics of the network, the centrality measures, Degree Distribution and Models of Networks and finally Community Detection.*

*Of course, it will be sought that during the project the explanation is as intuitive as possible to understand and it may be that during the analysis process certain parameters or metrics are not taken into account for the analysis but more than anything it is due to the response time that is can take to make calculations and for its complexity*.

## I. INTRODUCTION

As we know, being a data engineer involves a lot, really a lot of understanding of the data and you have to be prepared for any type of analysis required for your field of work.

Since the beginning of digital communication, people have looked for a way to be in contact, they have always wanted to be in constant communication regardless of the distance and as we have seen, people do not only have contact with a single person.

In the early 2000s, computing was already being part of the labor field, it was used for anything that was unthinkable at that time, but in the same way, various communication applications such as emails and networks began to be created. Social networks, one of the first social networks was Facebook, at that time it was only a "basement project" since, as we know, everything has its response time. By 2004 Facebook was officially recognized as a stable social network.

Today Facebook is one of the most immense social networks in the world, there are social networks such as Twitter, Linkedin, Instragram which have a purpose but in the end they are part of the social networks section. Only in Mexico in the year 2020 15.8 million active users were registered, which was a much lower number than the registered in 2019, of course, that number is a huge record but there is something very interesting among all this; the large amount and size of network that can be created with this social network.

With the passage of time and the advancement of technology, data analysis has taken much of the programming field and several methods have been created for data analysis, networks are a very important field of analysis for situations such as Facebook where you can see potential graphs for analysis, of course, it would be almost impossible to work with 15.8 million records and that we are only talking about the country of Mexico.

Below is the analysis of a small section of Facebook users which are represented for obvious reasons through a network. Both a general analysis and a deeper one will be shown.

I want you to understand something very important, the reason why the Facebook graph was taken was more than anything because of the ease that its use gives us, I mean that there are more networks such as Twitter or Google but the drawback is that they are very large networks that due to computational limitations simply generating the graph took hours to be able to give an image so it was chosen to use the Facebook network since the number of nodes was excellent for its practice, it was not an exorbitant number but not so much low to not have a good analysis.

## II. Network Characteristics

First of all, as mentioned before, the idea is to show the analysis of a network that comes from a social network, social networks such as Twitter and Google have a number between 80,000 and 120,000 nodes, so it is only possible to analyze with a good computer. . Because of this I chose to use Facebook data.

First of all, the most important step was to take a
Image of the graph to know what type of graph it is and see at a glance what the problems of its analysis could be, in the same way to contemplate how beautiful something as abstract as simply people connected in a social network can be seen.

As a main point it must be said that this graph is "Undirected", after knowing this I started with the general analysis, so we know the number of nodes with which we are working is 4039 which is a good number for the analysis. In the same way they follow it, it caught my attention but in the end it makes sense since a person not only has a friend as a single person, what I am talking about is the number of "edges" which I had as number 88234 edges.

As I mentioned before, being able to visualize the network helps us a lot for a quick anailisi and we can see that the Facebook network is a graph that is totally connected which led us to our next metric, the “shortest path”, for this metric the truth I did not know what to expect, the truth is, I thought it would give me a big number like 10, mostly because of the small sections at the ends but in the end as a result I got a shortest path of 4 which surprised me a lot since what it tells us is that if I want to get from person A to person B I have to spend a minimum or rather, on average, between 4 nodes to be able to complete my destination, in the same way it is worth mentioning that not only that I was impressed but also the number that I got the "diameter", as I mentioned before I thought that the number of the metrics were going to be high, in this case I thought that the diameter would be like a 20 but actually I got a diameter of 8 which as we know is the greatest distance between all the pairs points of it.

