## Introduction and Context of the Network

The purpose of this document is to report the process of creating a network of friends using the web crawling method. It also reports the results obtained with respect to the characteristics of the network and the algorithms implemented for link prediction.

The network corresponds to a network of friends on Facebook. The information was obtained from the author of this paper and considers the friends in common of the direct contacts.

## Mapping Process

The Web Crawling technique was used to map the network. A web crawler (also known as a web spider or web robot) is a program or automated script which browses the World Wide Web in a methodical, automated manner. [1]

To this end, much of the code from previous tasks was recycled to facilitate the process.

The main piece of code is a class named *facebook_friends*, composed of the following methods:

- *login*: Use the *chromedriver* executable to login to Facebook. Requires username and password to work.
- *go_profile*: uses *chromedriver* to go to the profile of the person logged in and enter the *friends* page.
- *scroll_down*: navigates to the *friends* page of the person who is logged in. It stops until it reaches the bottom of the page.
- *get_my_friends_list*: uses the browser's *inspect element* function to get all the user's friends. Returns the list of friends.
- *get_mutual_friends*: for each of the user's friends, generates a list of friends in common. Returns a list with friends in common for each friend.
- *create_dict_friends*: Generates a dictionary whose elements are composed of the user's friends (keys) and the friends in common with the user (values).

Because the dictionary generated users with my mutual friends but not me, "Alan Peraza" was added to the friends list of all my contacts. The *itertools* module was used to generate all pairs of friends.

The dictionary was moved to a list with the name `edgelist`.

## Basic Characteristics and Visualization

With the list of edges, a network could be generated using the networkx module, which is shown below.

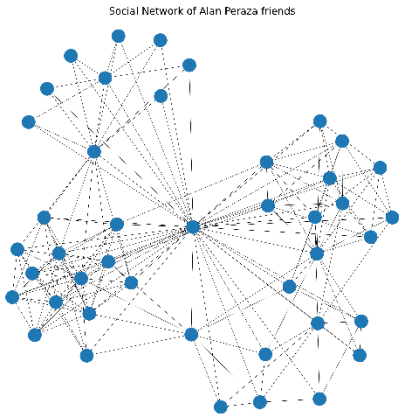


*Figure 1. Network generated by "edgelist" list*

Using the library, it was possible to generate some metrics of interest, which can be seen in the following table:

| Metric | Value |
|---|---|
| **Nodes** | 43 |
| **Edges** | 179 |
| **Average Degree <k>** | 8.33 = 8 |
| **Average Path Length** | 1.8 |
| **Density** | 0.2 |

It is important to consider that I have 72 friends on Facebook, but only 42 of them are shown in the network. This is because some of them have private settings regarding who can see their friends. The script is designed so that when one of these cases is encountered, it ignores the contact.

The average degree of 8 tells us that on average (in my network of friends) each friend has contact with 8 people I know (friends in common). The average path length is, as expected, short, since we are talking about a really small network.

It has a density of 0.2. Several authors mention that real networks are sparse [2]. Given the nature of the network and how it was created, the value obtained does not represent a surprise.

The degree distribution shows that the network is a small and indistinguishable from any known distribution (such as Poisson or Normal).

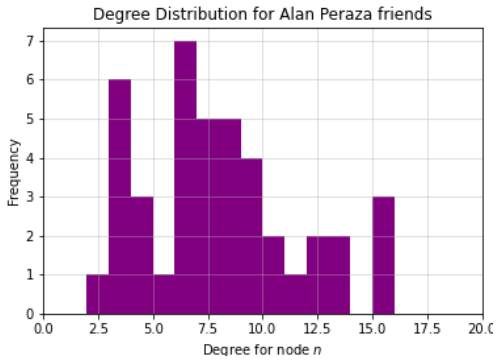


*Figure 2. Degree distribution for Alan Peraza's network of friends*

## Conclusions and Future Work

Web crawling methods proved to be effective in obtaining the data necessary for the creation of the network. Although some impediments (software, rules and Facebook's own policies) represented certain challenges, I consider that the work achieved correctly represents the reality of my circle of friends.

The original idea was to use an API authorized by Facebook, but some setbacks made its use complicated. It would be interesting to learn about this and more social network contact network structures with certified APIs.

# References

[1] “Web crawler,” Science Daily.

[2] A.-L. Barabasi, Network Science. Cambridge, England: Cambridge University Press, 2016.