

# CAN SOCIAL MEDIA MAKE YOU A BETTER DEVELOPER?

Matteo Biglioli

Social Network Analysis

# 938199

✉ bigliolimatteo@gmail.com

GitHub bigliolimatteo



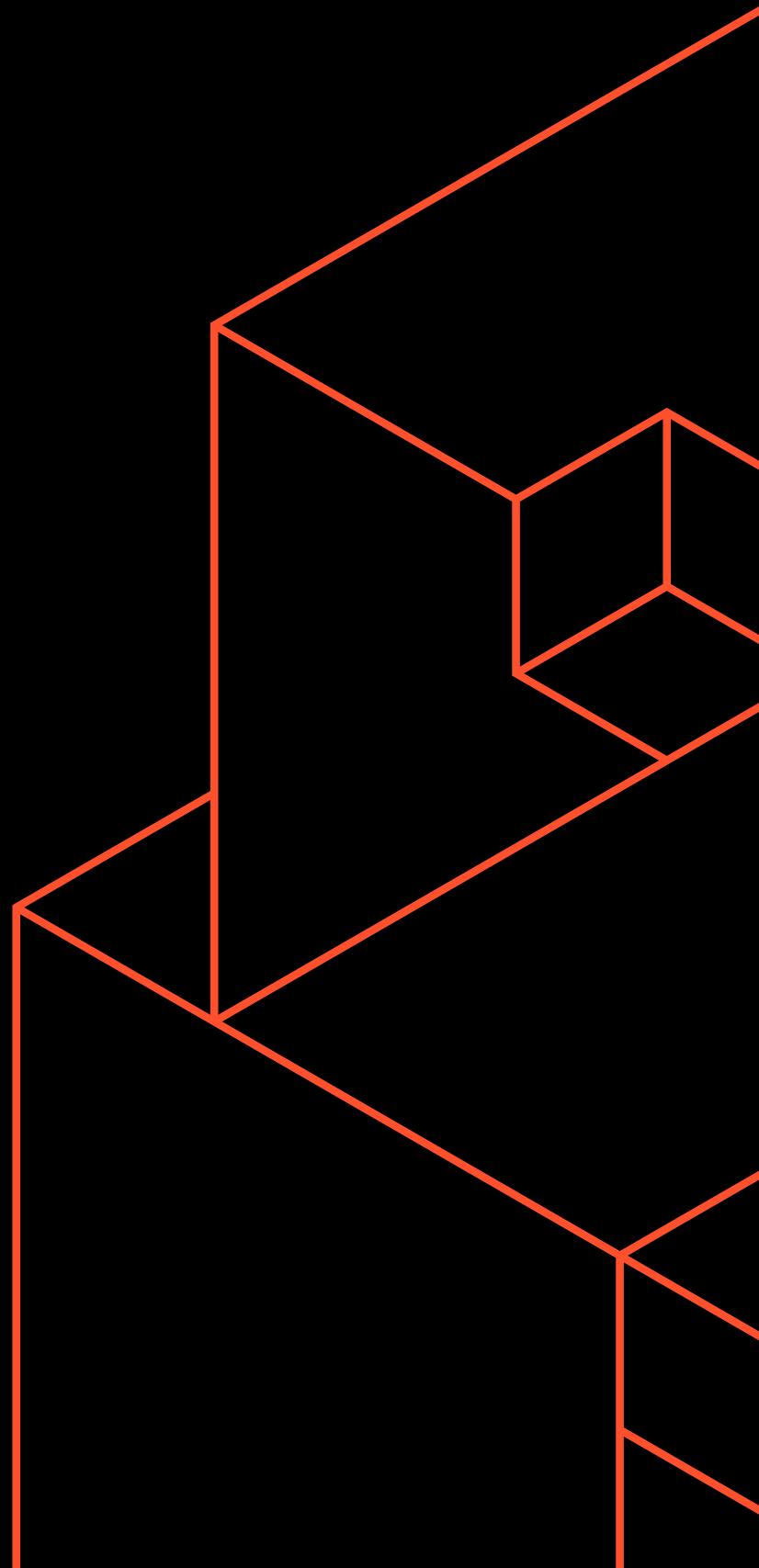
# INTRODUCTION

Github is one of the most important platform for what it concerns programming and programmers, frequented by more than 73 million developers that worked on over 200 million repositories.

Asides from development-related features, it offers a comprehensive social experience by implementing different logics typical of a modern social network.

In this project we will analyse a community of developers linked to a specific node, my personal GitHub account.

The data in this project were collected via the [Github API](#).



# DISCLAIMER

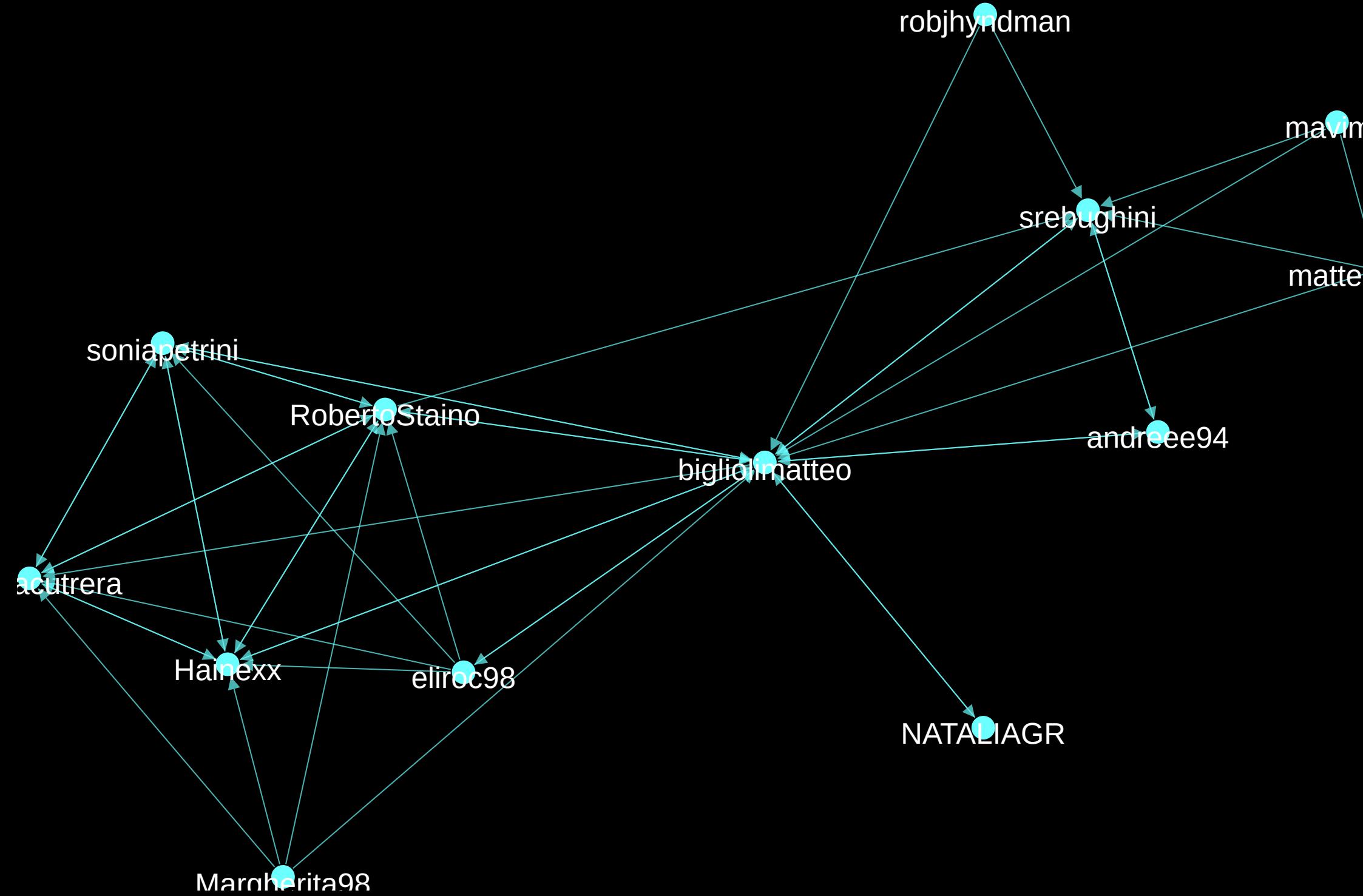


Due both to rate limit and dimensionality issues,  
we were not able to perform an analysis on the  
whole network of Github.

We generated three different surrounding  
networks including nodes which have a distance  
from our center less than 3\*.

\*This number, selected through different experiments, ensure that our network remains in the 10 thousand nodes limit.

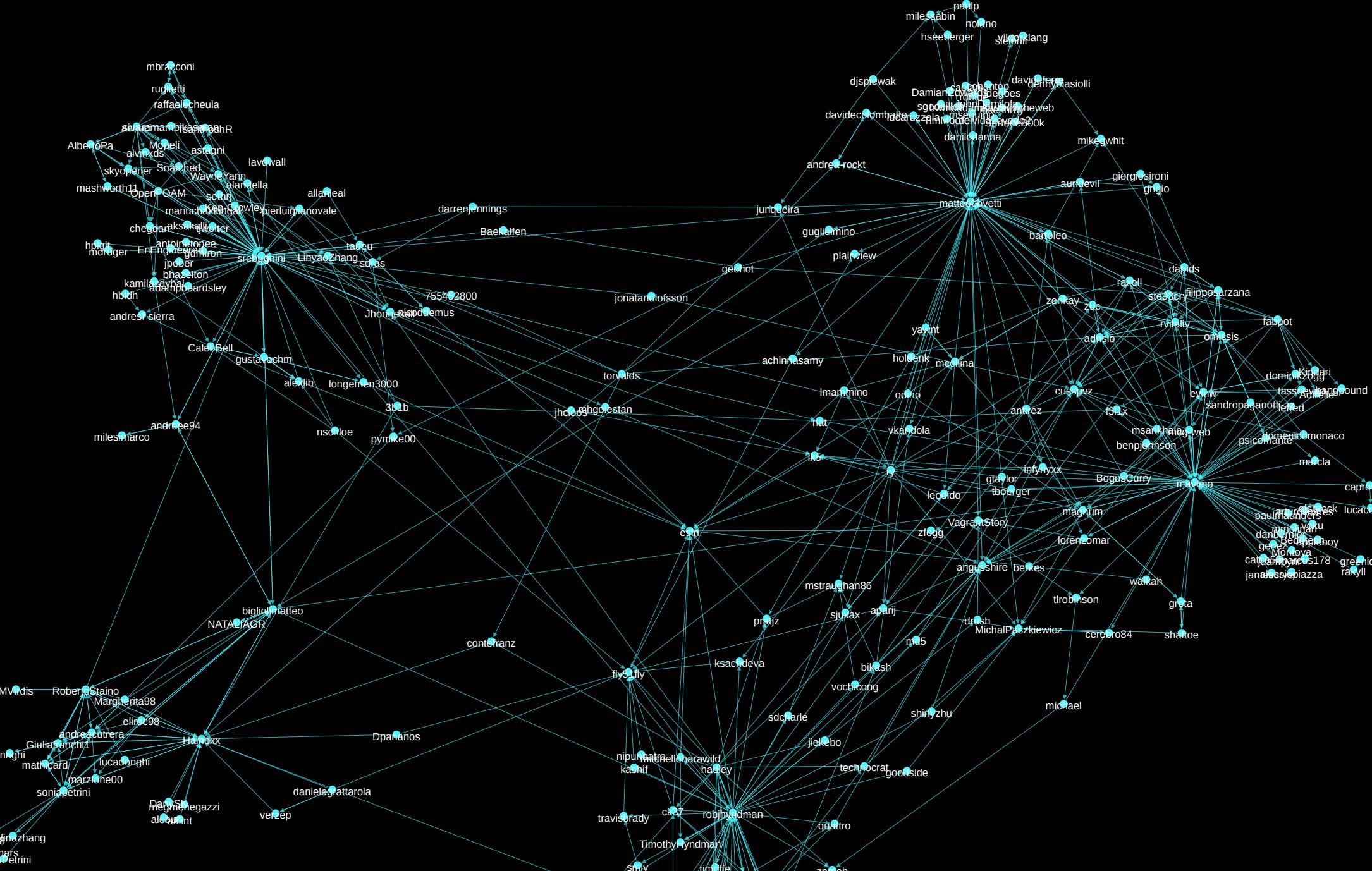
# MY NEIGHBORHOOD



- Max distance: 1
- Number of nodes: 13
- Number of edges: 45
- Average in/out-degree: 3.462
- Degree distribution:
  - Mean: 6.923
  - Median: 6.0
  - Std: 4.632
  - Max: 19
  - Min: 2

Density: 0.288

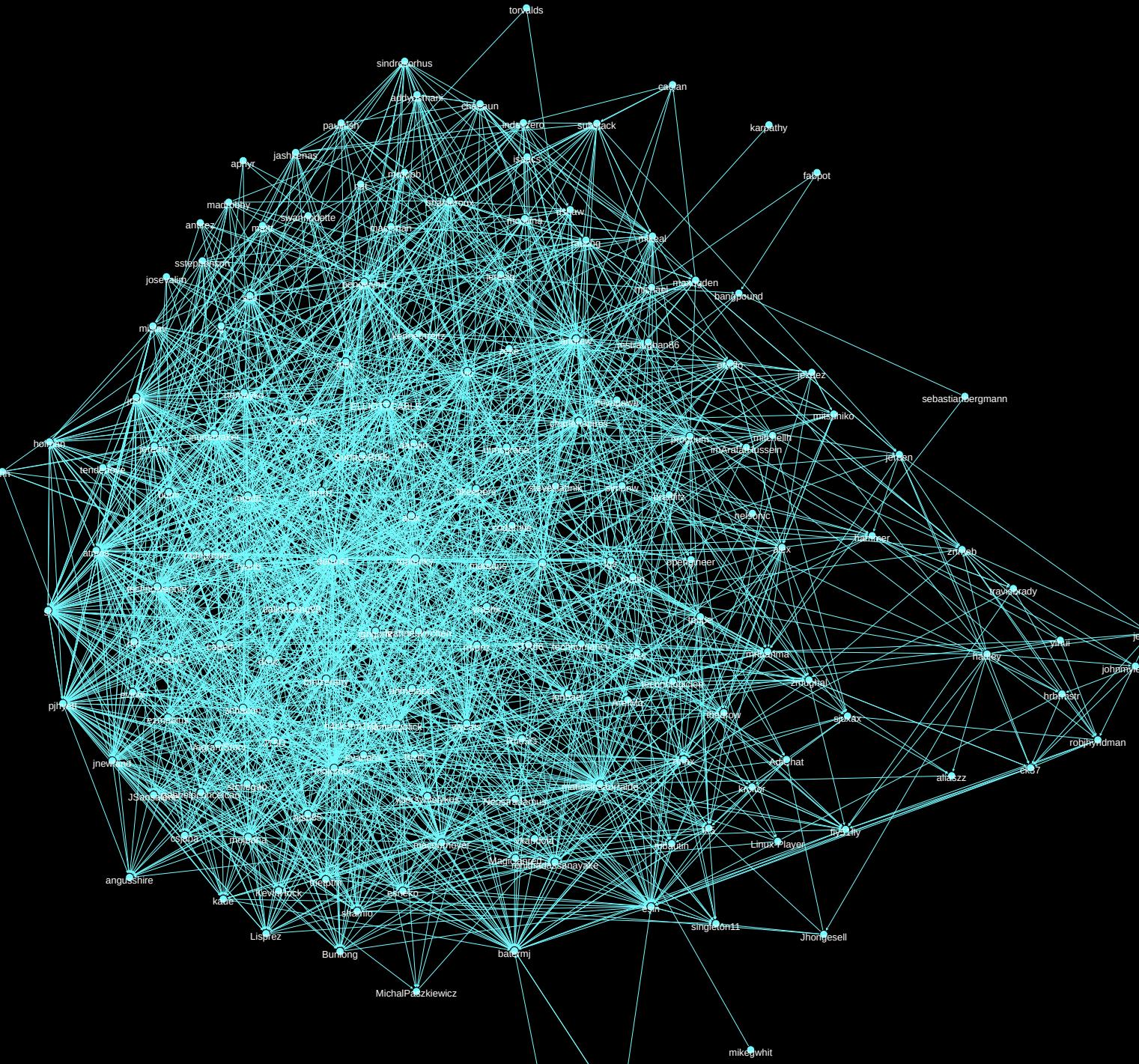
# MY NEIGHBORHOOD



- Max distance: 2
- Number of nodes: 239
- Number of edges: 656
- Average in/out-degree: 2.745
- Degree distribution:
  - Mean: 5.489
  - Median: 3.0
  - Std: 8.964
  - Max: 74
  - Min: 1

