An analysis of the Instagram network

Mattia Manna, 1886817
Paolo Zilviano, 1916518

SAPIENZA
UNIVERSITÀ DI ROMA

Contents

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Why social media? (1)

- Nowdays social media is a cross-section of our community; we all use it, to interact with other people or using the content offered.
- The great importance that social media have gained in people's lives, think of some tweets that influenced the stock market, has meant that over the years they became the subject of studies.

Why the social media? (2)

As the literature says social network are:

- very sparse
- scale free
- there is an ultra small world effect
- there are hubs.

Analyzing social media we investigate how a network of people is organized:

- how many people they usually follow
- with how many people they interact

Instagram was chosen due to the way it is constructed, because allows to gather info on users that aren't directed friend.

Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Research questions

1. Does the ultra small world effect exist?
2. Is Instagram a scale free network?
3. Is it sparse network?
4. Are there hubs?
5. Is the dunbar number respected?

Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

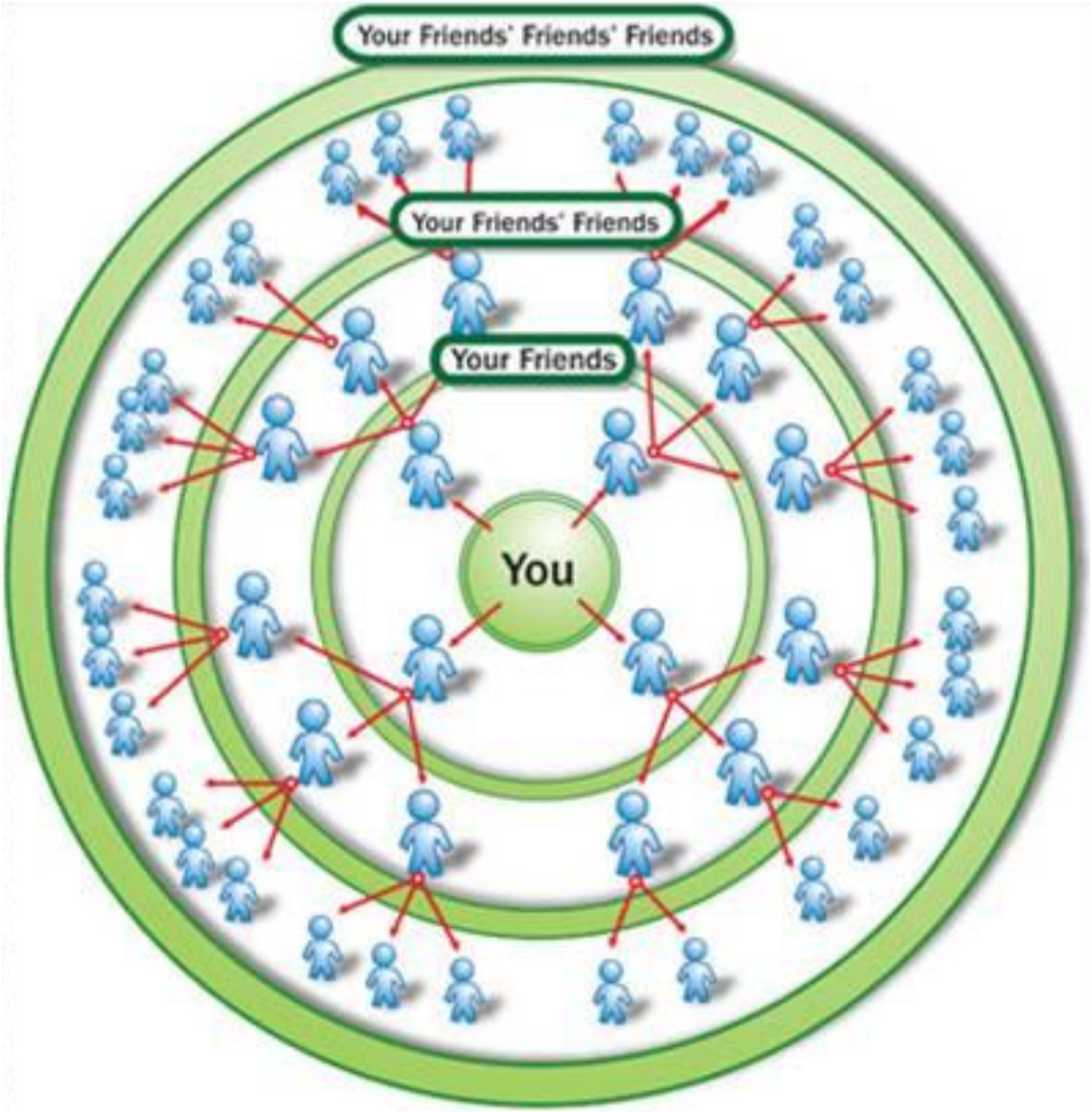
Data collection (1)

To gather informations about the instagram network we used web scraping.
We used two chrome's extensions to extract info about instagram accounts (IG Follower Export Tool, IG Exporter).