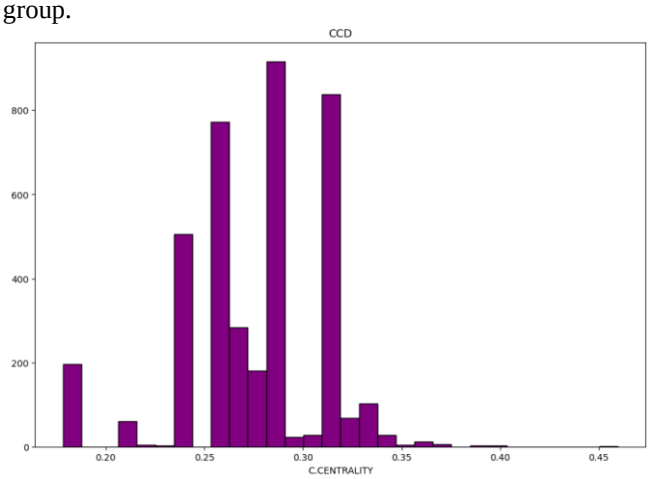
In the same way, I wanted to analyze the peripheral and the central nodes with which I was surprised, more than anything it seems something very strange but as the main node I only obtained 1 and peripheral I obtained 197 nodes as peripherals. In the same way I obtained the "cluster coefficent" as 0.6055467186200876 which if I expected a close number since as we see in our graph there are different groupings.

Finally, the density of our graph as the main idea we must know that we have to obtain a low number for this type of graphs since they are from social networks and I obtained 0.010819963503439287 as a result.

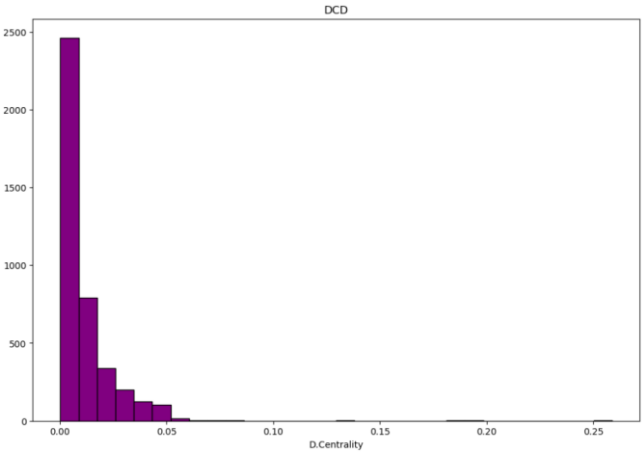
As a personal conclusion about this first section I can say that the results obtained on this general analysis were in the end good since they make sense, the only thing that caught my attention a lot was the shortest path and the diameter which if it was clear to me that there is no than leaving with the first expectation we have.

## III. Centrality Measure

During this process the Closeness centrality Distribution (CCD) and the Degree centraliry Distribution (DCD) were used, to begin with, I did not know what to expect about the behavior of these graphs but as we can see for the first graph (CCD) we can notice that there are some results with 0.40 and 0.45 that on second thought may be the nodes that help connect each separate group.

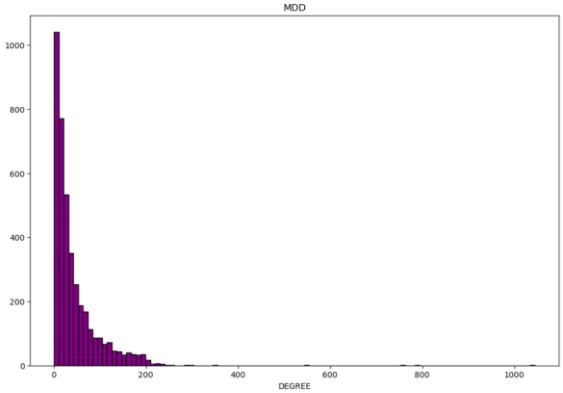



In the same way, for our second graph (DCD) we can notice its behavior and we can conclude how many nodes have a connection with another, that is, if we notice correctly there are some nodes around .13, .19 and .26 which tells us that they are nodes who do not have many connections with others