Density: 0.012

# MY NEIGHBORHOOD

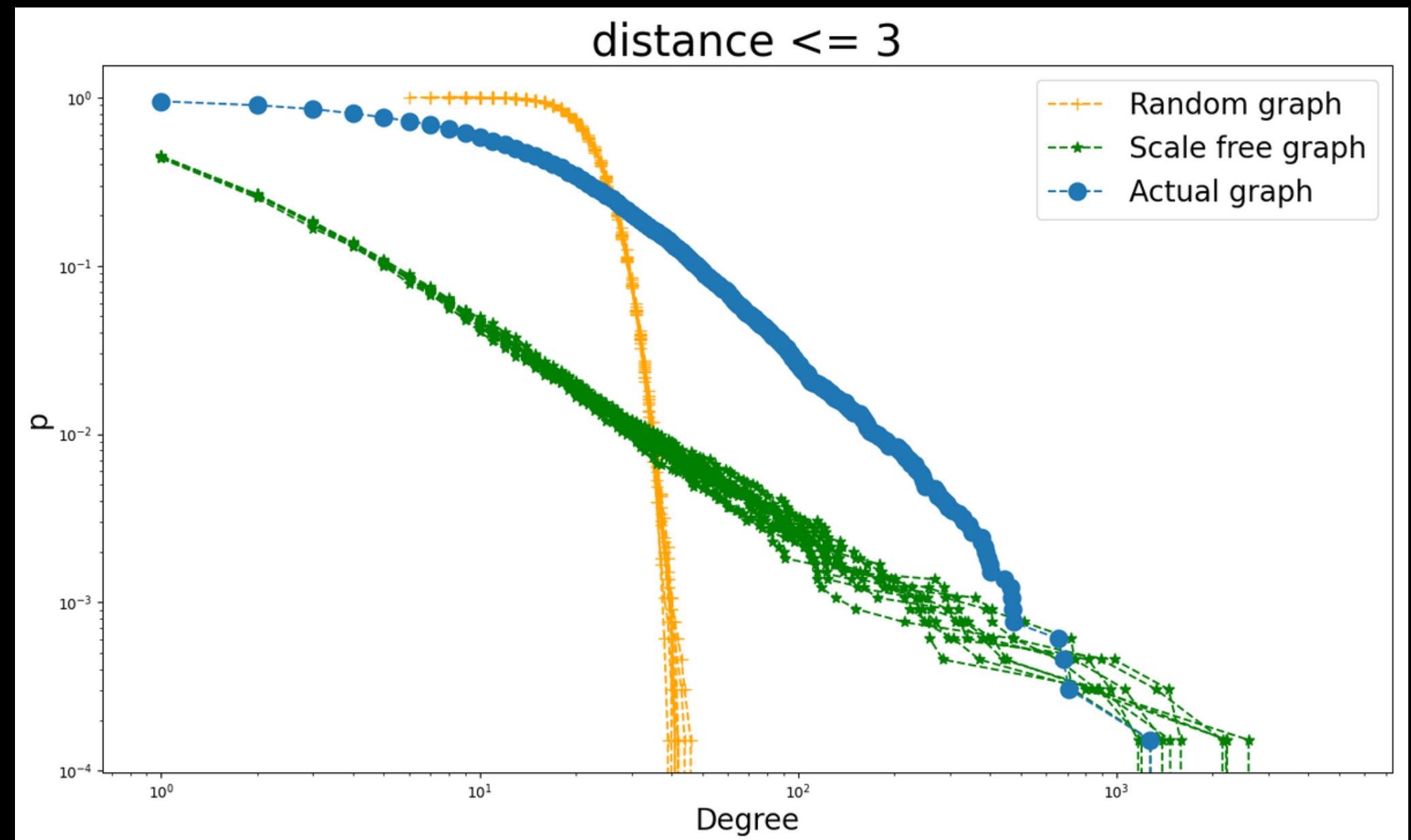
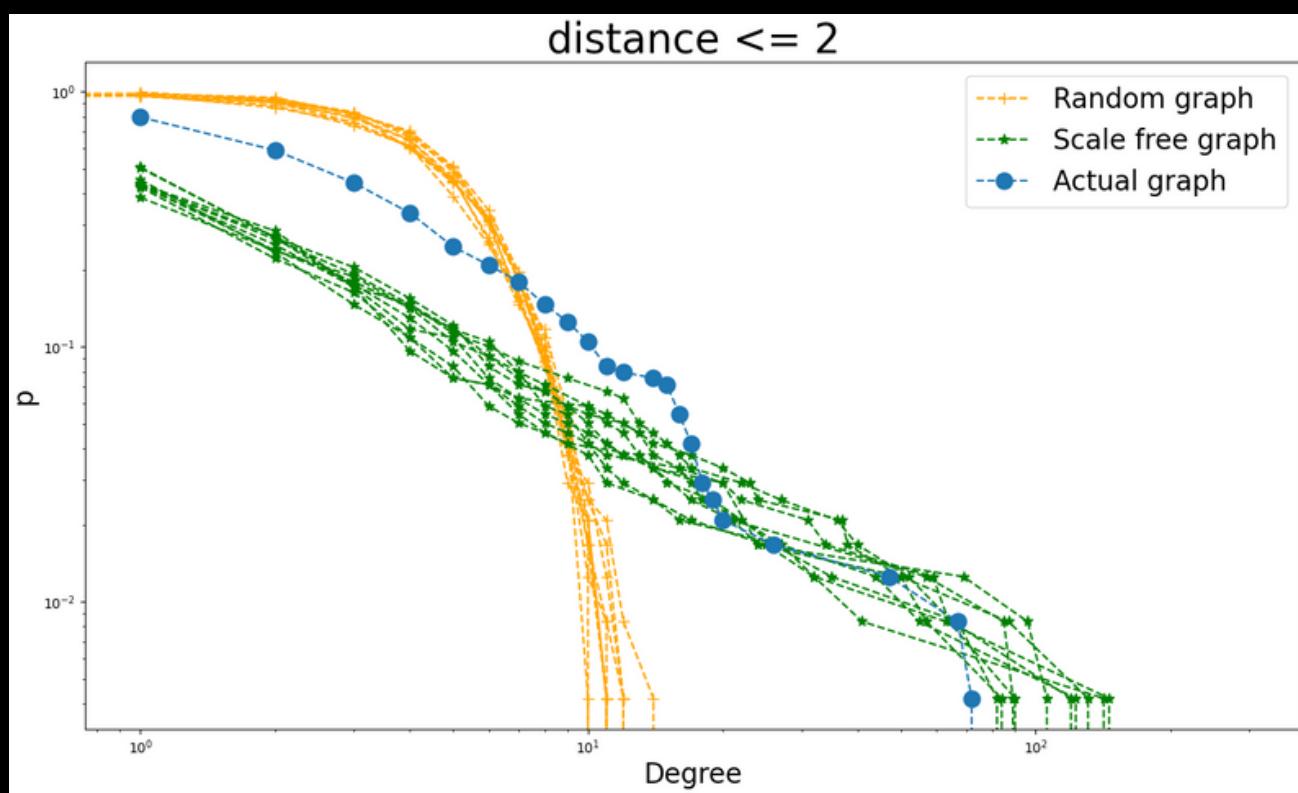
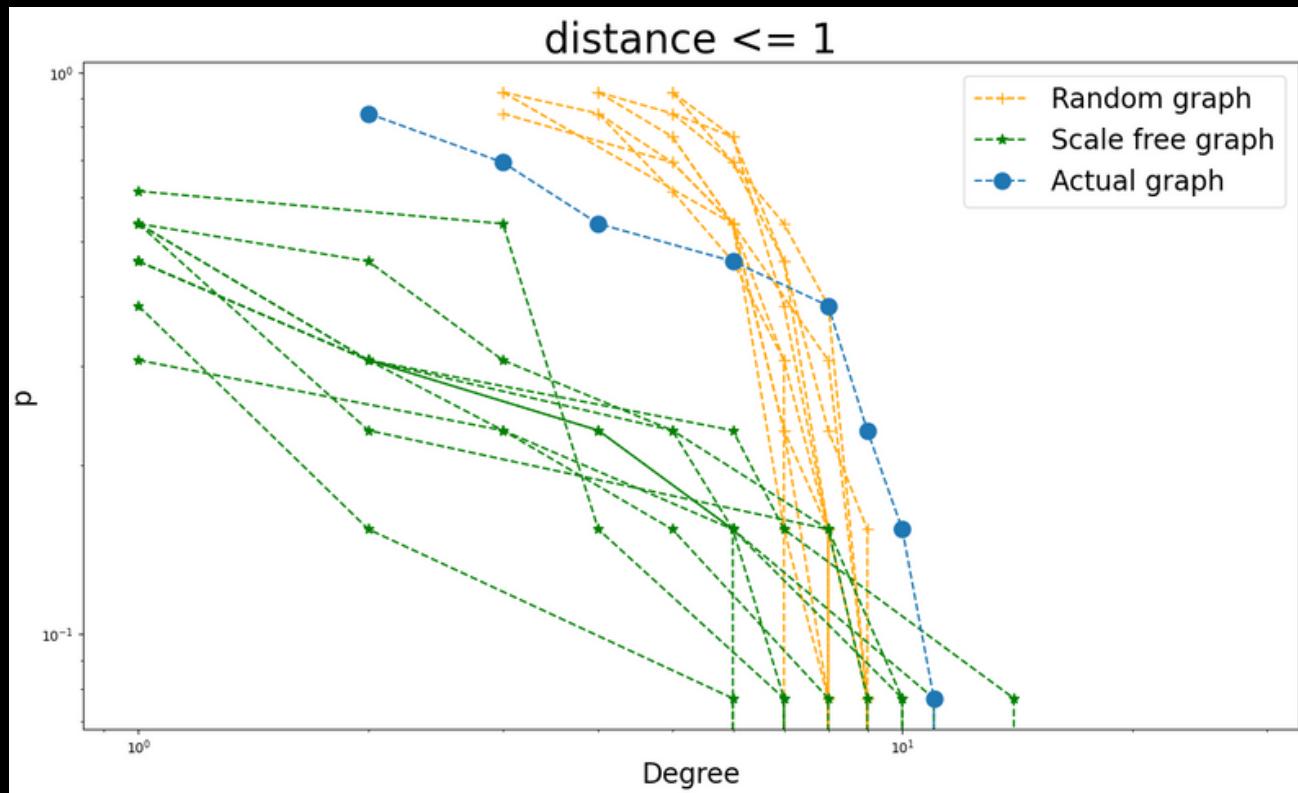


Only showing nodes with degree > 100

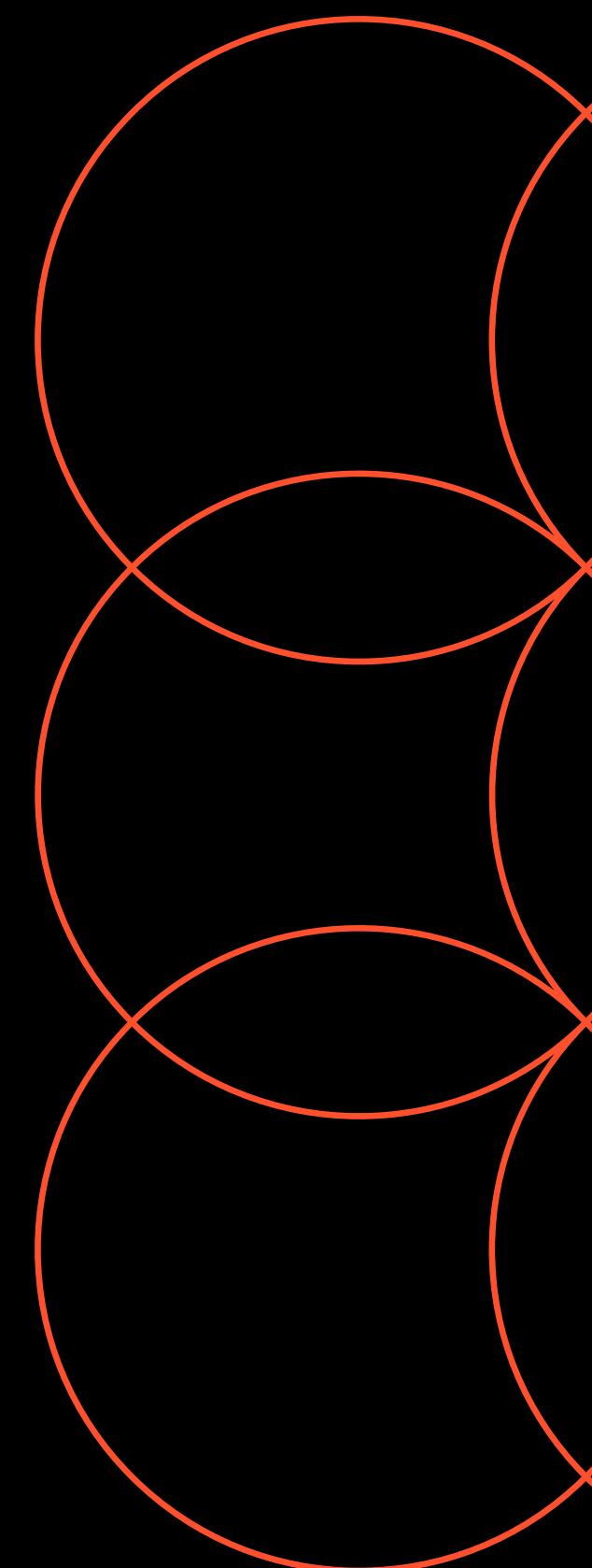
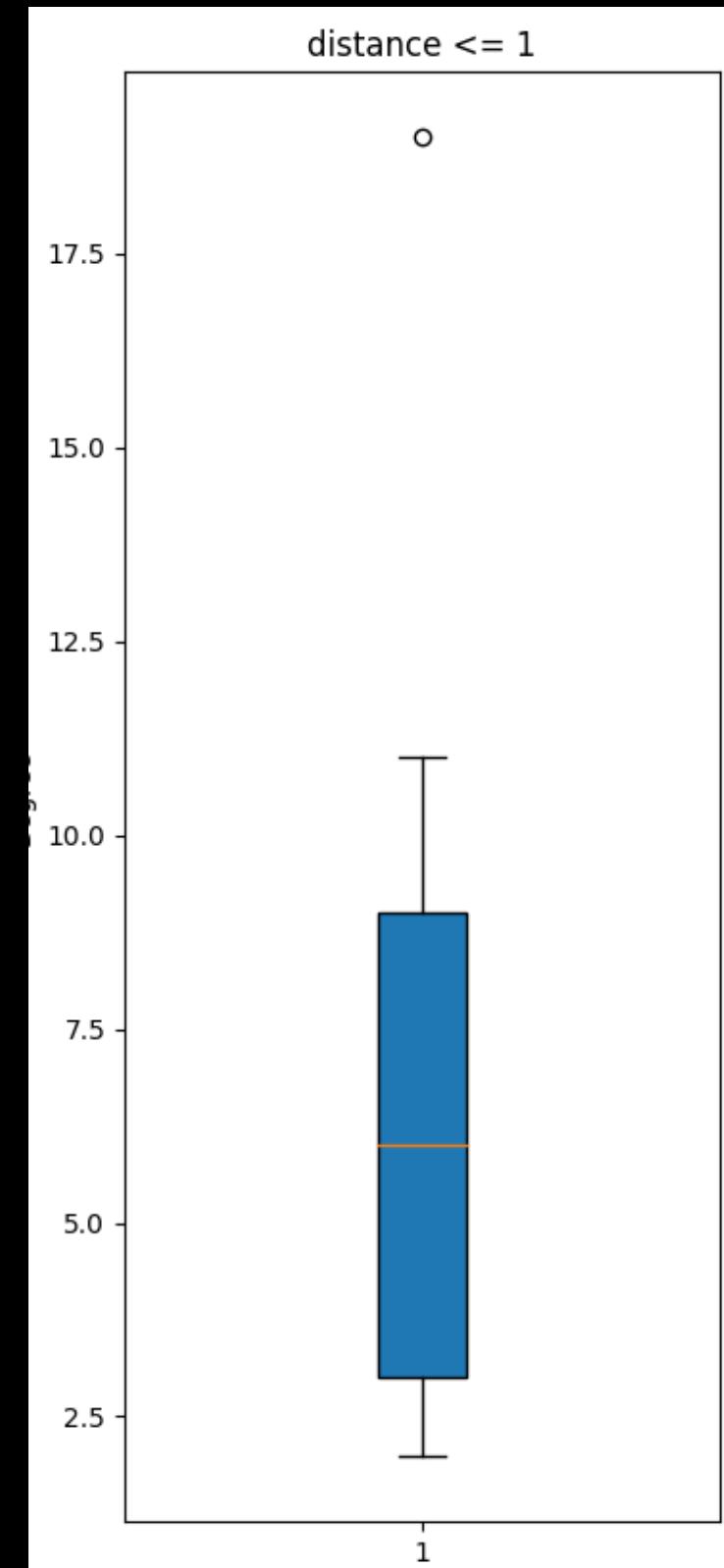
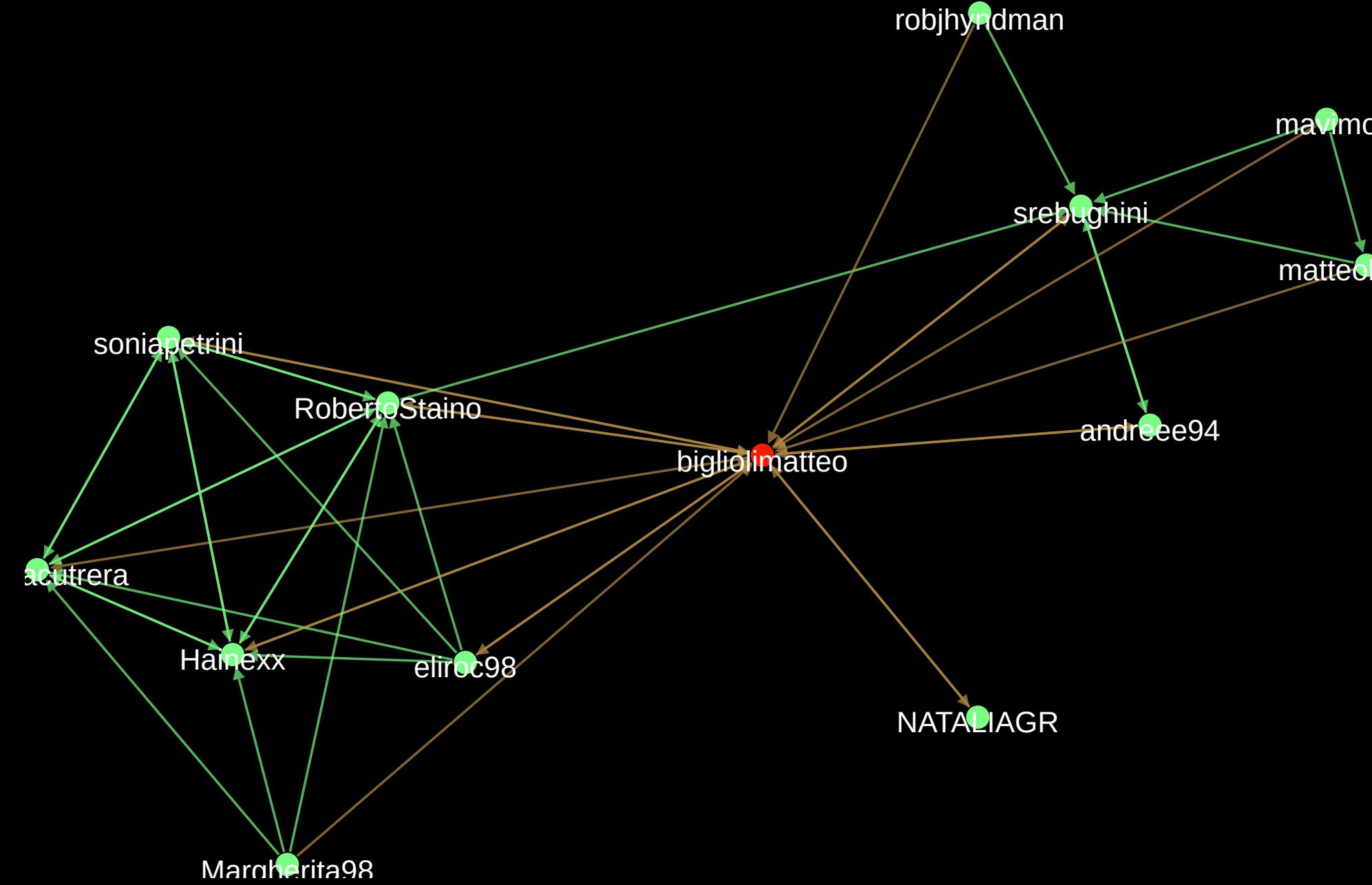
Density: 0.002