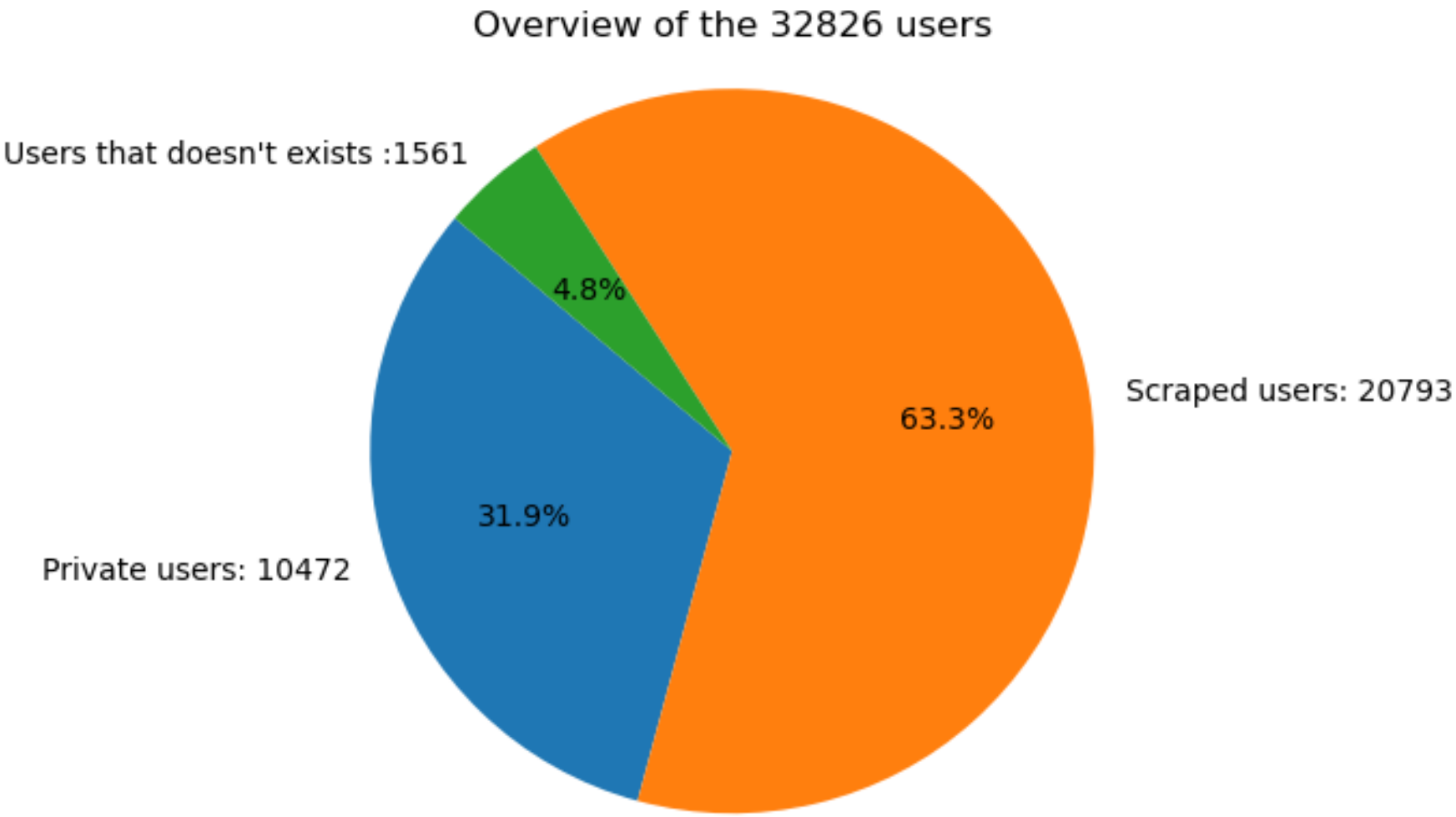
What we did was to try to take a portion of this network.
To build a subgraph that made sense we extract the informations about:

1. Tier1: my followings
2. Tier2: following of my followings
3. Tier3: following of the followings of my followings

Data collection (2)



Data collection (3)