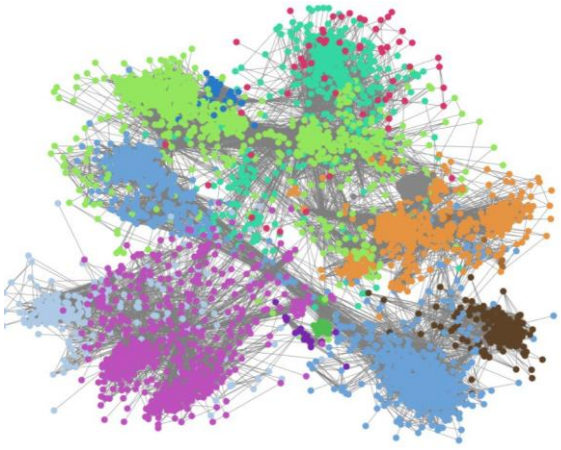

## IV. Degree Distribution and model of networks

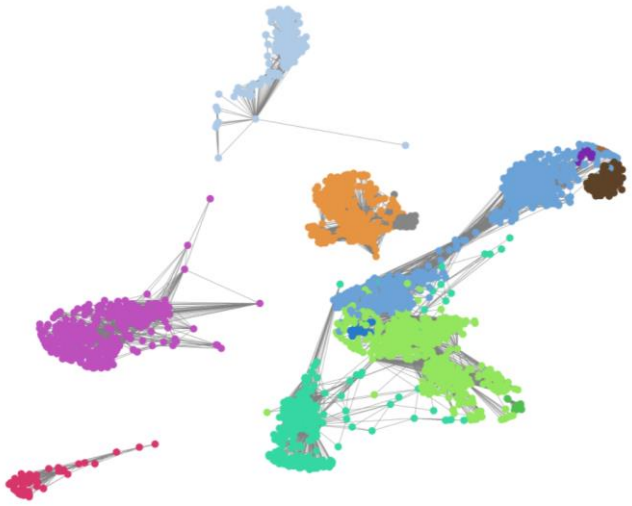
For this section I was very surprised by the results since at first glance we can see that there are a large number of nodes with a considerably normal degree but if we approach well some nodes in the range of 300 - 1000 have a degree that is high for this type. of situations, this means that there are some nodes that serve as a center and normally these nodes are the ones that have the most friends in social networks, that is, if there are nodes with a high degree it means that it is a center and that it has a connection with a large number of nodes, in the same way I wanted to give it a more exact touch in terms of the bins to be able to have something a little more exact for my criteria.




## V. Community Detection

During this last step I was not sure how to do it since in my opinion it is one of the most complex due to accuracy, but more than anything they were a problem of my own knowledge, at the end of everything during this process I really liked the result obtaining as a number of 13 communities and having as a visual result this graph where we can see the 13 communities of different tones, it is worth mentioning that it seems that there are less than 13 but this is due to similar tones such as blue, the truth when seeing the first graph I thought that There would be about 10 communities because we can go with the idea of this number because of the different groups that we can see, in the end I got a closer number.





## VI. Conclusion

During this Project I noticed that working with graph is easy and at the same time has its complications, I do not mean that it is easy to manipulate them, it was all thanks to Netwokx and it is worth mentioning that from this library I learned more than I already knew and different ways to manipulate a network, honestly it is very difficult to work from scratch with networks but in the end comes satisfaction.

In the project I was able to reinforce certain knowledge and in the same way I was able to realize there are many implications for the analysis, we should not leave with the first result we obtain.

Regarding the implications that I had, the first thing was the size of the network, it was a network with an acceptable number of nodes for my computer and honestly if it took a few minutes to do certain calculations, I also want to talk about the community network, in this process if I have to reinforce various knowledge, at first I obtained a practically useless network since it did not tell me anything, I had to support myself from previous tasks and some videos to be able to have an acceptable result and honestly the result I obtained was good although it could be Better, I could have made the communities more colorful, more separated in groups to have a better visualization, after all, the communities can be noticed.

I liked this project a lot because, as I already mentioned, I was able to reinforce knowledge and in the same way it felt satisfactory to obtain acceptable results, I was able to learn more about how to improve the visualization of a graph and its metrics.

## VII. References


[1] Leskovec, J., 2021. SNAP: Network datasets: Social circles. [online] Snap.stanford.edu. Available at: [Accessed 23 June 2021]

[2] "Reading and writing graphs — NetworkX 2.5 documentation", Networkx.org, 2021. [Online]. Available: https://networkx.org/documentation/stable/reference/readwrite/index.html. [Accessed: 25- Jun- 2021].

[3] I. Despot, "Understanding Community Detection Algorithms with Python NetworkX", Memgraph.com, 2021. [Online]. Available: https://memgraph.com/blog/community_detection-algorithms_with_python_networkx. [Accessed: 25- Jun- 2021].

[4] "networkx.algorithms.community.quality.modularity — NetworkX 2.5 documentation", Networkx.org, 2021. [Online]. Available: https://networkx.org/documentation/stable/reference/algorithms/generated/networkx.algorithms.community.quality.modularity.html. [Accessed: 25- Jun- 2021].

[5] Klein, Douglas. (2010). Centrality measure in graphs. Journal of Mathematical Chemistry. 47. 1209-1223. 10.1007/s10910-009-9635-0.