- Max distance: 3
  - Number of nodes: 6558
  - Number of edges: 77066
  - Average in/out-degree: 11.751
  - Degree distribution:
    - Mean: 23.503
    - Median: 13.0
    - Std: 43.732
    - Max: 1365
    - Min: 1

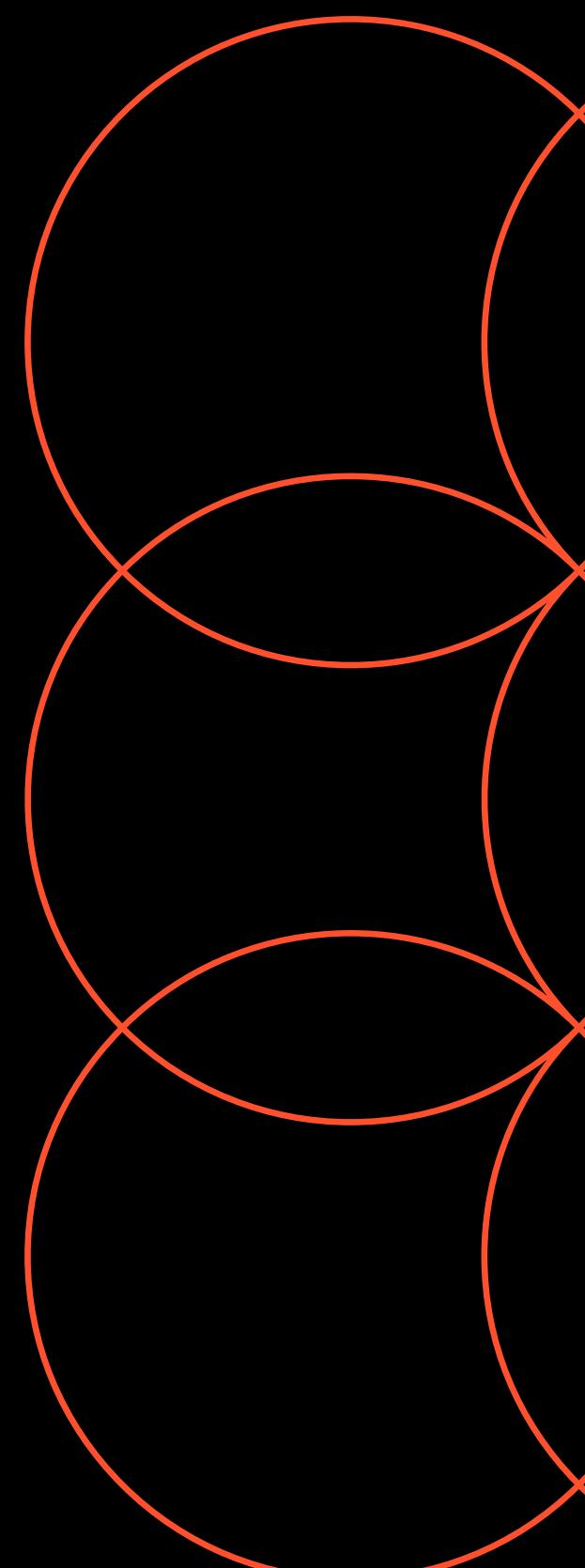
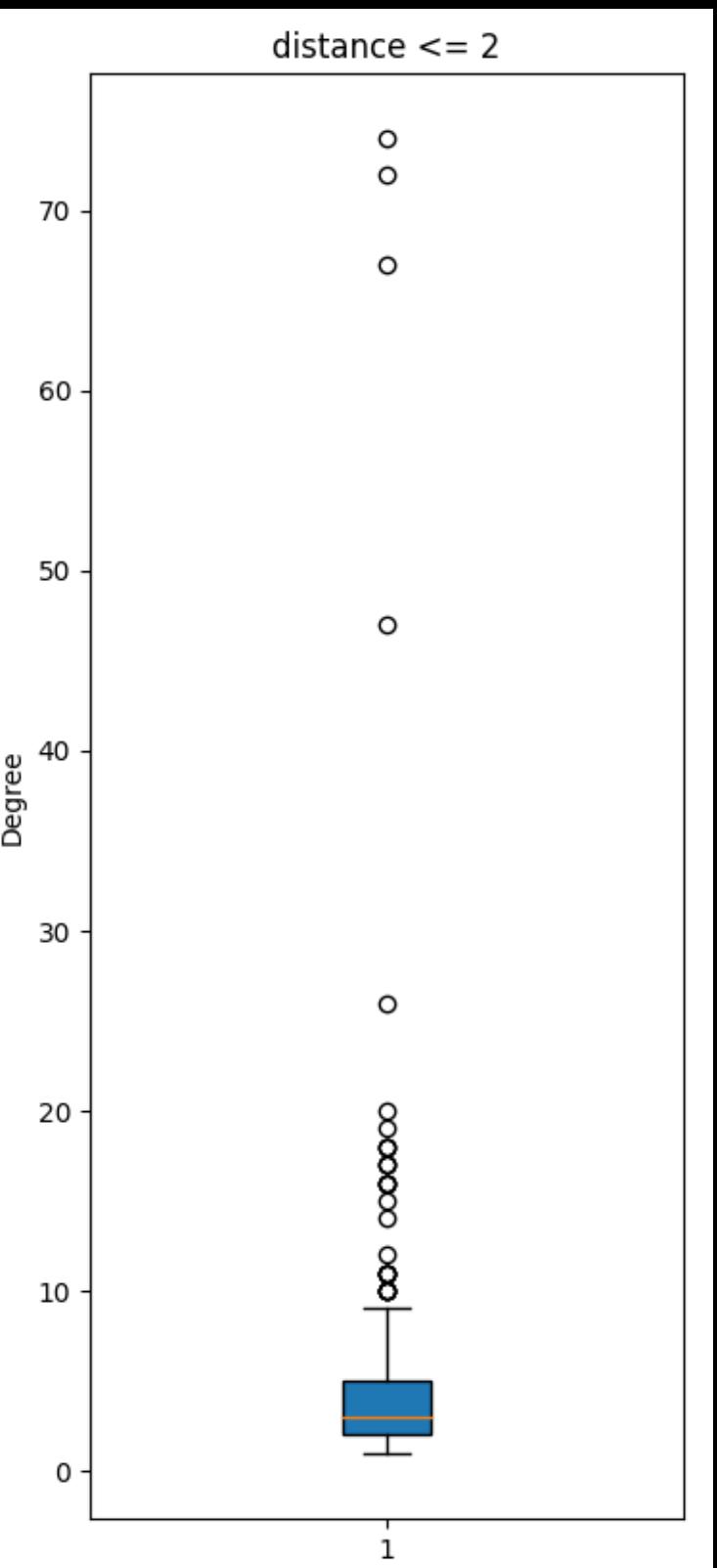
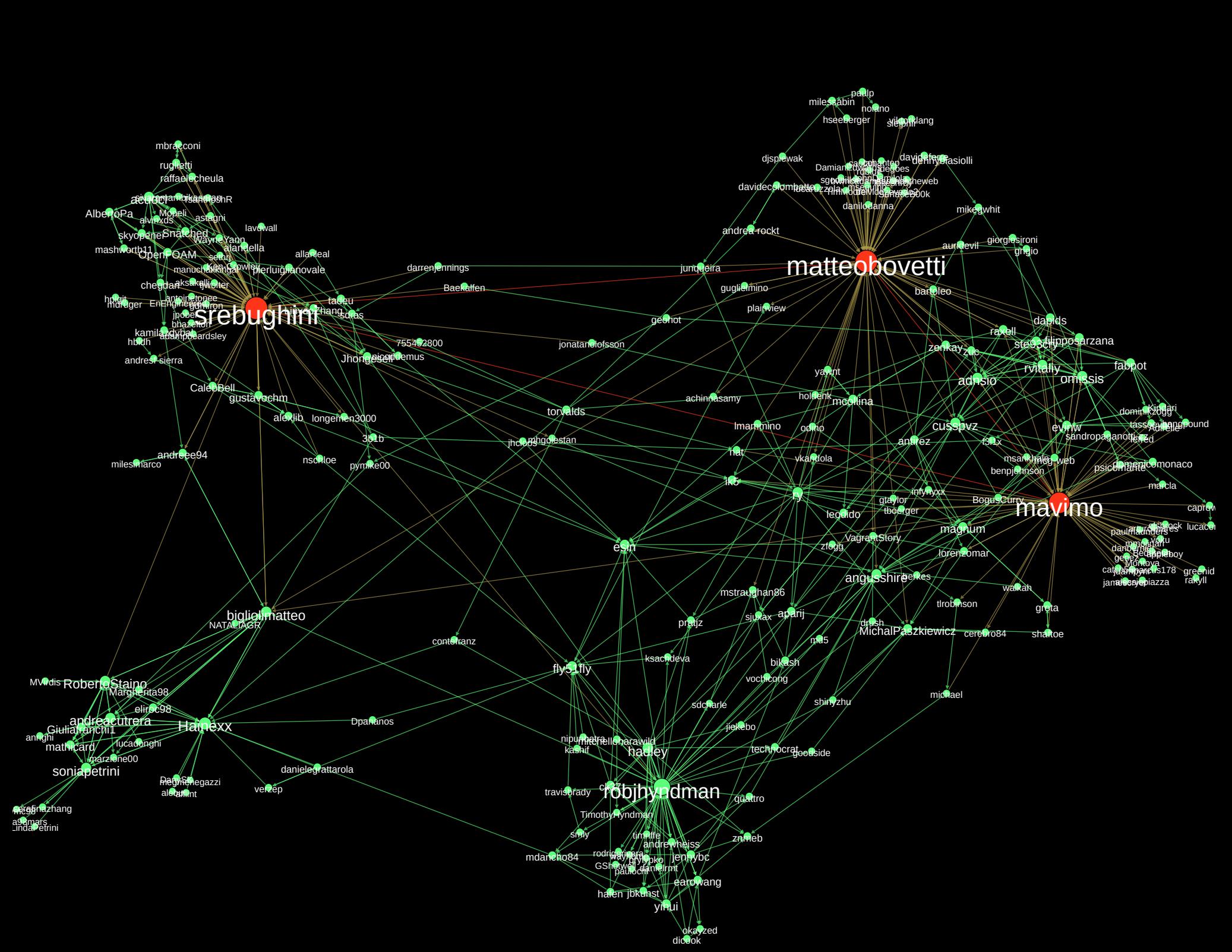
# ECCDF



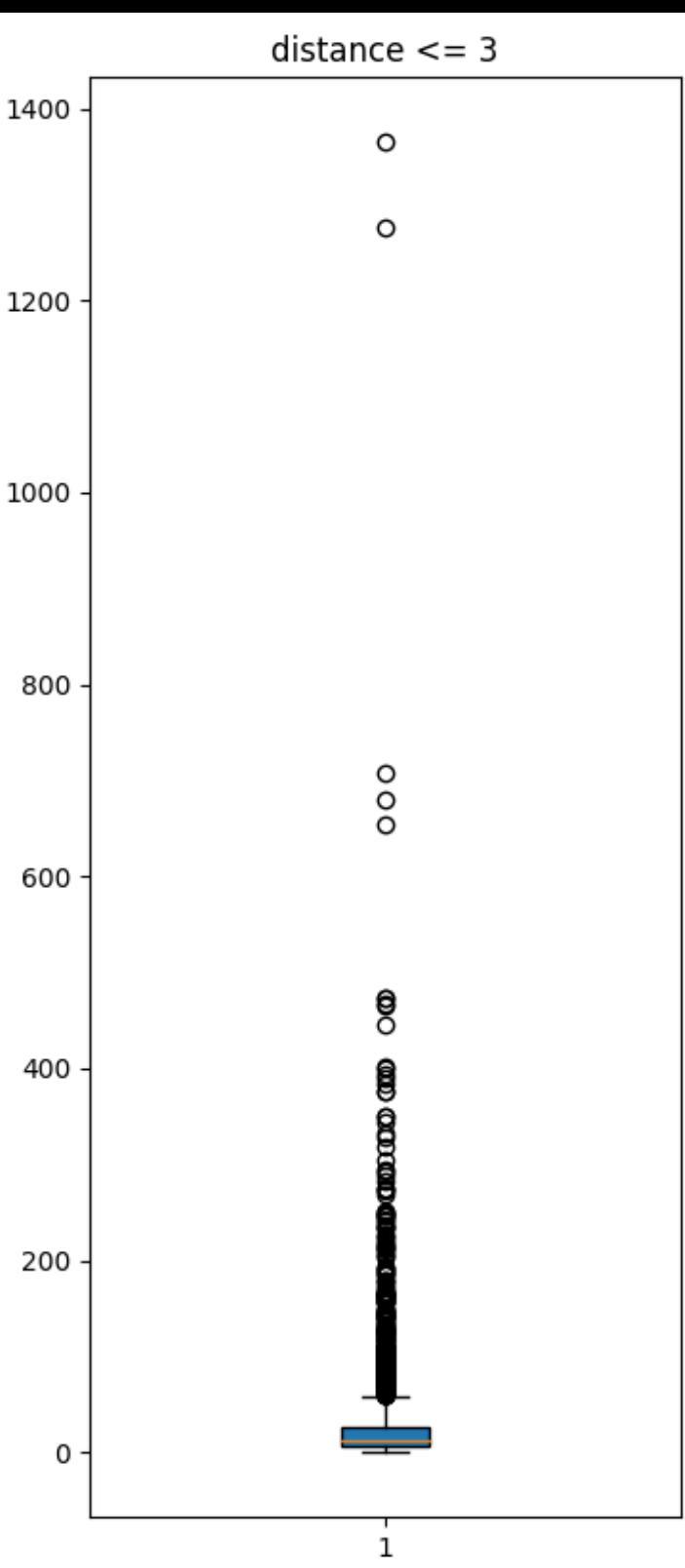
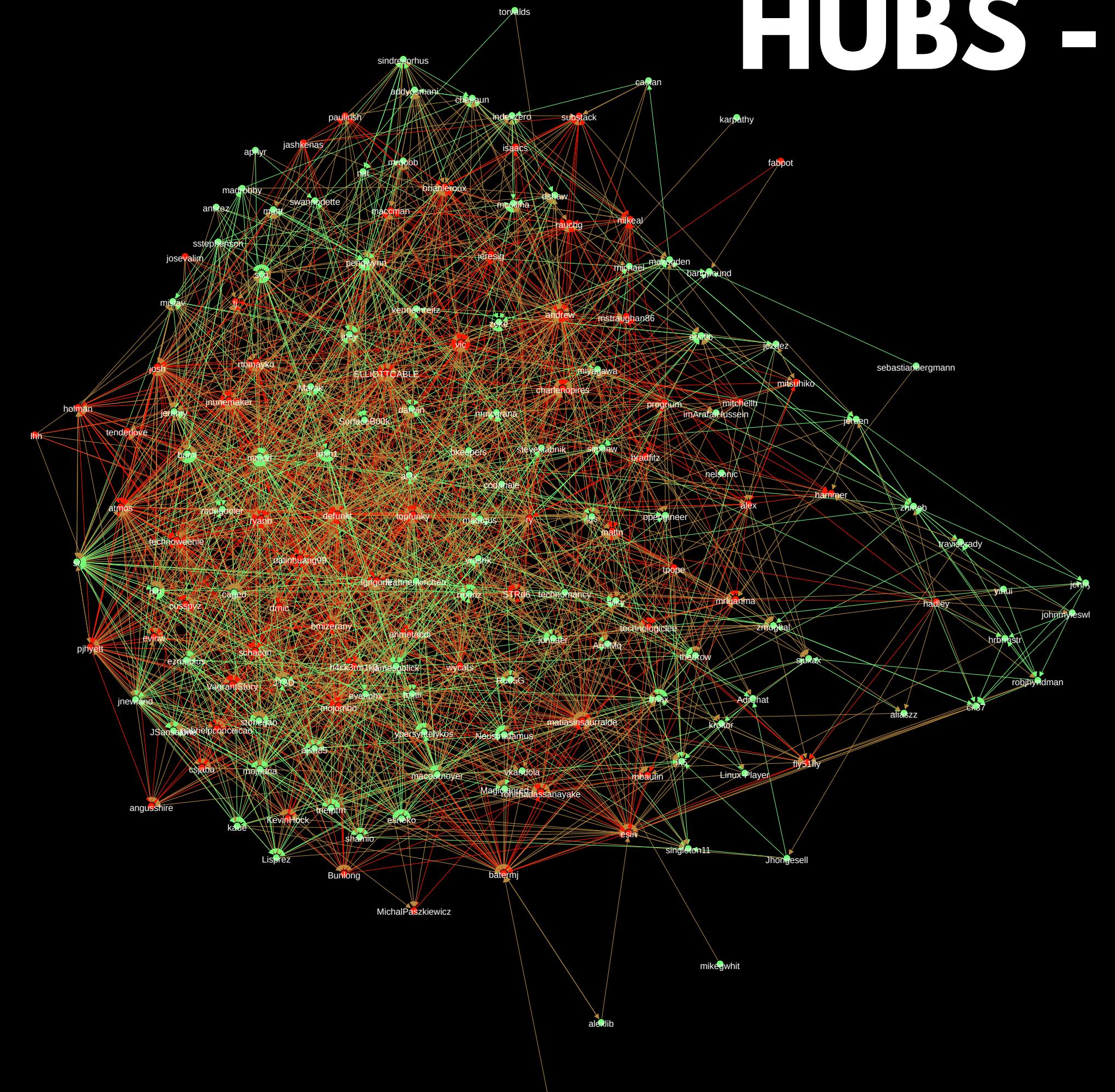
# HUBS - 99%