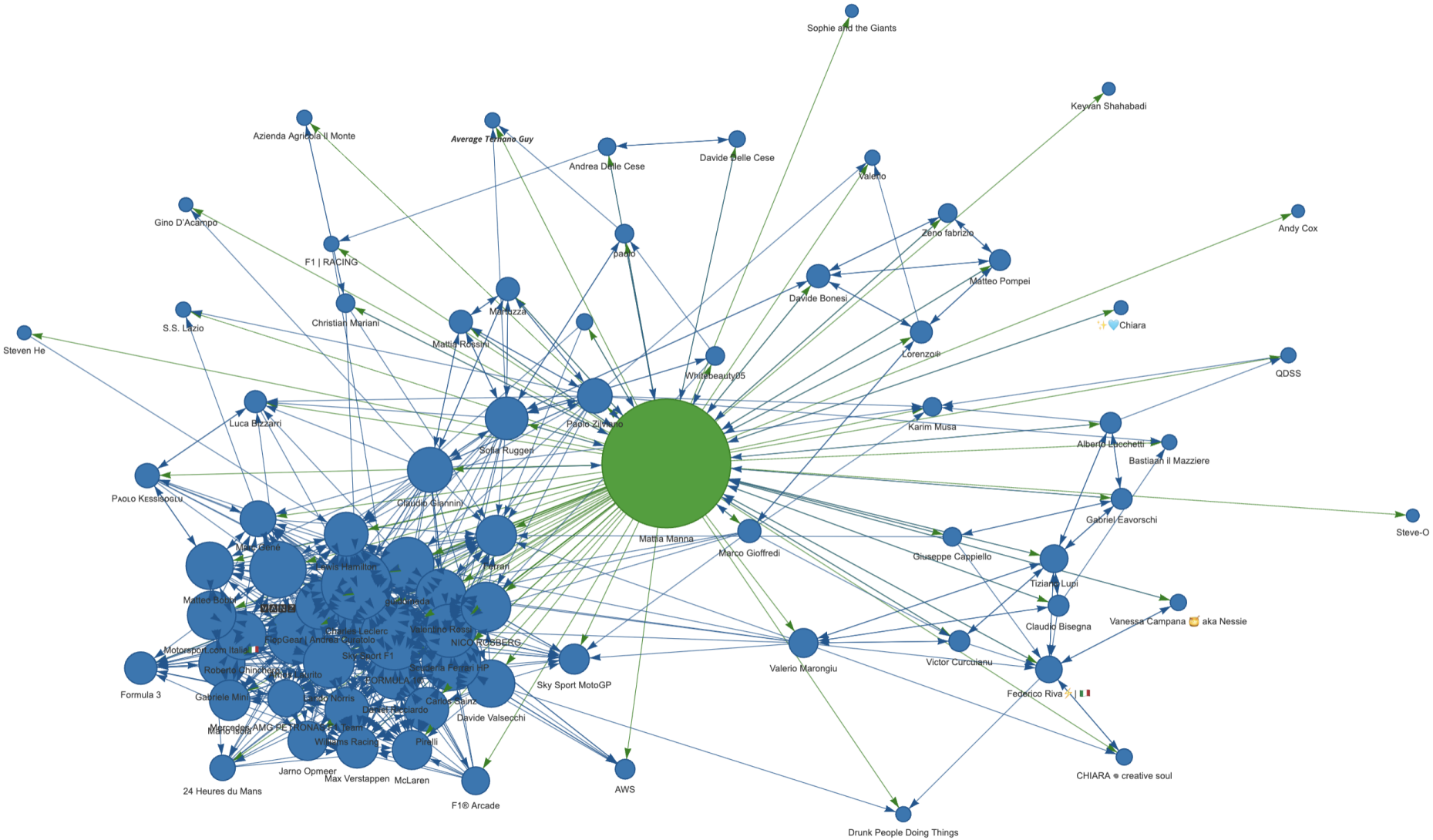


Data preprocessing and network construction

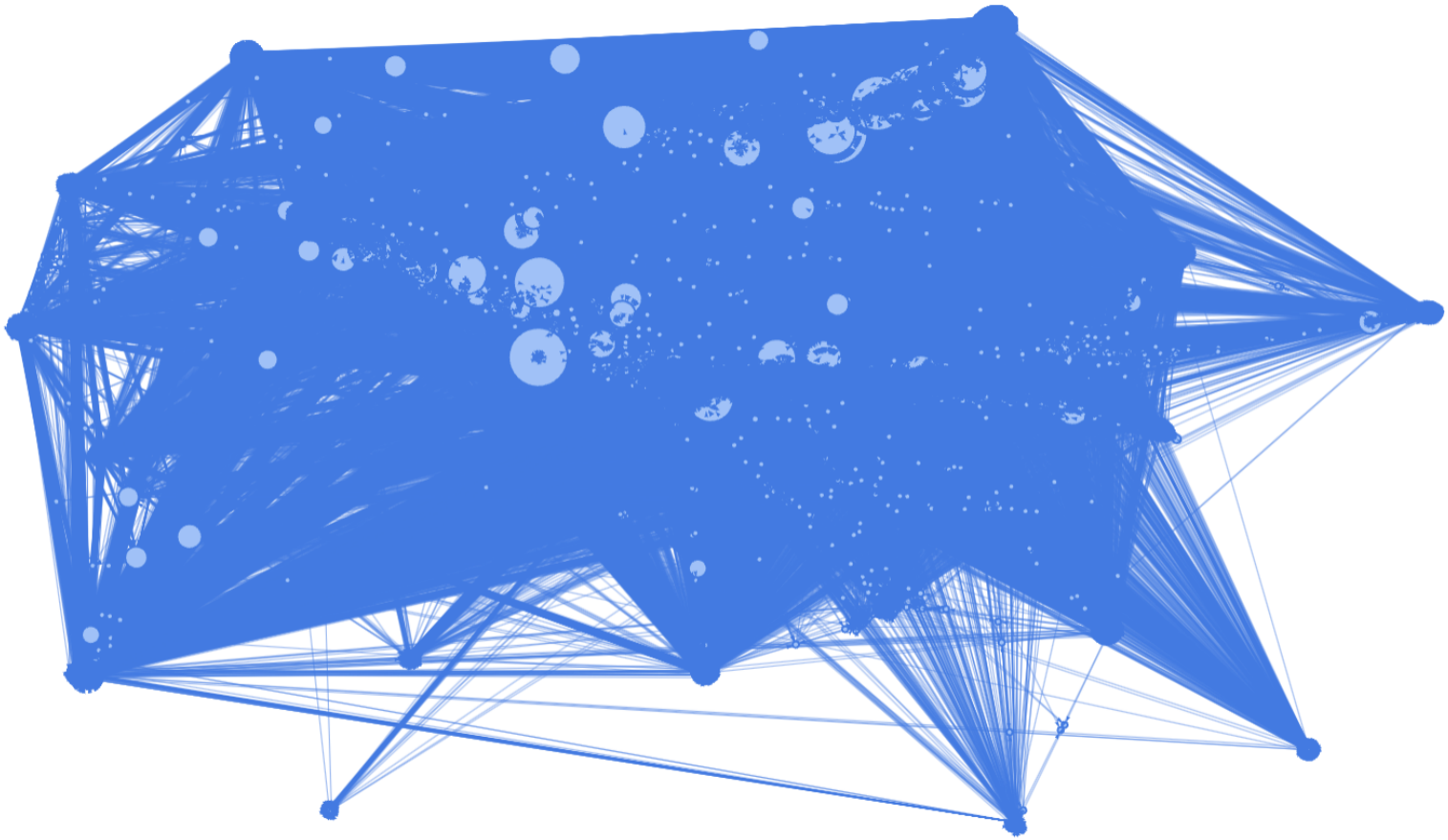
1. We collected every csv file given by the exporting tools
2. Then through a python script we created 2 datasets, one for the nodes (in order to have a list of the users) and one for the directed links (to register who follow who).
3. After we cleaned those two dataset removing useless links and nodes, for instance duplicates and nodes that are outside the 3rd tier.
4. Then importing those two datasets in R we build the subgraph of instagram.

In the end we build a graph that is made of 19992 nodes and 1094224 links.

Partial graph



Resulting graph



Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Ultra small world effect (1)