# HUBS - 99%



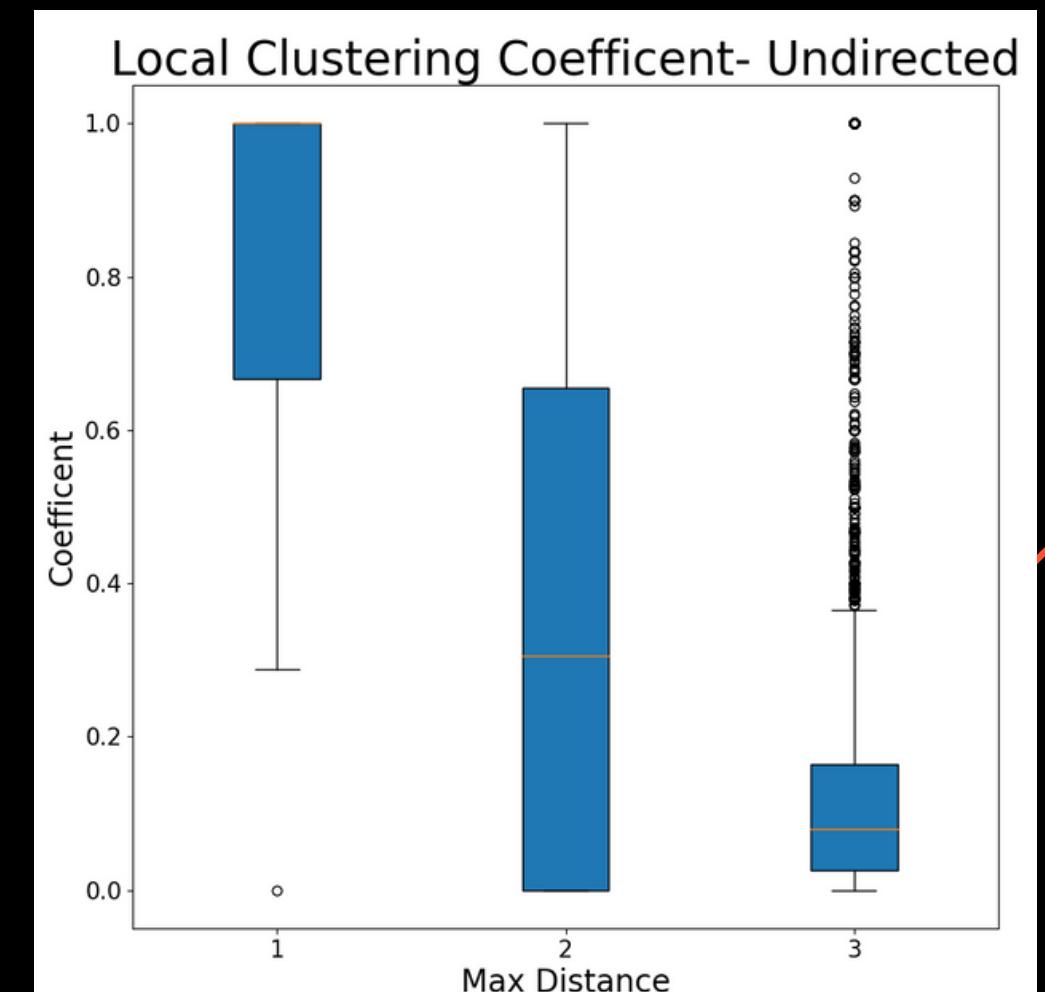
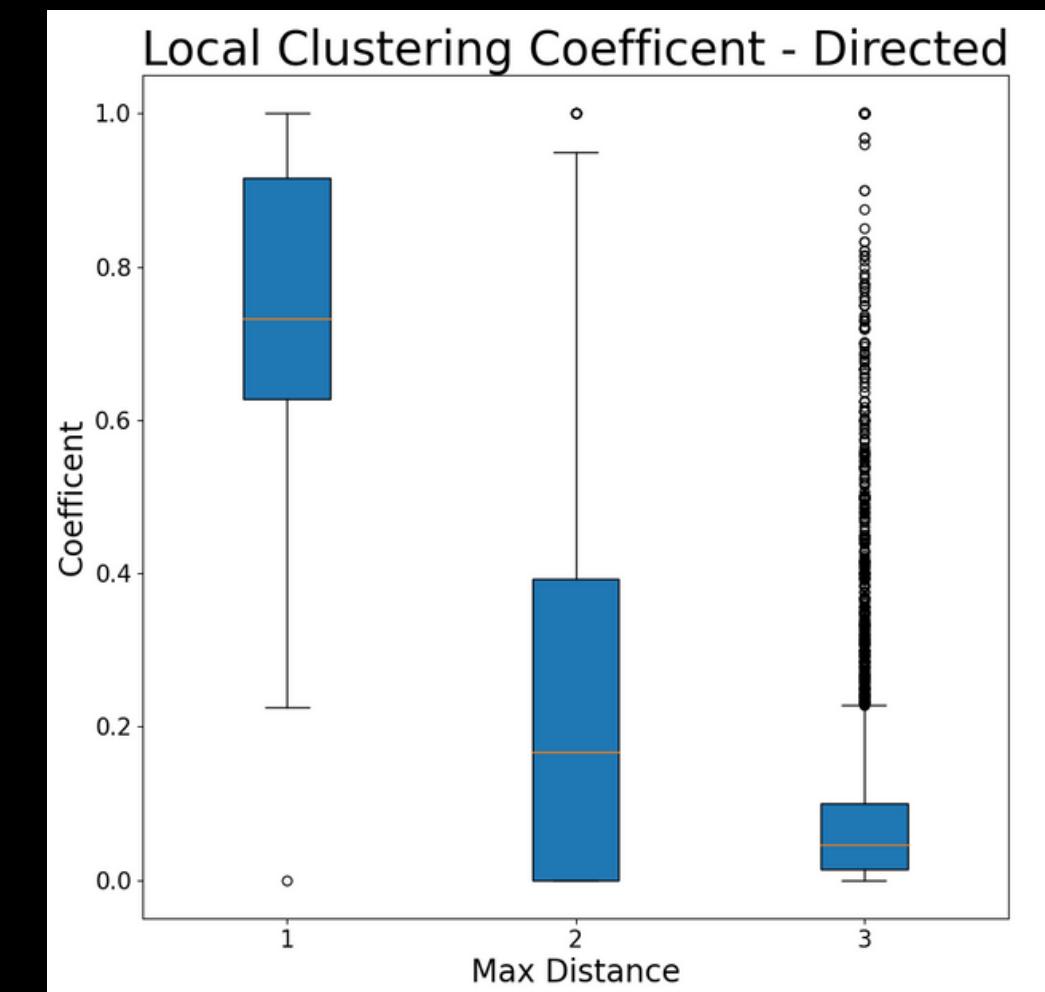
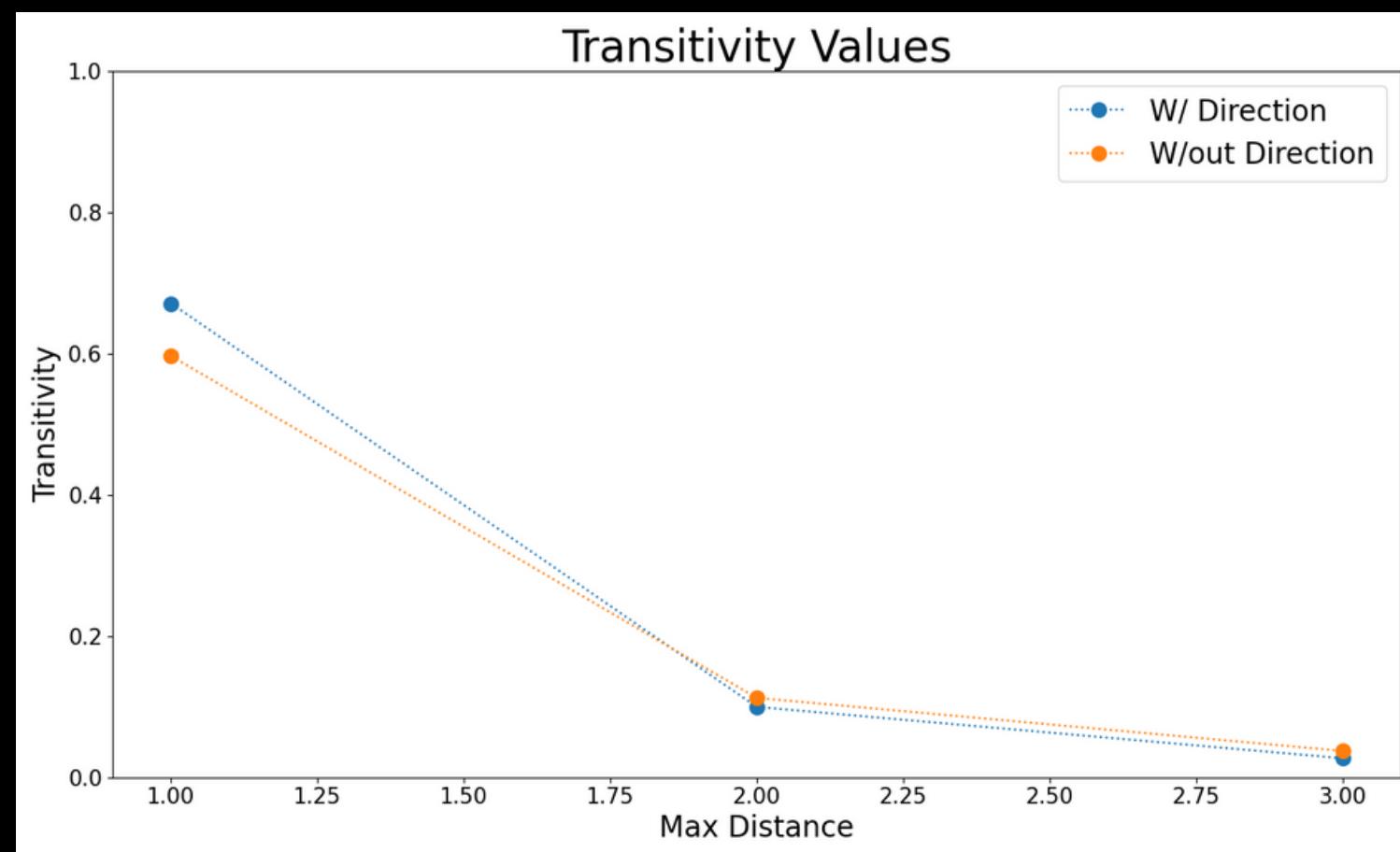
# HUBS - 99%



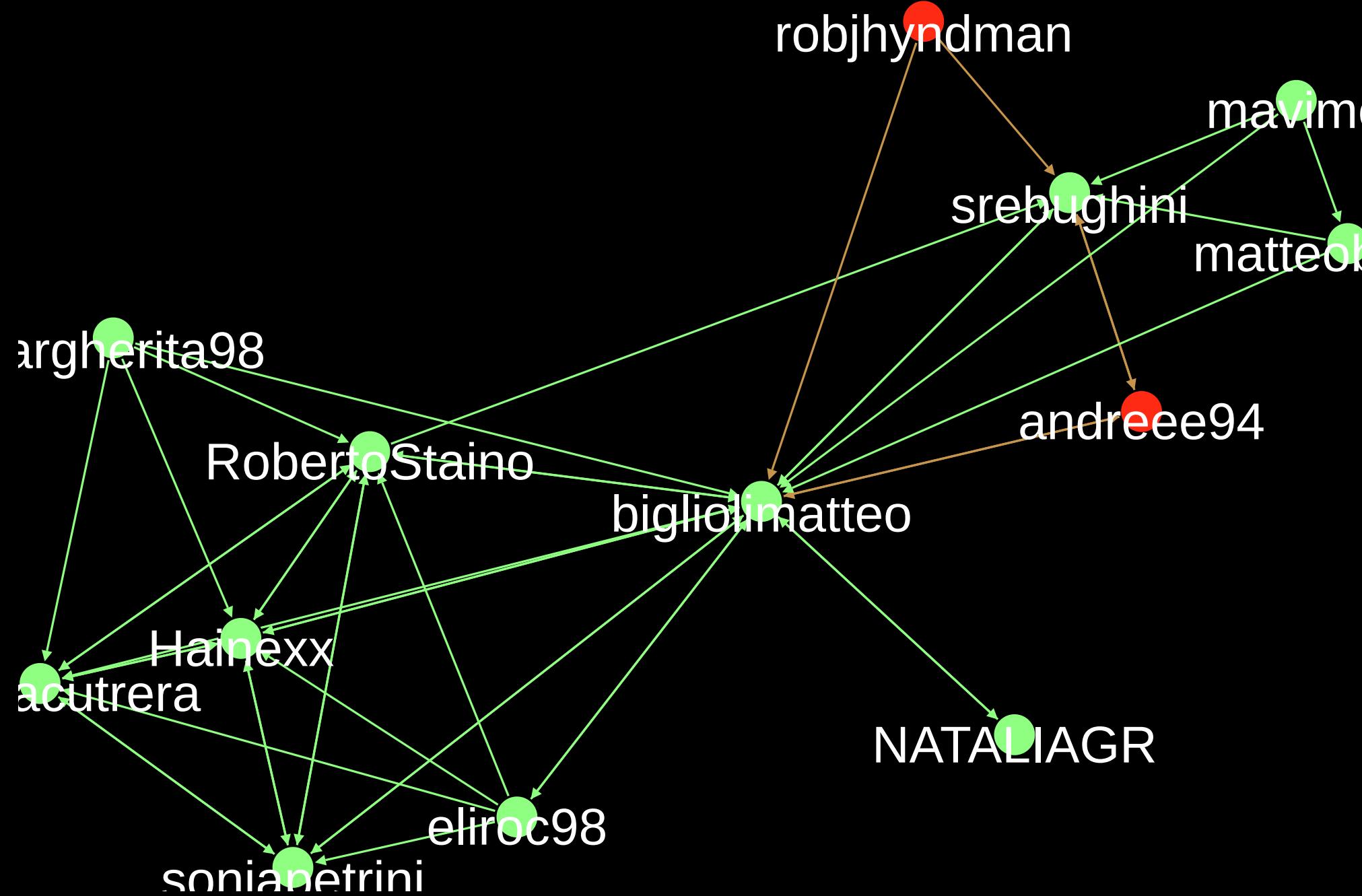
# TRANSITIVITY

Global

Local



# TRANSITIVITY

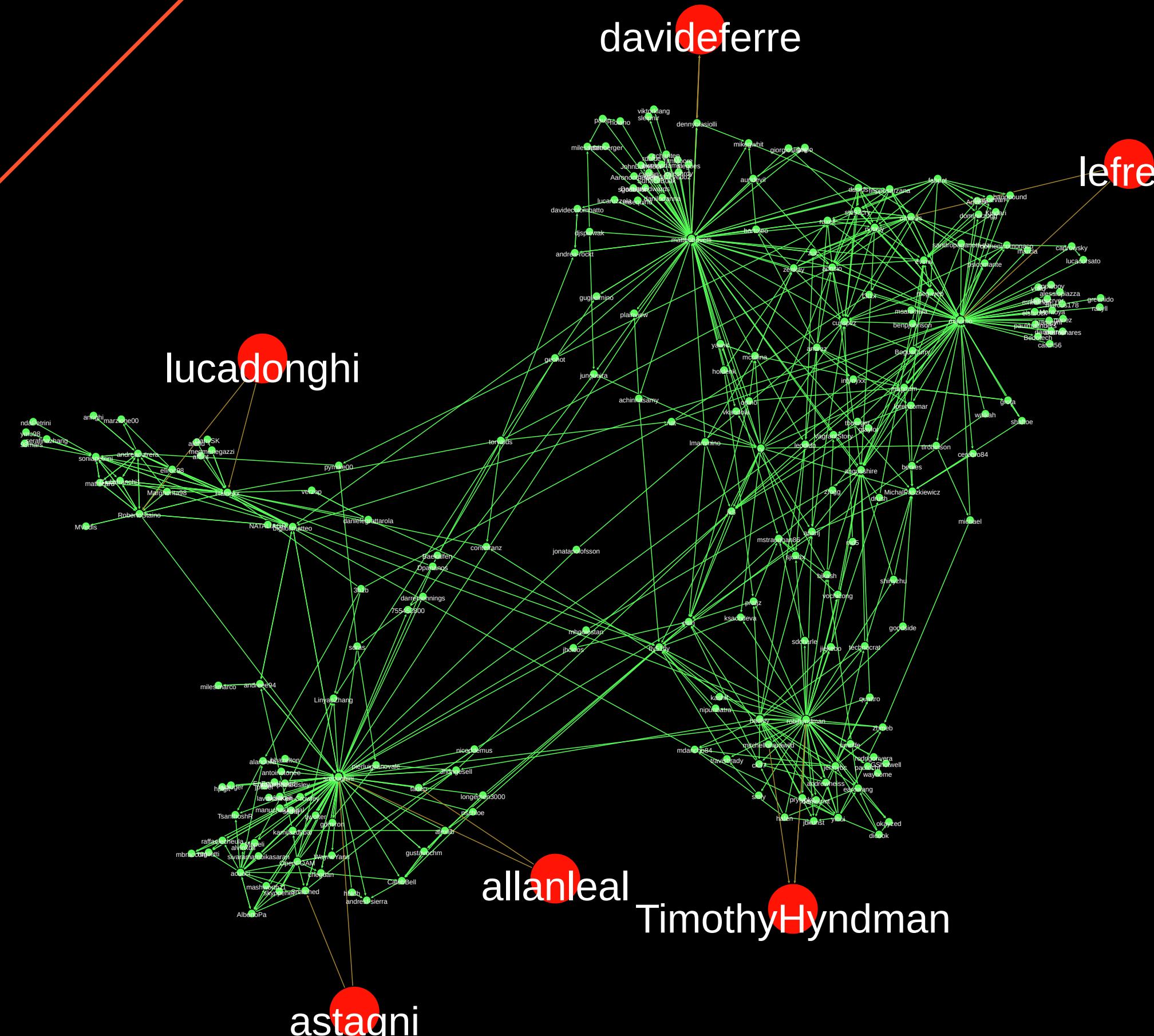


Nodes with local clustering value greater than **0.95**

Note that they have just a few connections

\*Taking direction into account

# TRANSITIVITY



Nodes with local clustering  
value greater than 0.95

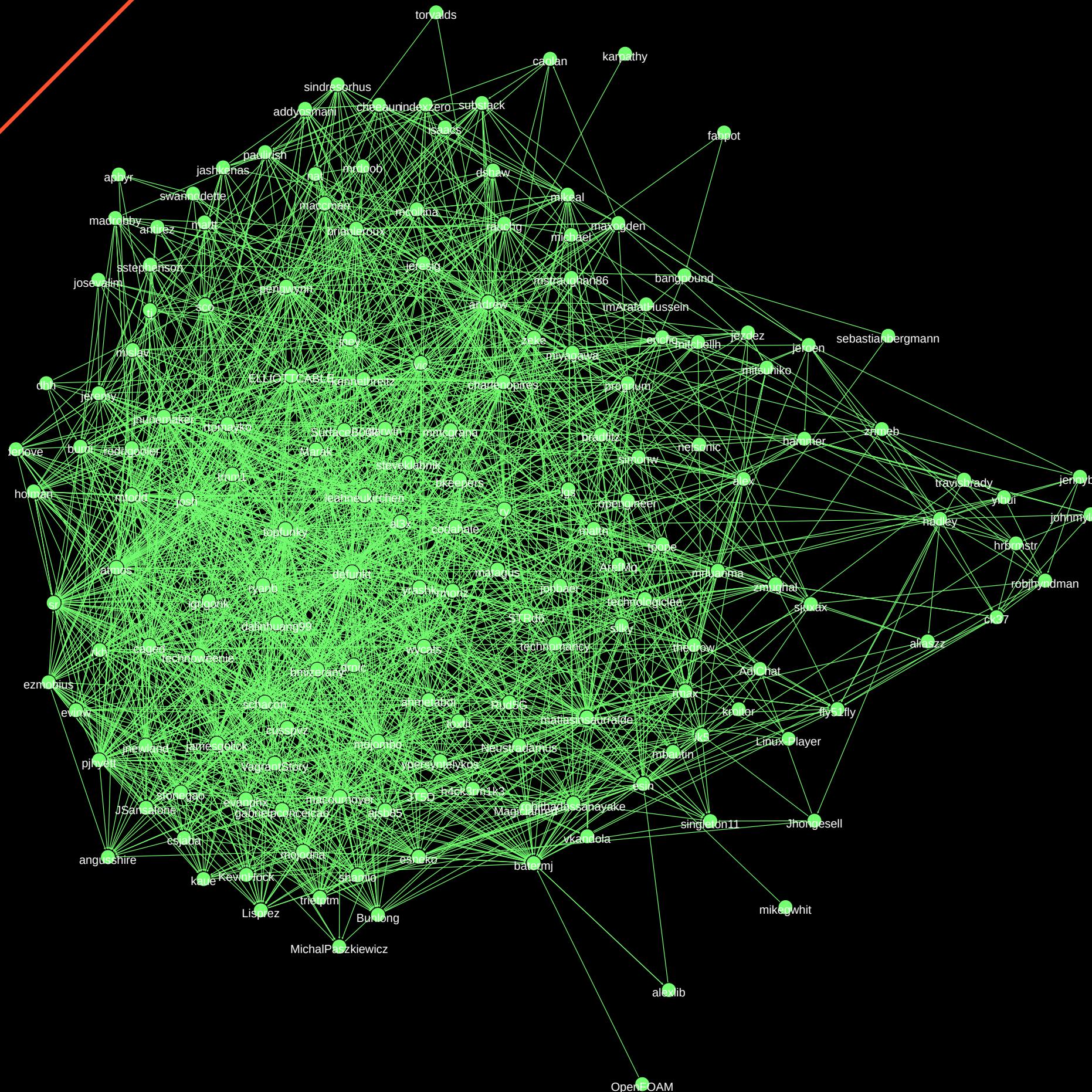
Note that they have just a few  
connections

\*Taking direction into account

# TRANSITIVITY

# Nodes with local clustering value greater than 0.95

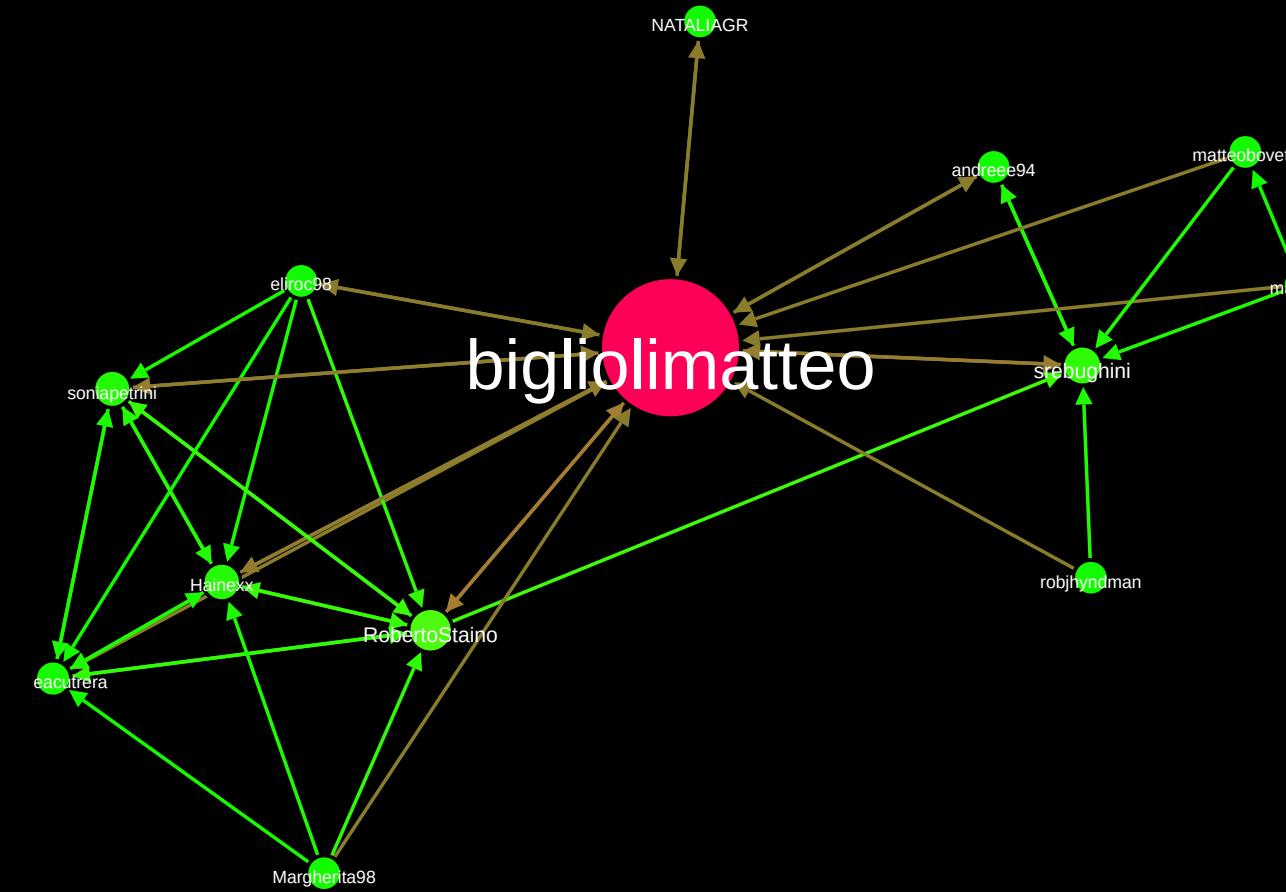
**Note that they have just a few connections**