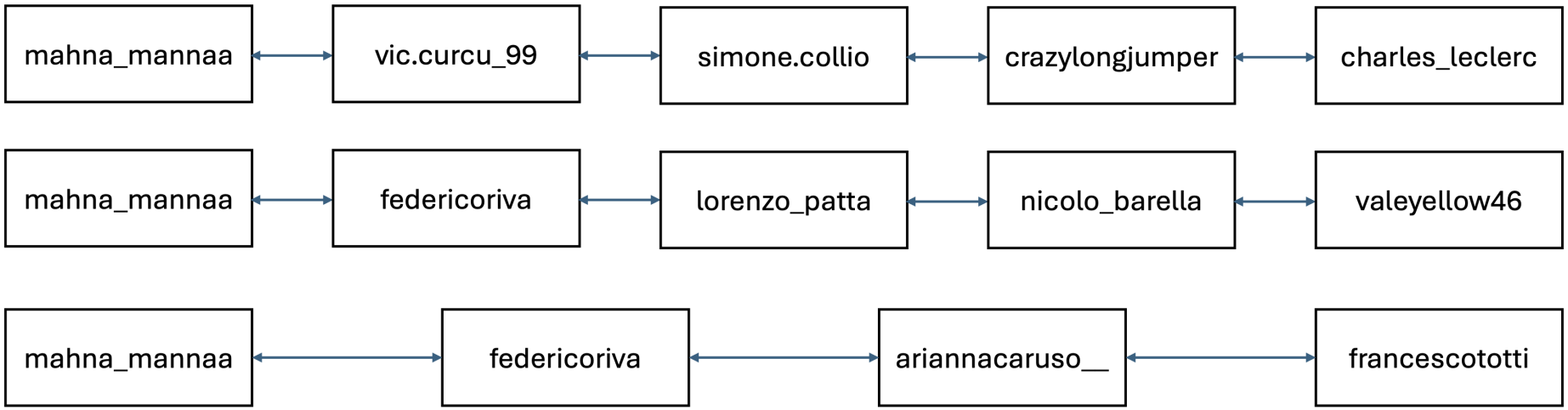
- In order to find out if there is the ultra small world effect in our network we look at the average path length
- The average path length is the average number of links in the shortest path between any two nodes in the graph

```
```{r}
avg_path_length <- mean_distance(graph1_directed,directed = T)
avg_path_length
```
```

```
[1] 3.753972
```

We found out that the average path is of 3.75, so we are very close to have the ultra small world effect.

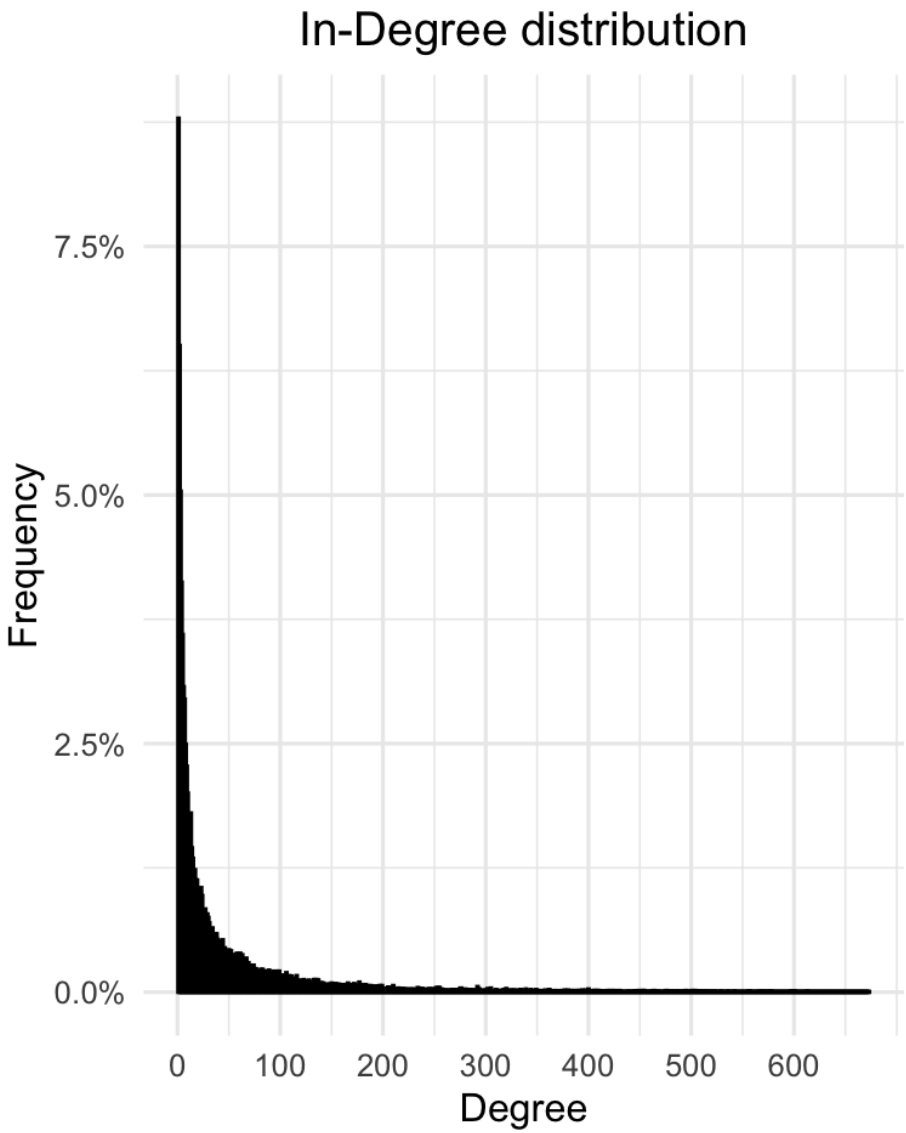
Ultra small world effect (2)