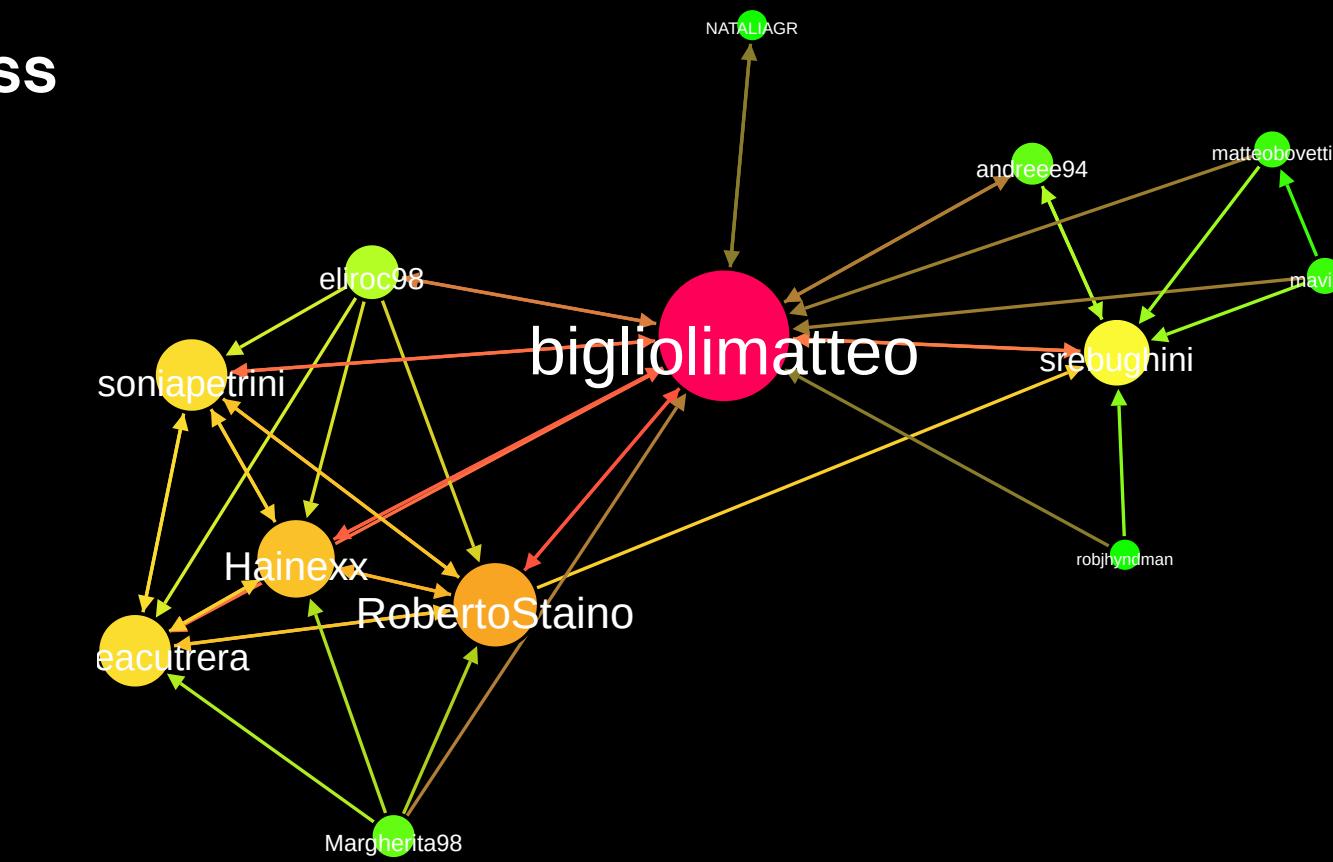
\*Taking direction into account

# CENTRALITIES

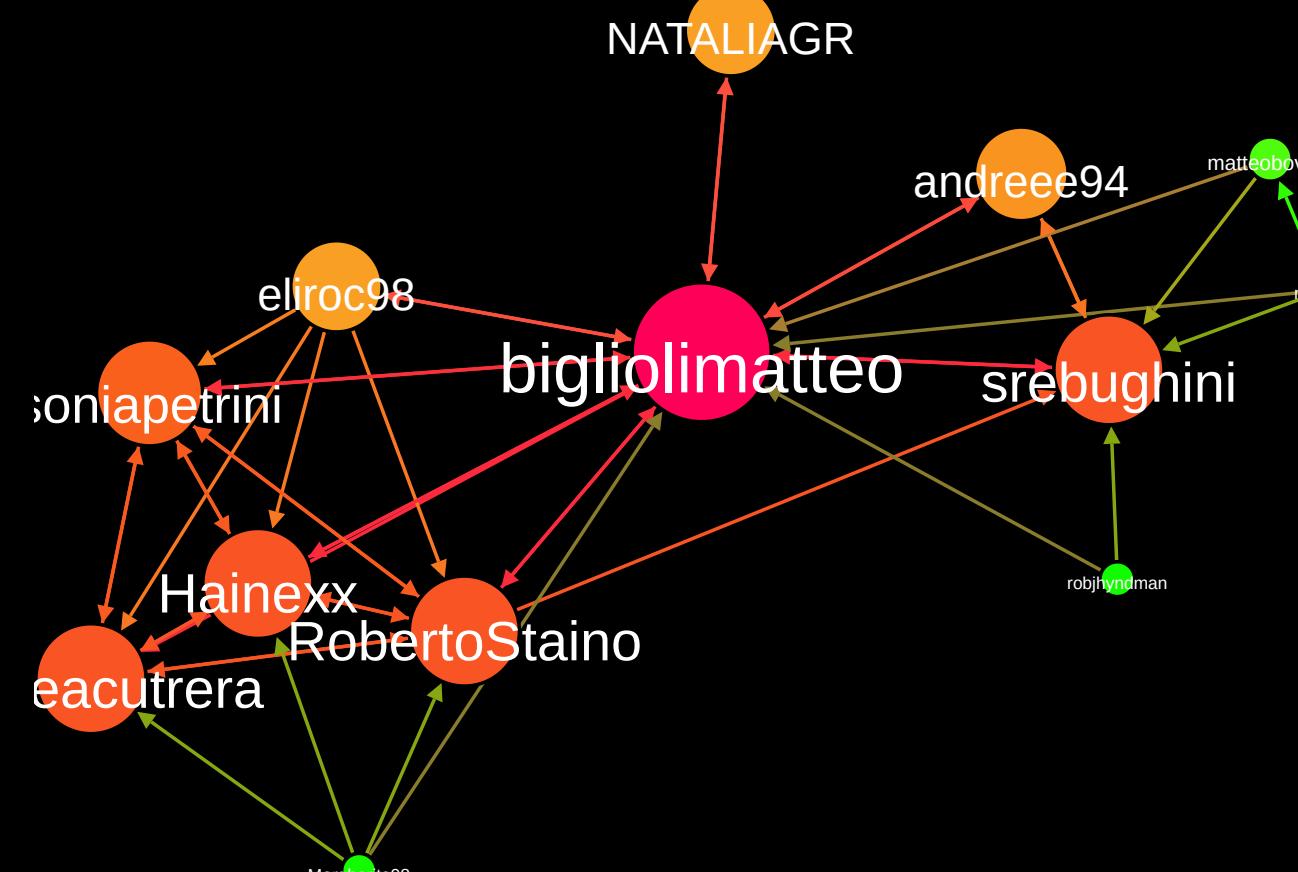
Degree



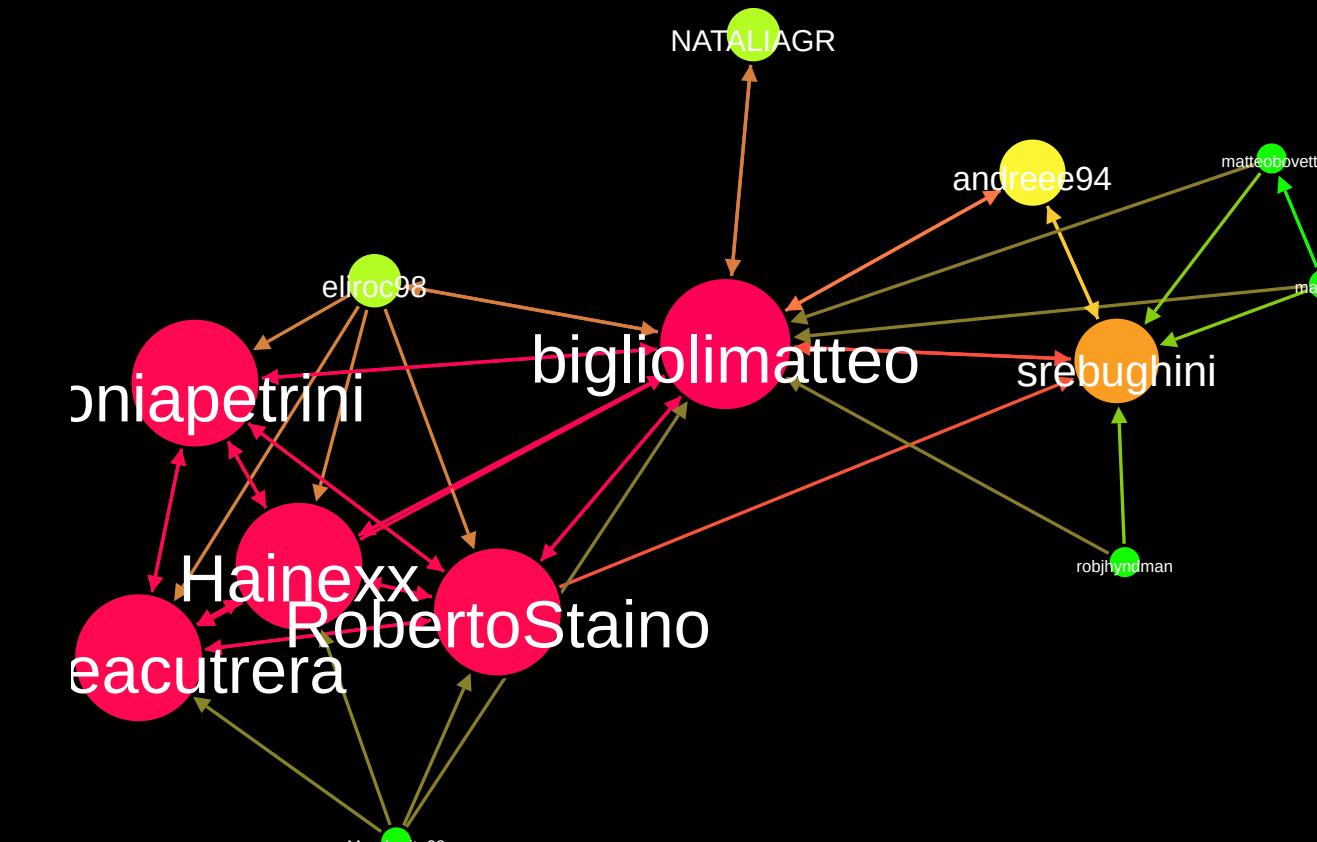
Betweenness



Closeness

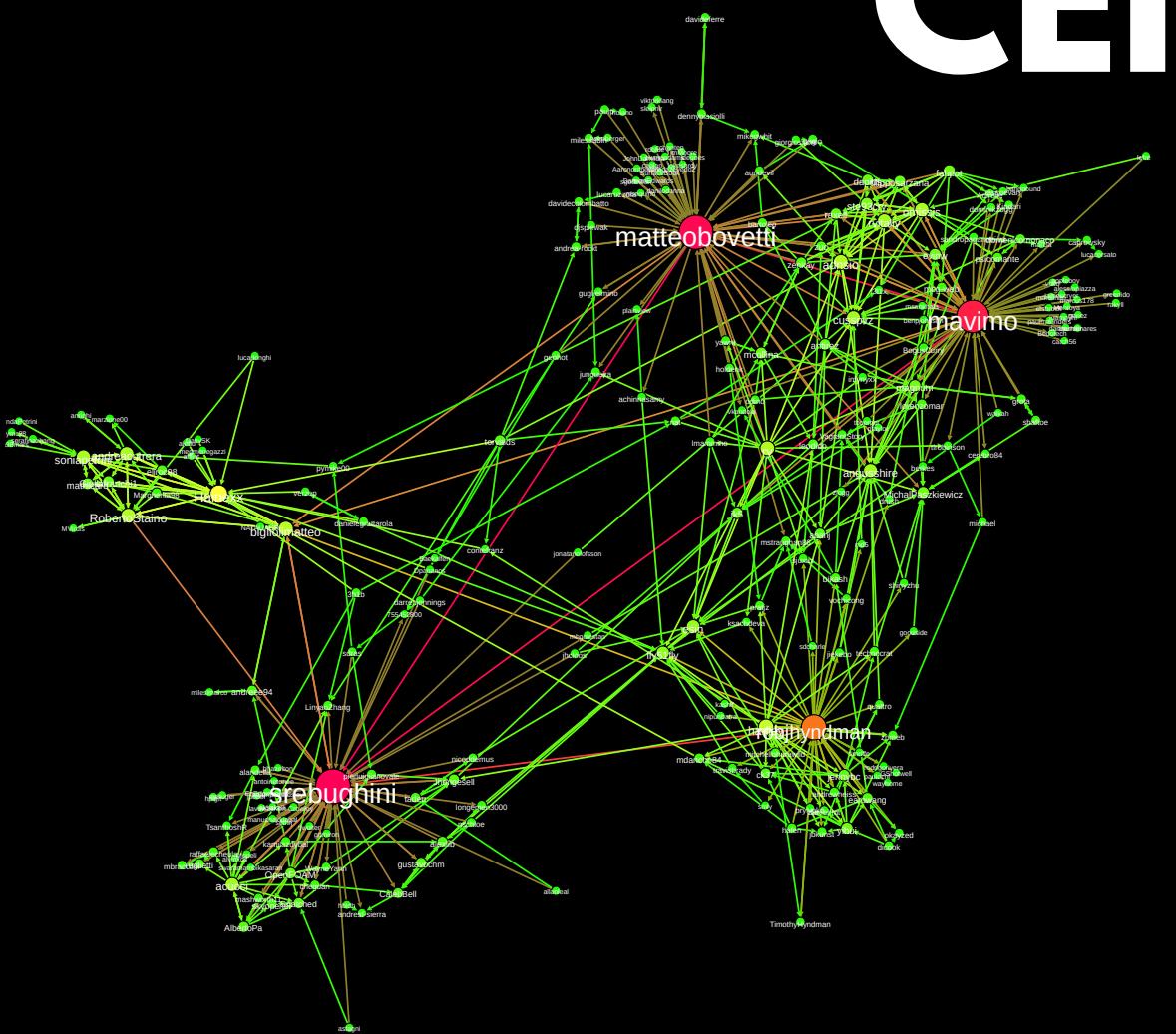


Eigenvector

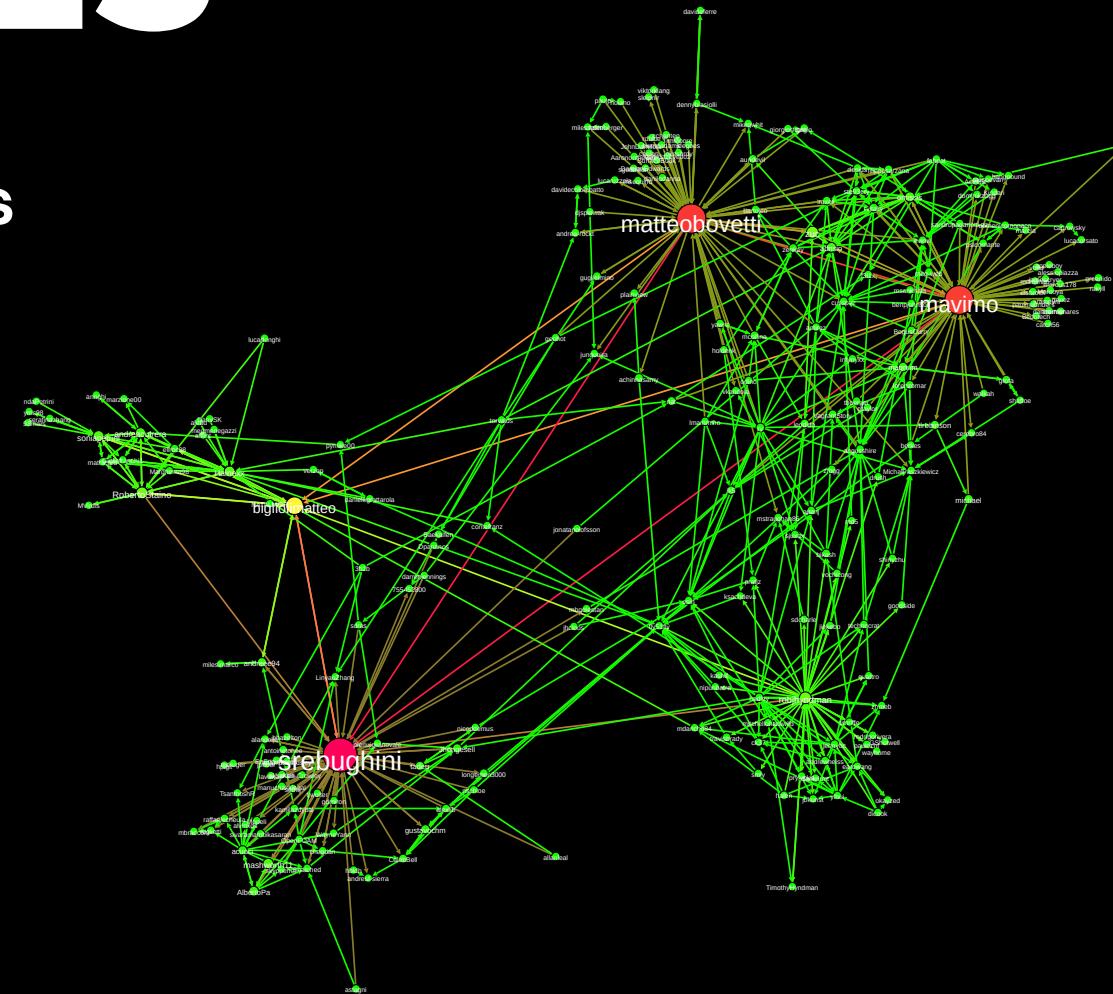


# CENTRALITIES

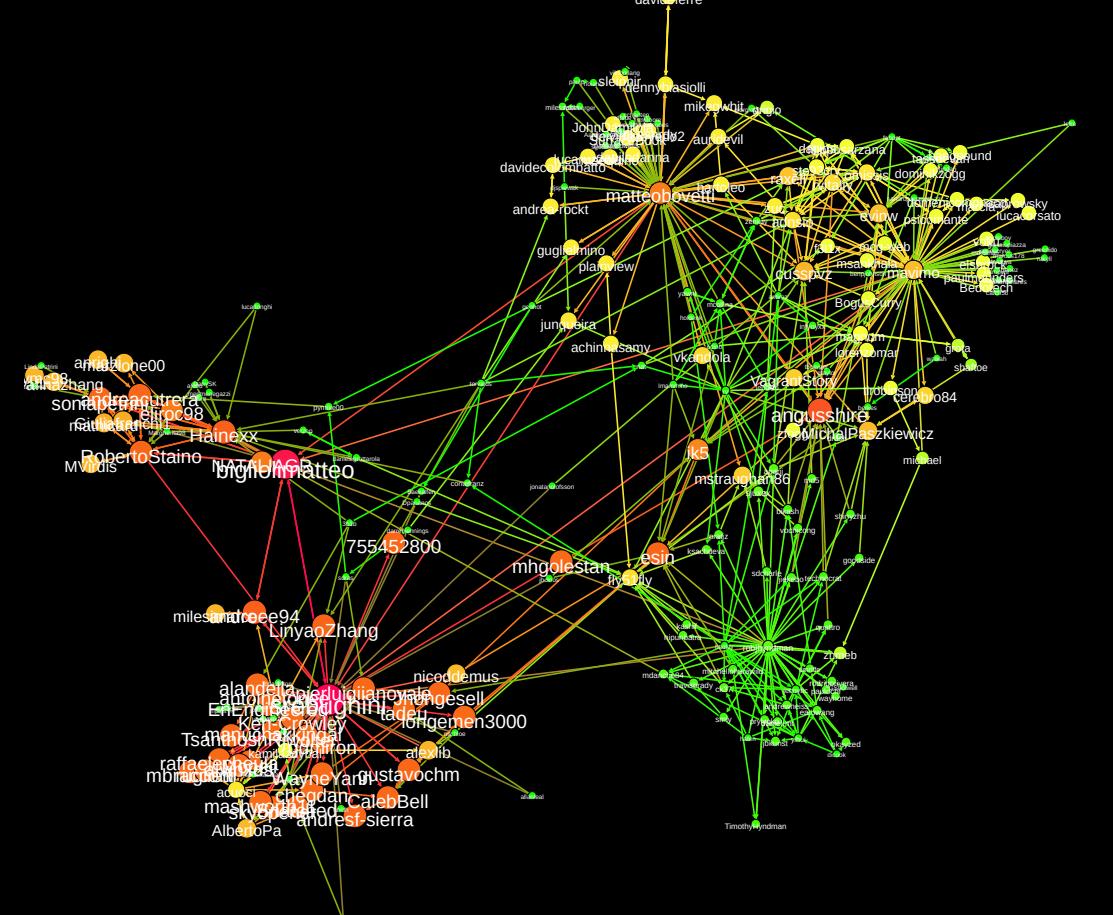
Degree



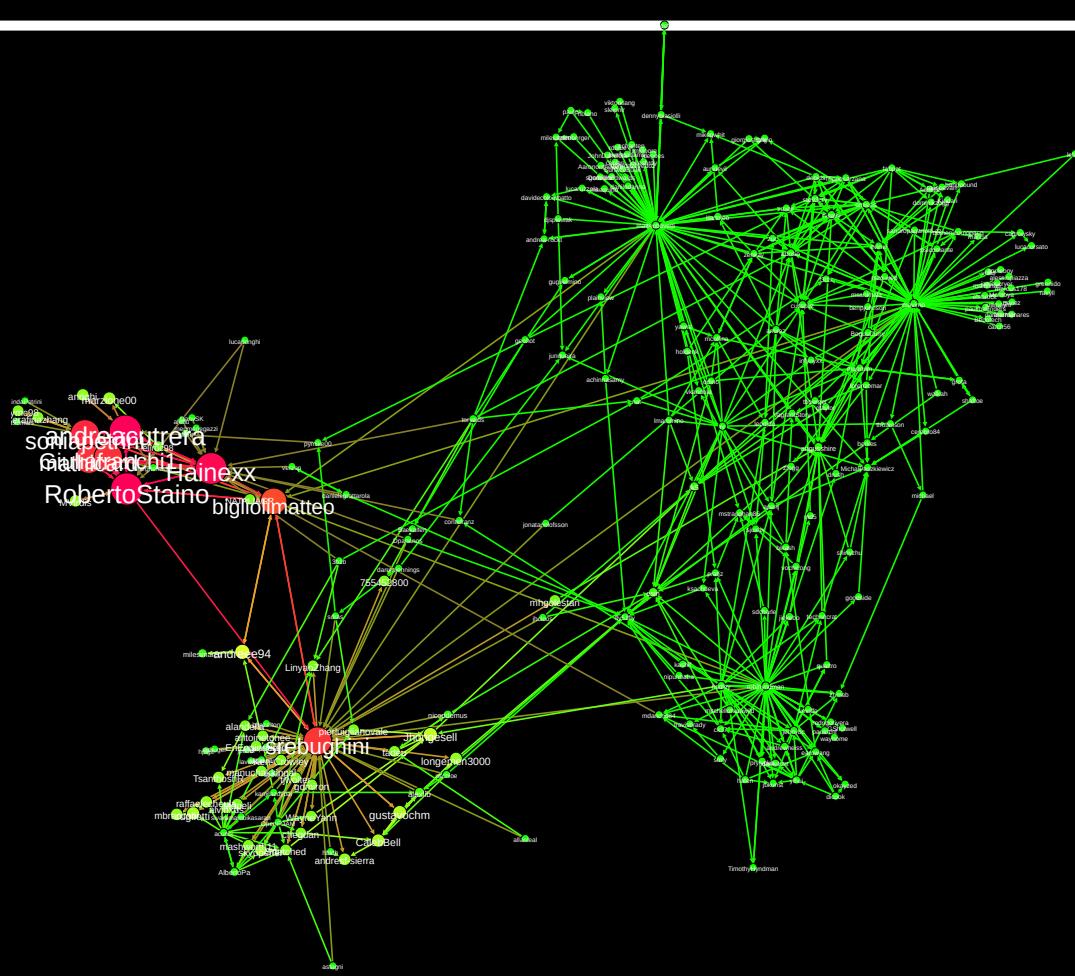
Betweenness



Closeness

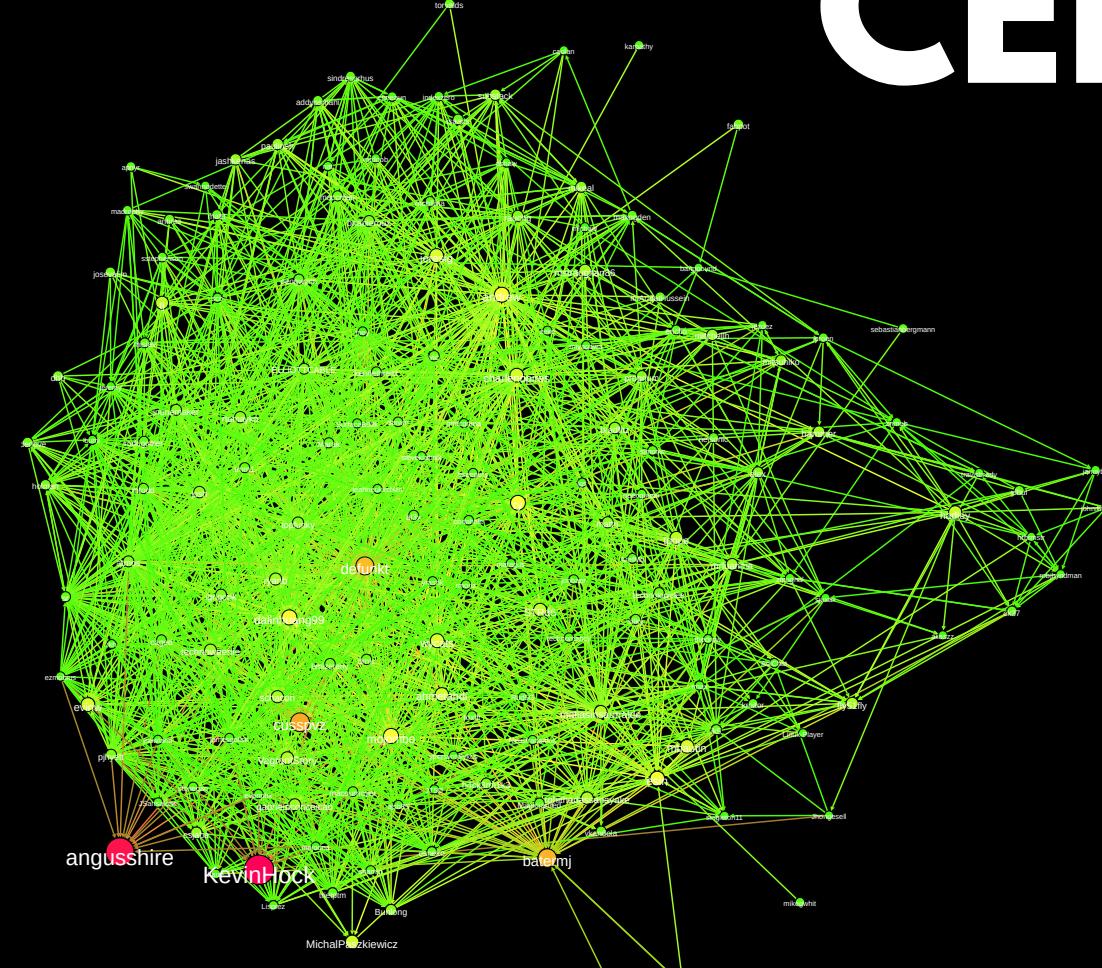


Eigenvector

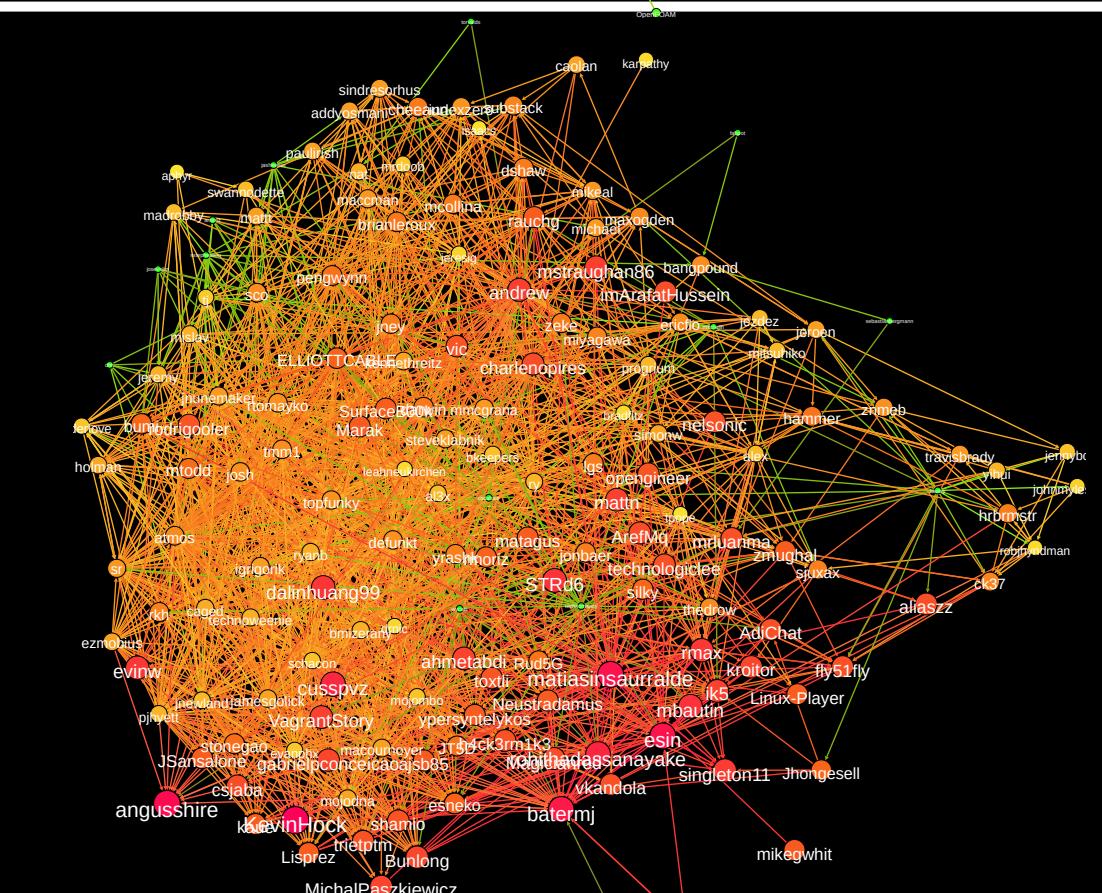