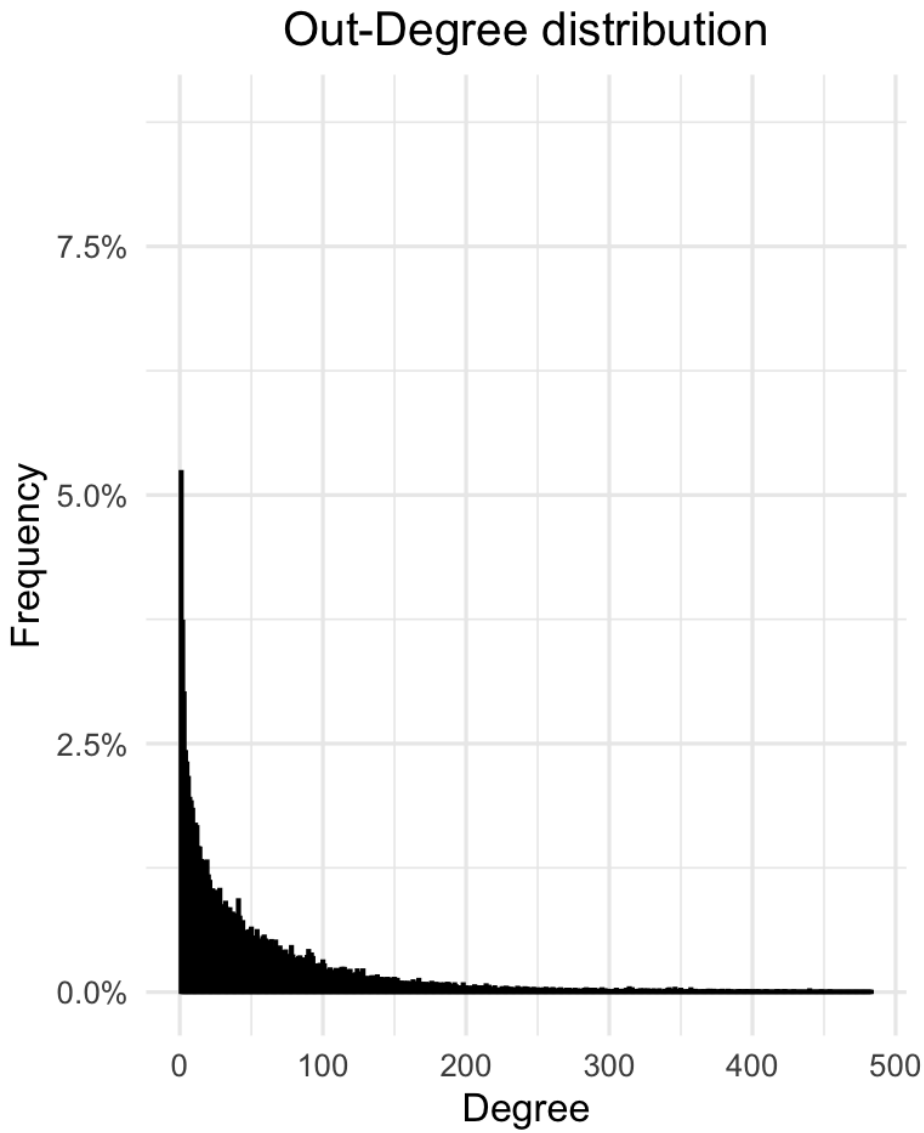


Outline

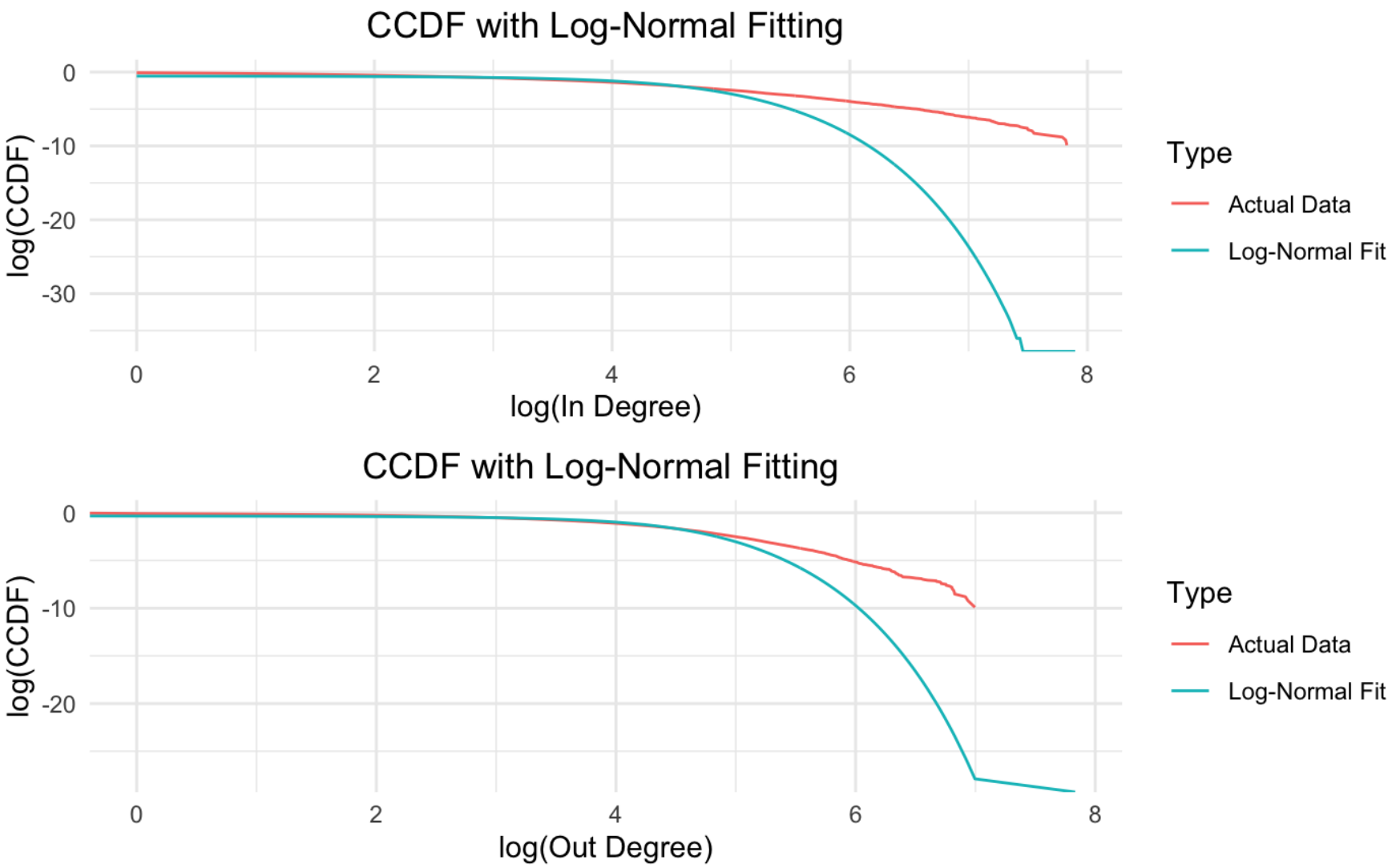
1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Degree distribution (1)





Degree distribution (2)



Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Hubs (1)

- Even though we don't have a scale-free network, we can still find hubs

- Selecting the top 5% of the accounts with the highest incoming links, we found out that those represent the 41% of the incoming links

Hubs (2)

User	InDegree	User	InDegree	User	InDegree
f1	2689	natgeo	1797	fernandoalo_oficial	1401
charles_leclerc	2508	danielricciardo	1733	kendalljenner	1384
lewishamilton	2482	landonorris	1688	redbullracing	1371
cristiano	2412	nasa	1678	ferrari	1367
chiaraferragni	2102	carlossainz55	1645	433	1349
scuderiaferrari	1900	willsmith	1549	motogp	1346
instagram	1886	dualipa	1531	iamzlatanibrahimovic	1328
valeyellow46	1868	badgalriri	1493	mercedesamgf1	1325
leomessi	1811	champagnepapi	1490	mclaren	1321
maxverstappen1	1800	kingjames	1407	belenrodriguezreal	1312