# CENTRALITIES

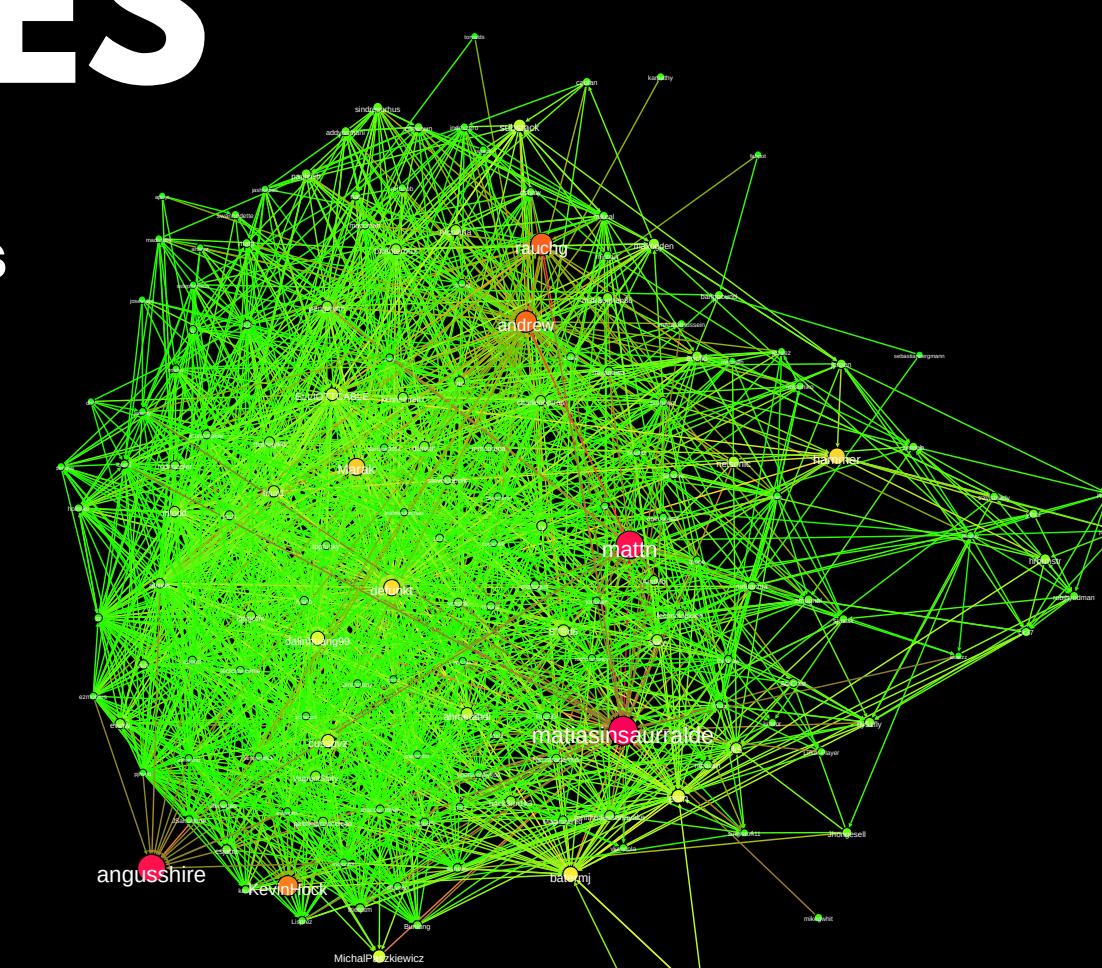
# Degree



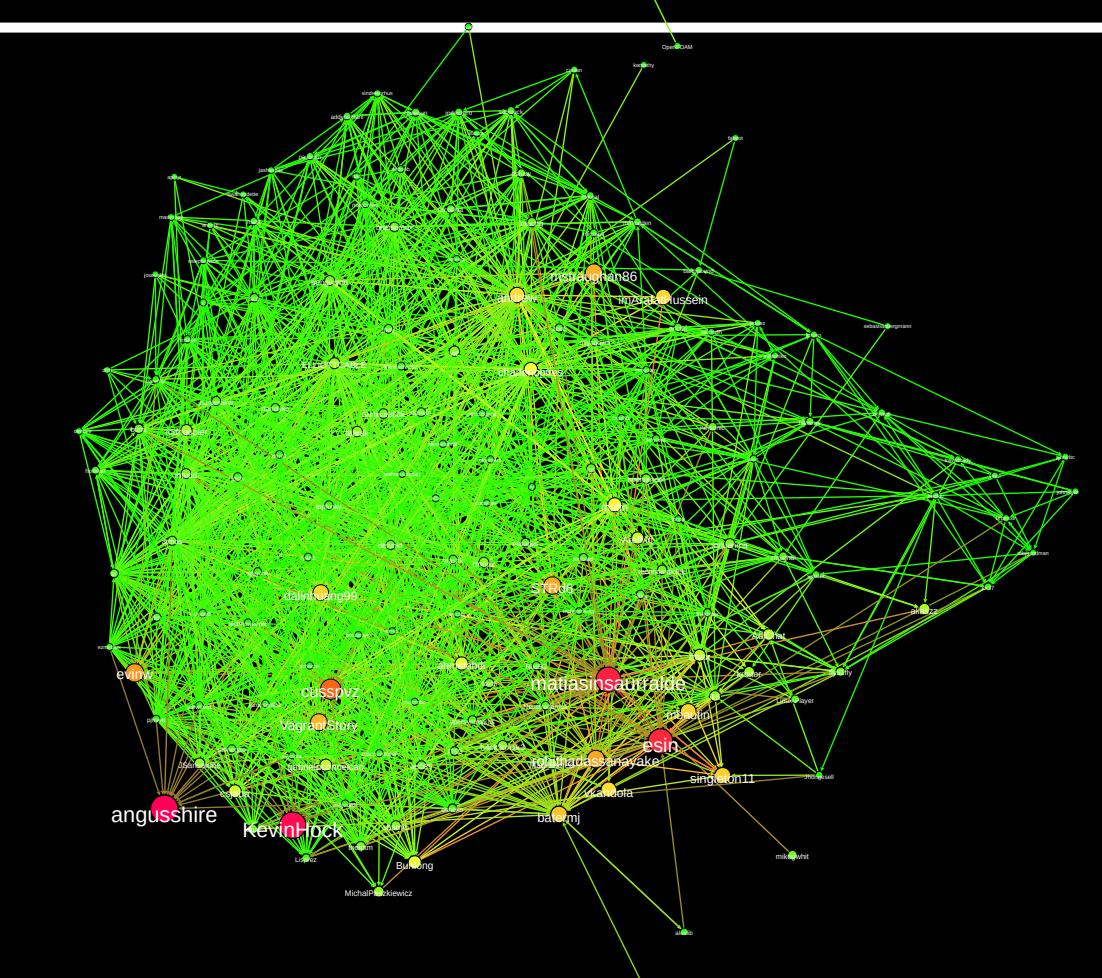
# Closeness



# Betweenness

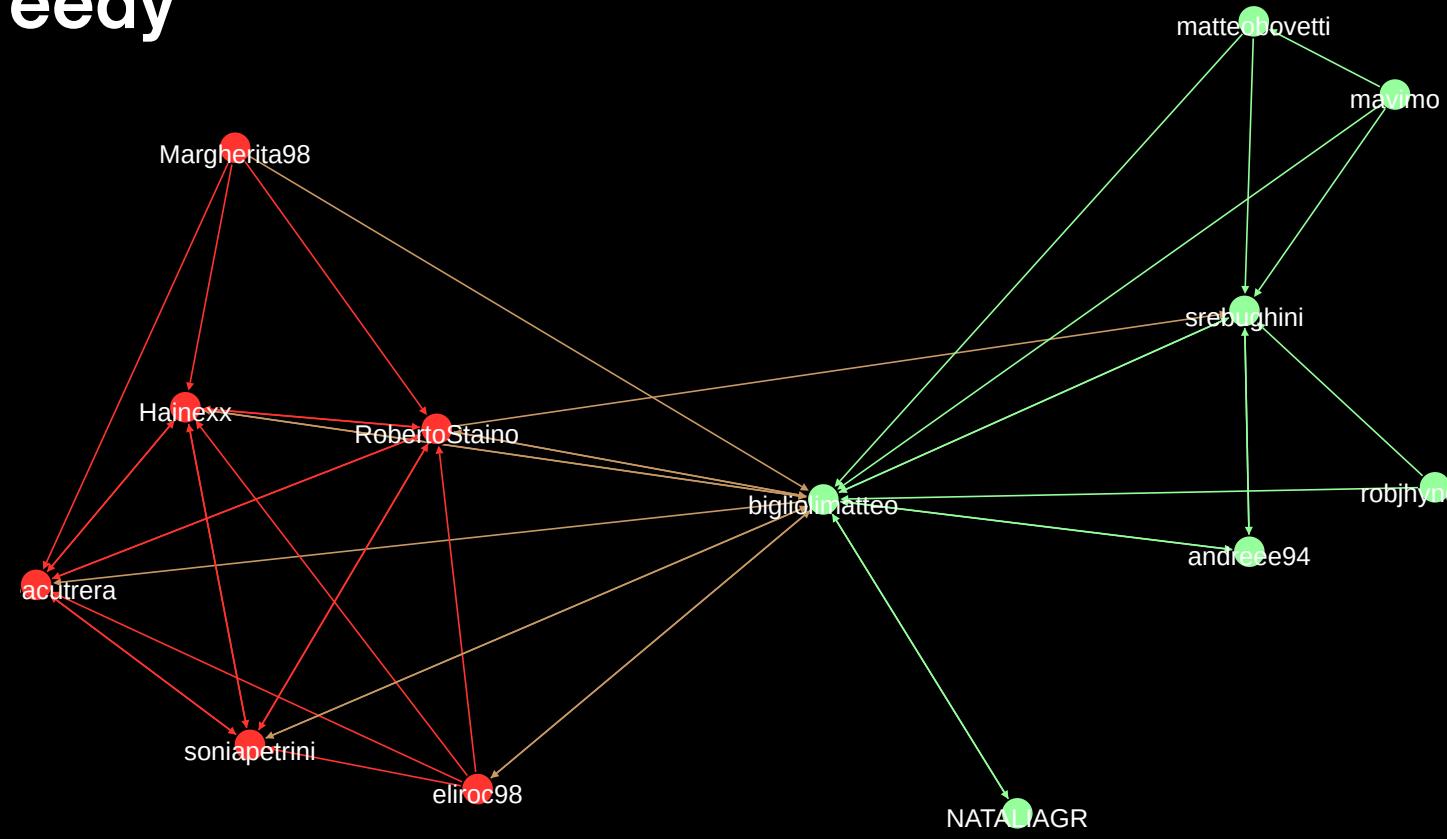


# Eigenvector

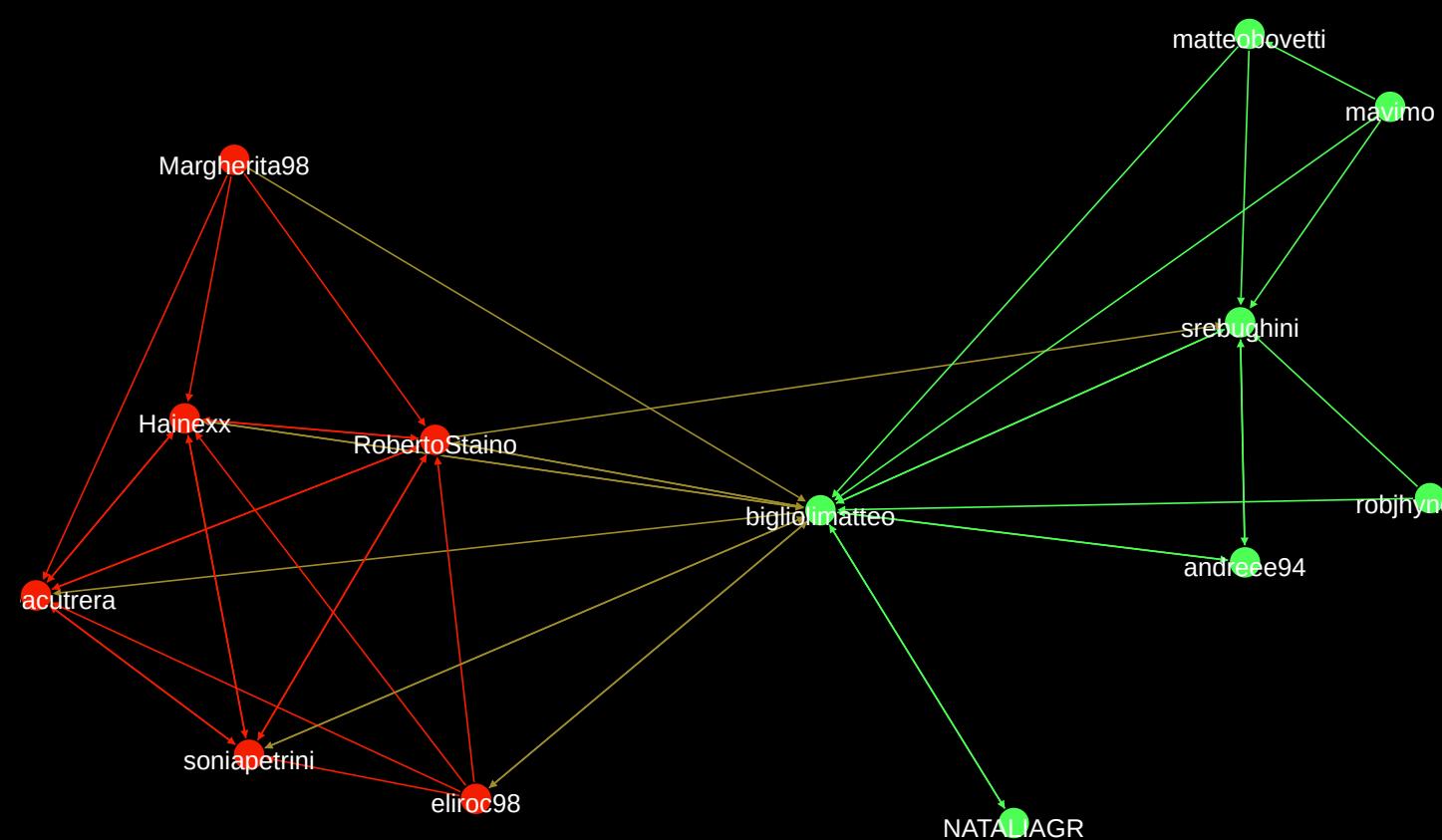


# COMMUNITIES

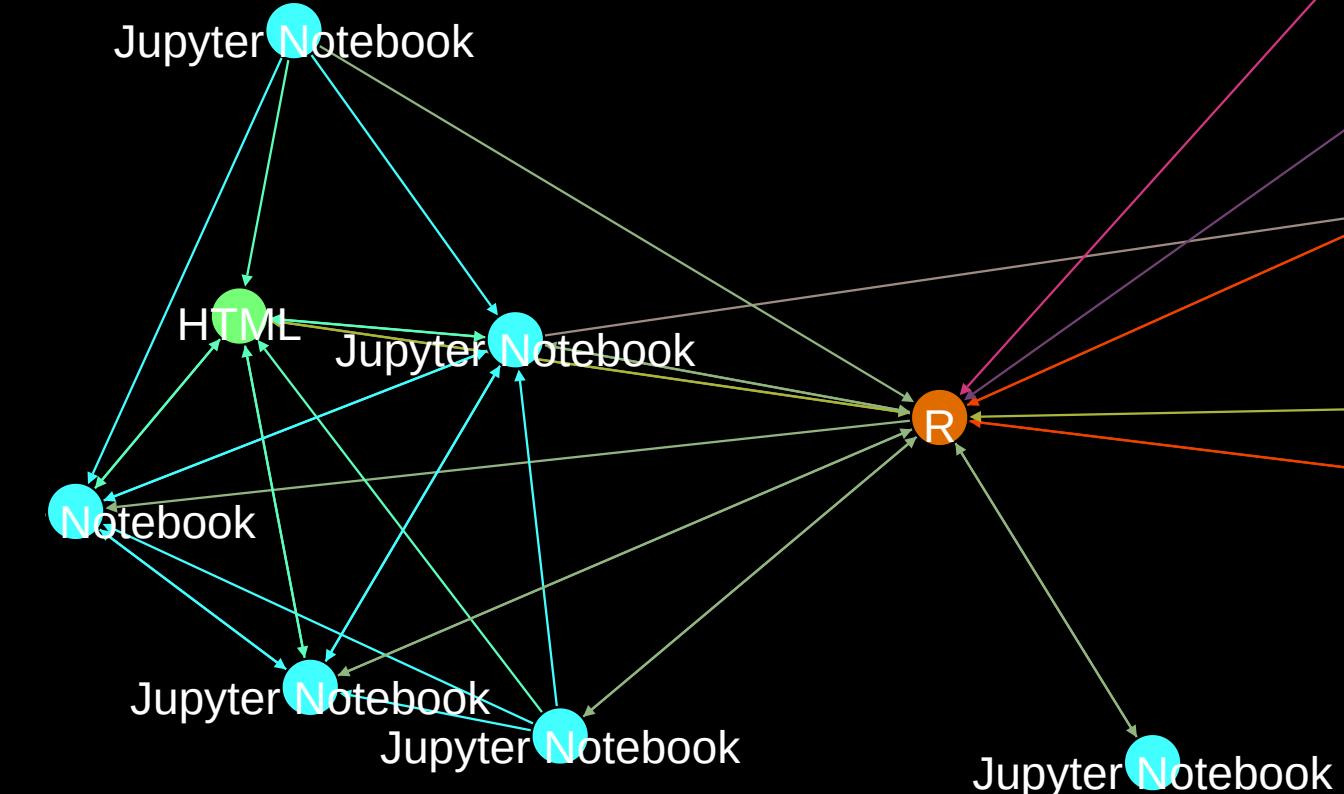
Greedy



Louvain



Preferred Language\*



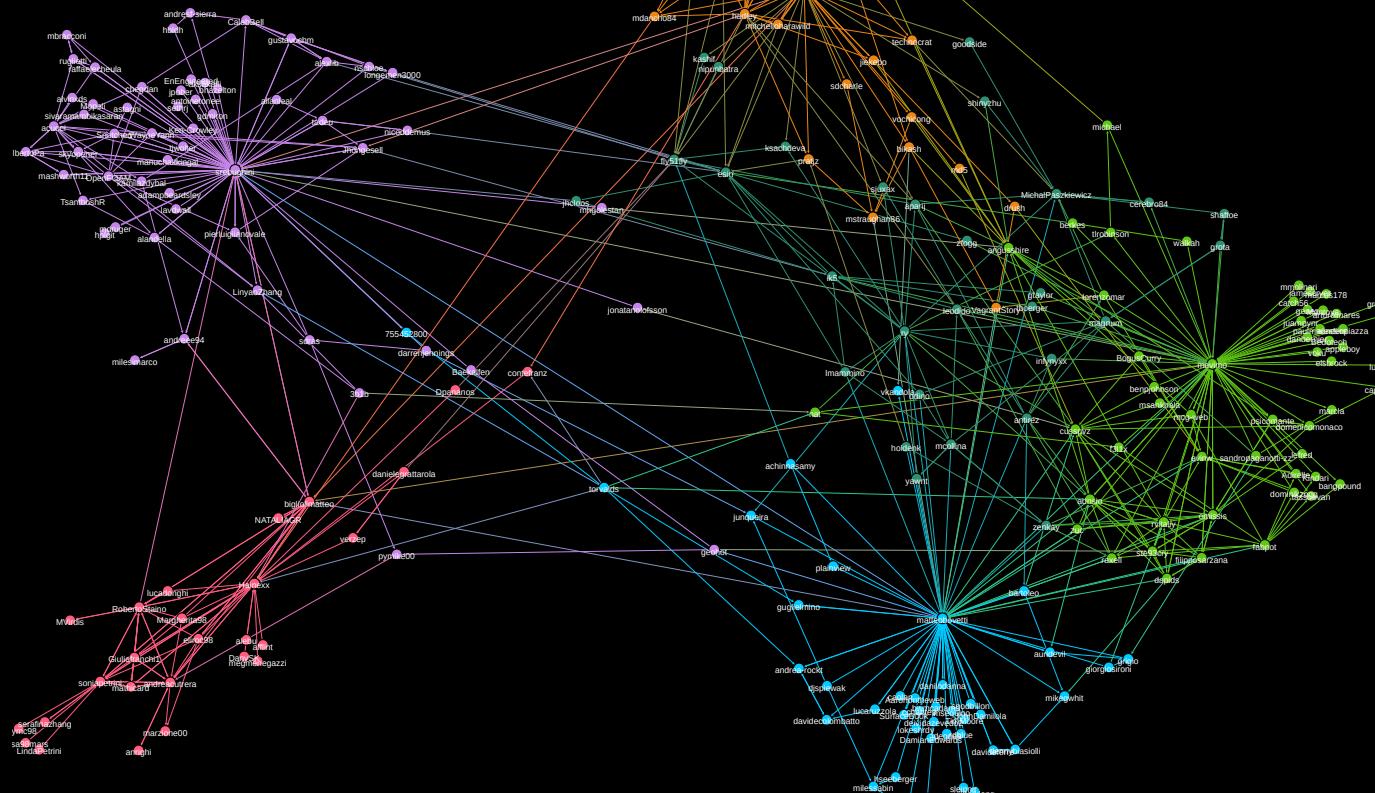
Assortativity

0.006

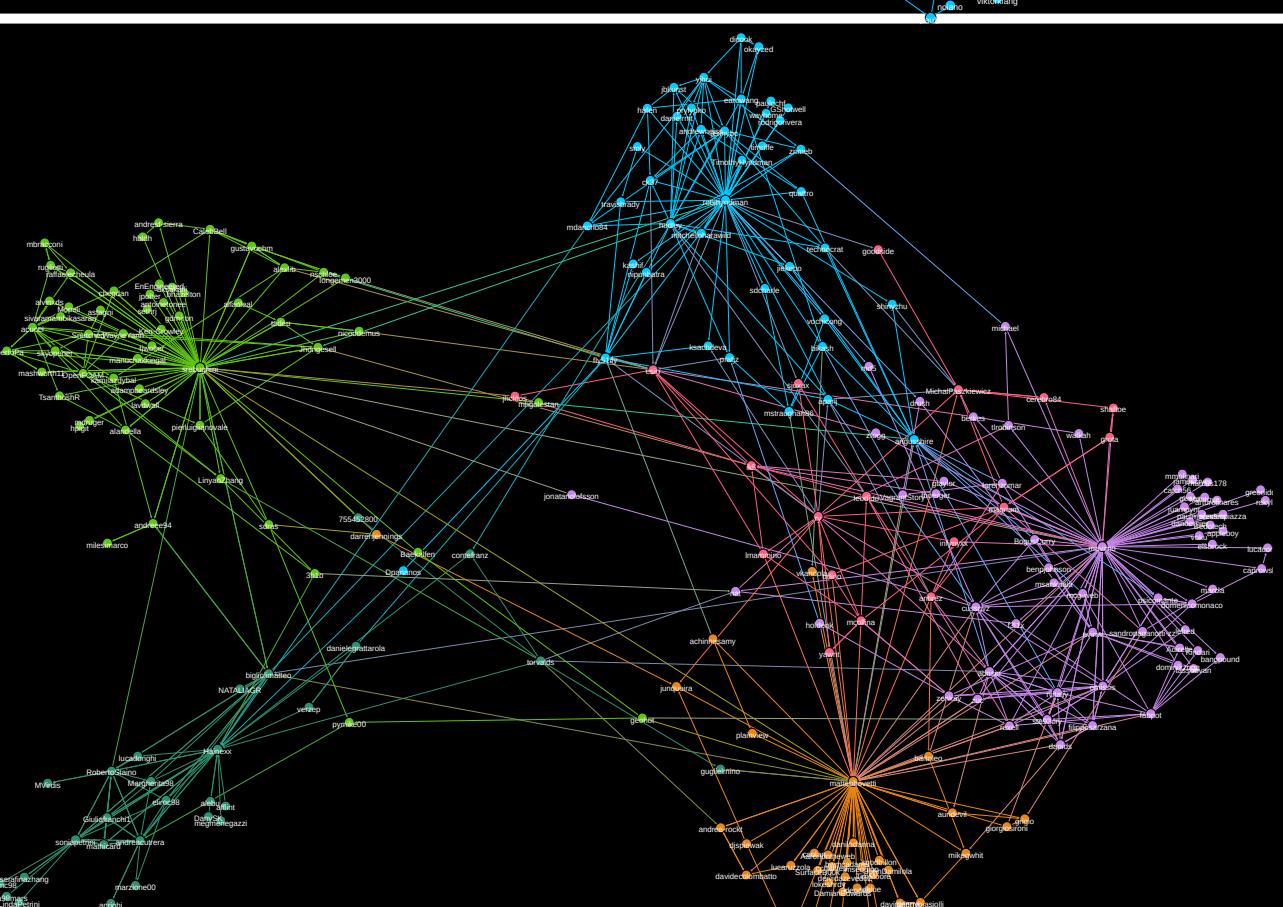
\*Based on the number of characters of a given language

# COMMUNITIES

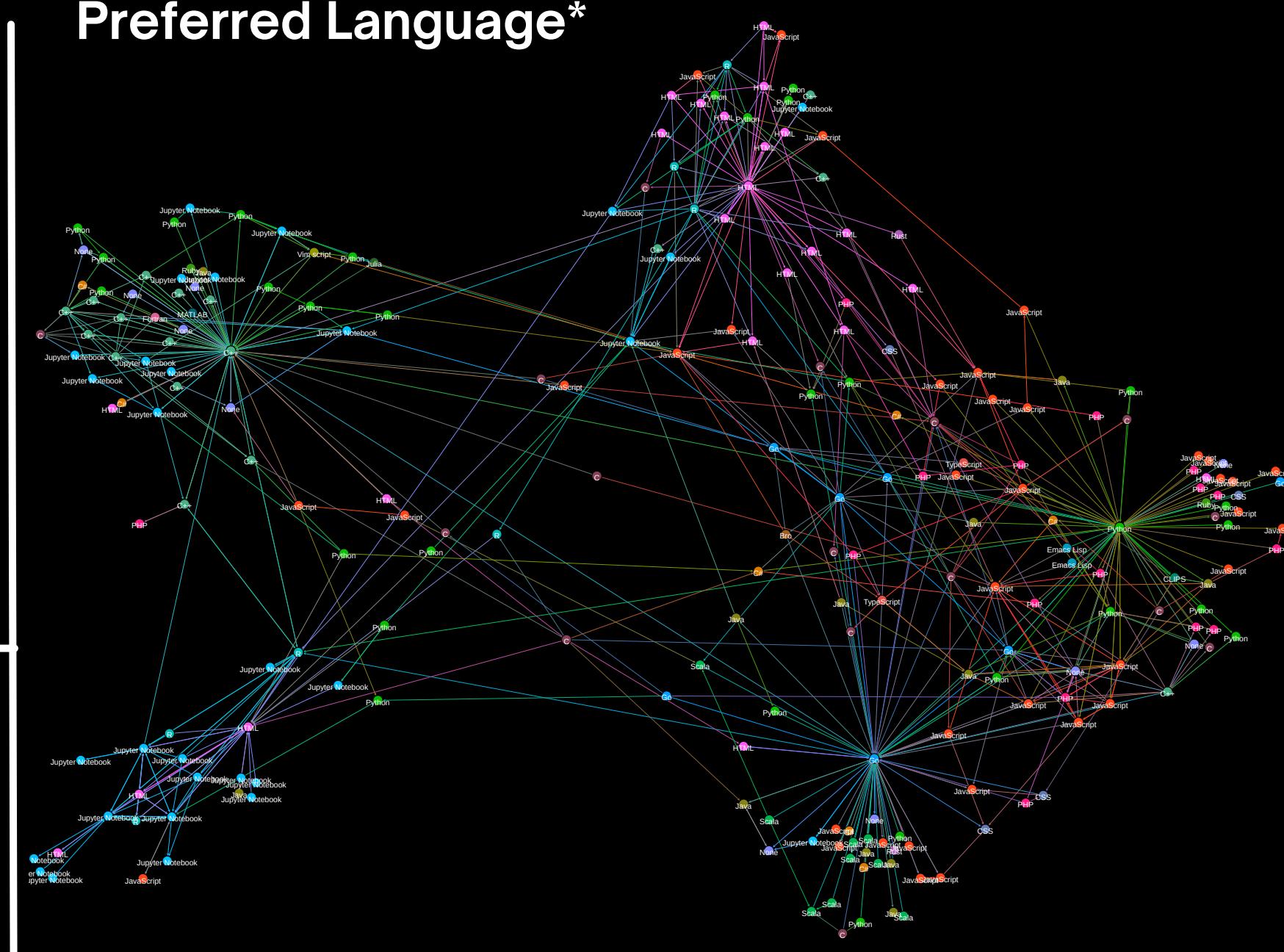
# Greedy



Louvain



## Preferred Language\*



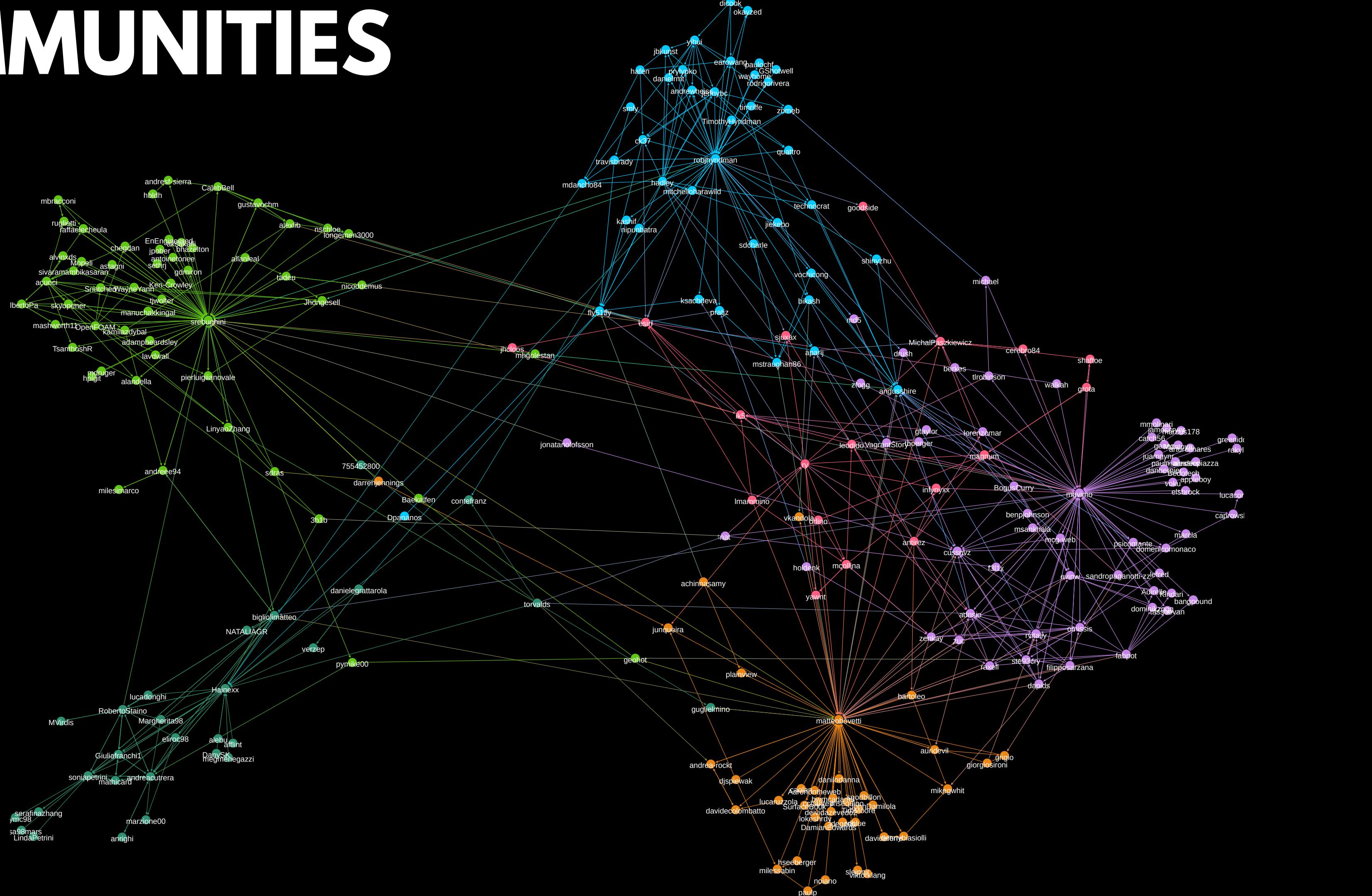
# Assortativity

**0.085**

\*Based on the number of characters of a given language

# COMMUNITIES

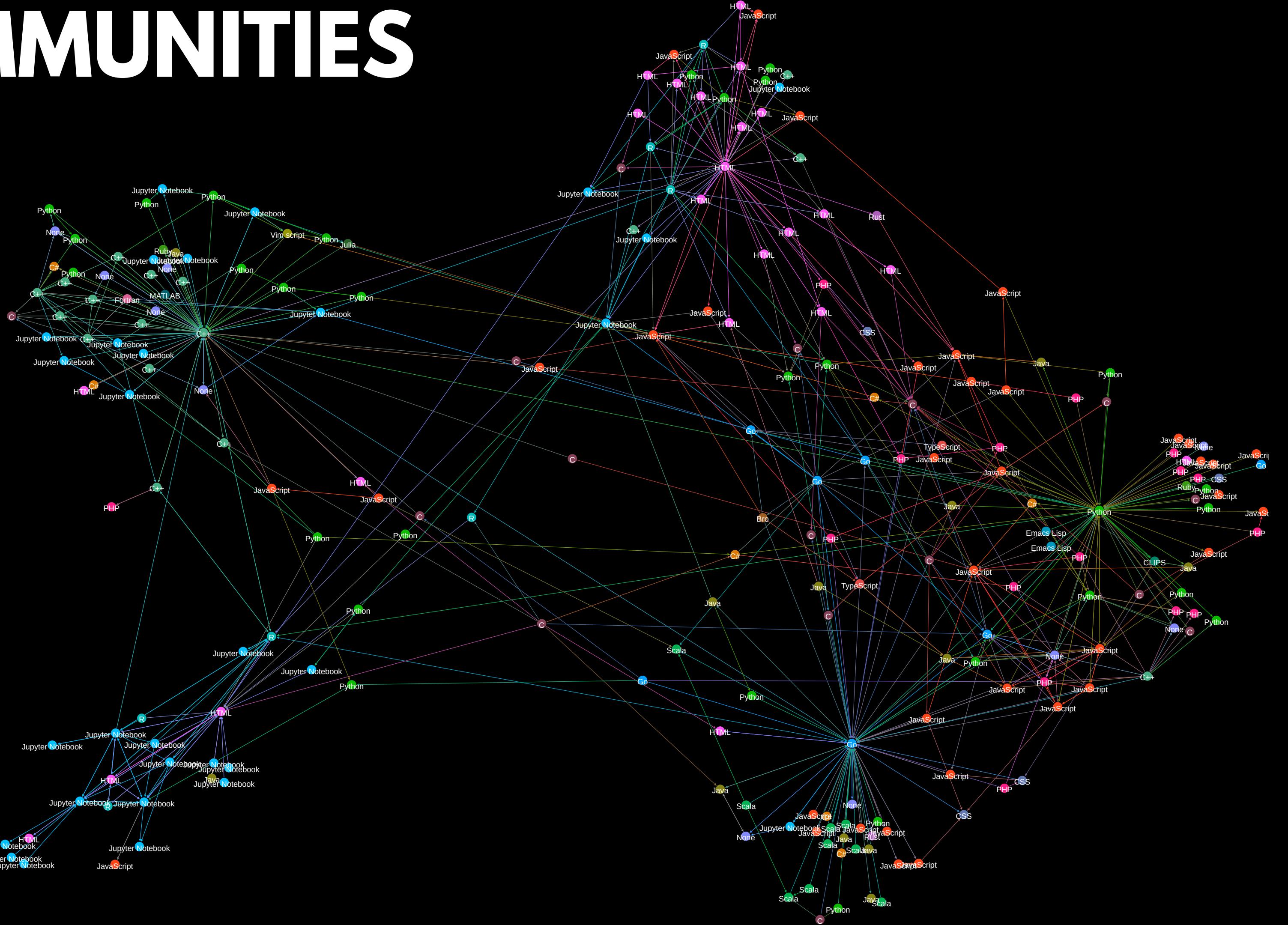
# Louvain



# COMMUNITIES

Preferred

Language\*



# COMMUNITIES

Timeout and Bans

In distance  $\leq 3$



# NEXT STEPS

- Improvements on the concept of "*preferred language*"
- Analyse a bigger portion of graph with GraphQL
- Benchmark with graph from other developers