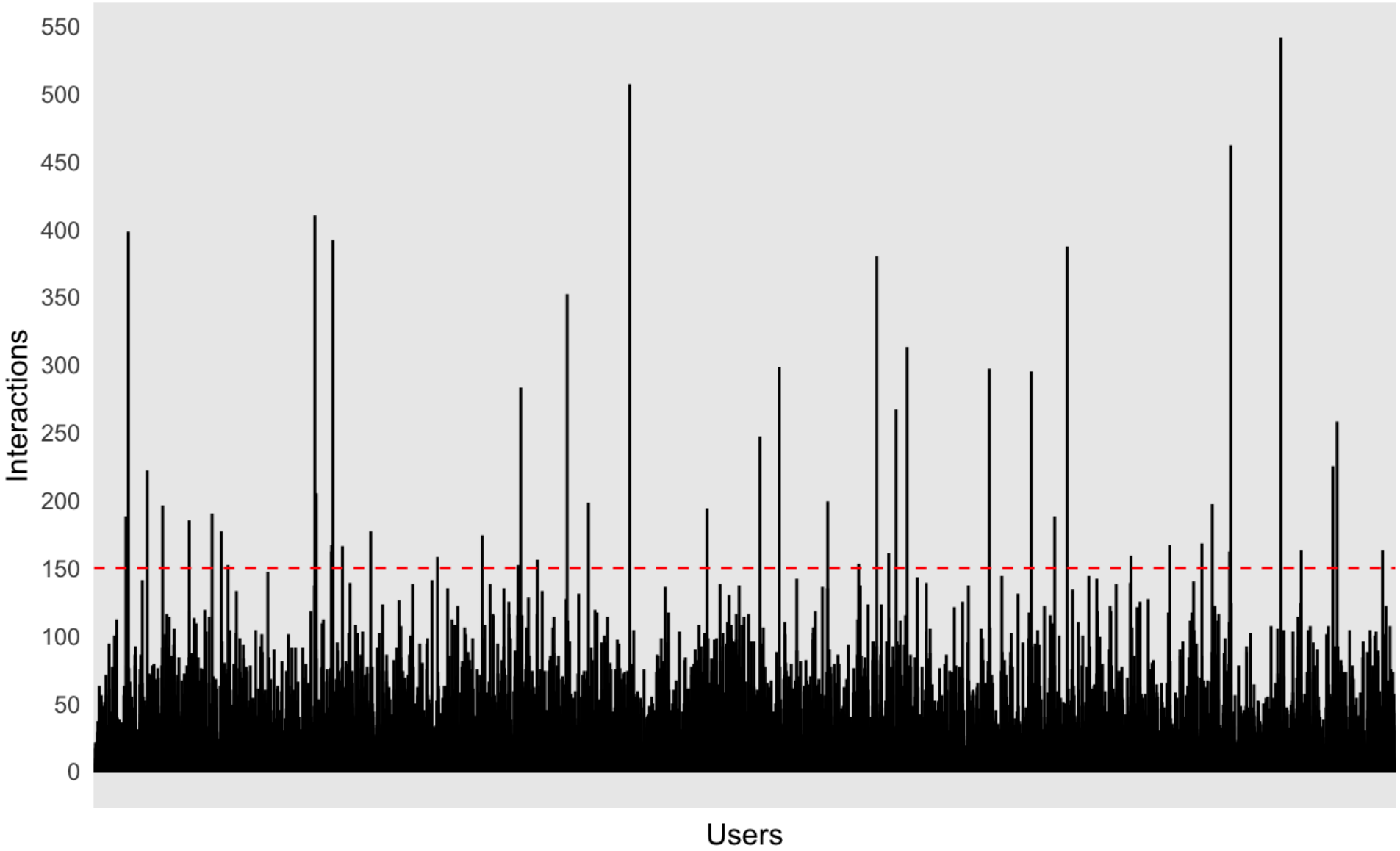
Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Dunbar number (1)

- The dunbar number states that people do not interact with more than 150 humans in the every day life.
- In our work we considered as interactions when two users follow each other.

Dunbar number (2)



Outline

1. Introduction
2. Research questions
3. Data collection and preprocessing
4. Ultra small world effect
5. Degree distribution
6. Hubs
7. Dunbar number
8. Conclusions

Other graph's information

```
[1] "Density: 0.00273788671088464"
[1] "Is the graph connected?: TRUE"
[1] "Average degree <k>: 109.466186474602"
[1] "Average path lenght: 3.75397179553626"
[1] "Diameter: 10"
```

Final notes

- **Incomplete Data:** The final graph that was used for the analysis is missing a small part of the whole collective, despite we scraped for 4 months. This of course is an issue, but we didn't have the resources to scrape 24/7. Maybe without this problem the analysis would have been more reliable.
- **Interactions During Collection:** It wasn't possible to build the graph in a freezed timespan. So possible interactions could have happen during the scraping but, again, we didn't have the tool to operate this way. However the instagram network has been tangible for almost a decade, so its structure and the one of its subgraphs shouldn't have changed too much in few months.
- For these reasons, the results of our analysis will no longer be related to the entire instagram network but to its subgraph

Conclusions on the instagram subgraph

1. Does the ultra small world effect exist? Yes
2. Is Instagram subgraph a scale free network? No
3. Is it a sparse network? Yes
4. Are there hubs? Yes
5. Is the dunbar number respected? Yes

Thank you for your attention!