

Evolution of climate change perception on Twitter

Focusing on Greta Thunberg Impact

Riccardo Bergamasco

Francesca Civolani

Martino De Nardi

Matteo Migliorini

Carlotta Segna

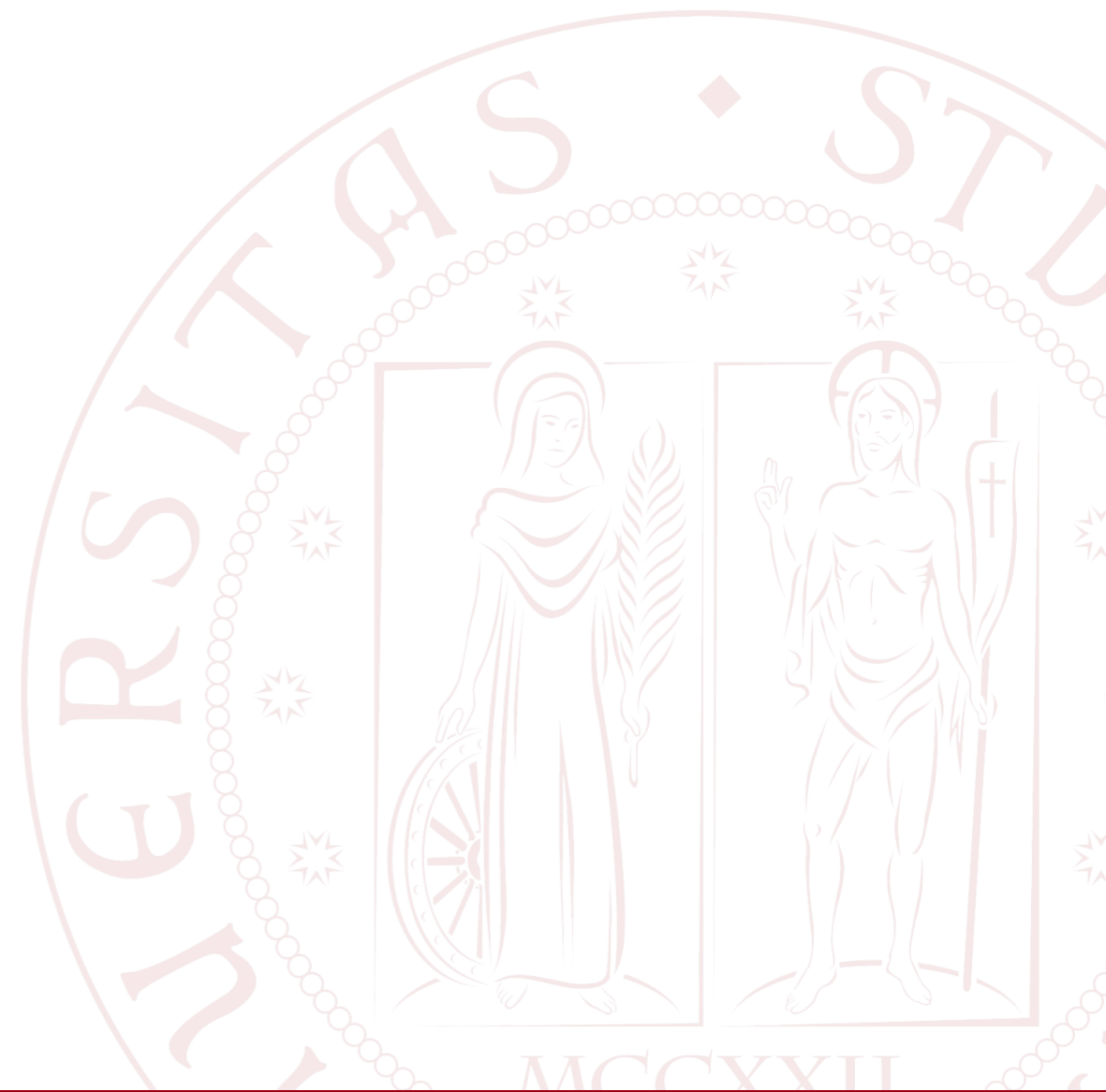
Domenico Solimini

January 31st, 2020



Overview

1. Introduction
2. Data acquisition
3. Network definition
4. Network structure
 - Parameters
 - Scale free property
 - Degree correlation
5. Community detection
6. Results and conclusions



Introduction

- In the last few years the analysis of social network has become increasingly important
- This is also valid for climate change that has received new lifeblood since it came into play Greta Thunberg
- Greta has mobilized millions of people both on social media and in the squares with the famous events called *"fridays for future"*
- For this reason we have decided to put this phenomenon under our attention



Mission: understand if and how Greta has influenced the spread of this topic on Twitter

Introduction

We have selected **two different time frames**, one before and one after the day she joined the social media




then we have created a **network made by hashtags about climate change**

Reason why we have chose to analyze Twitter:

- It is the most used social media for **political and social discussion**
- People can express their opinion concisely and quickly
- Provides a huge amount of **easily accessible data**

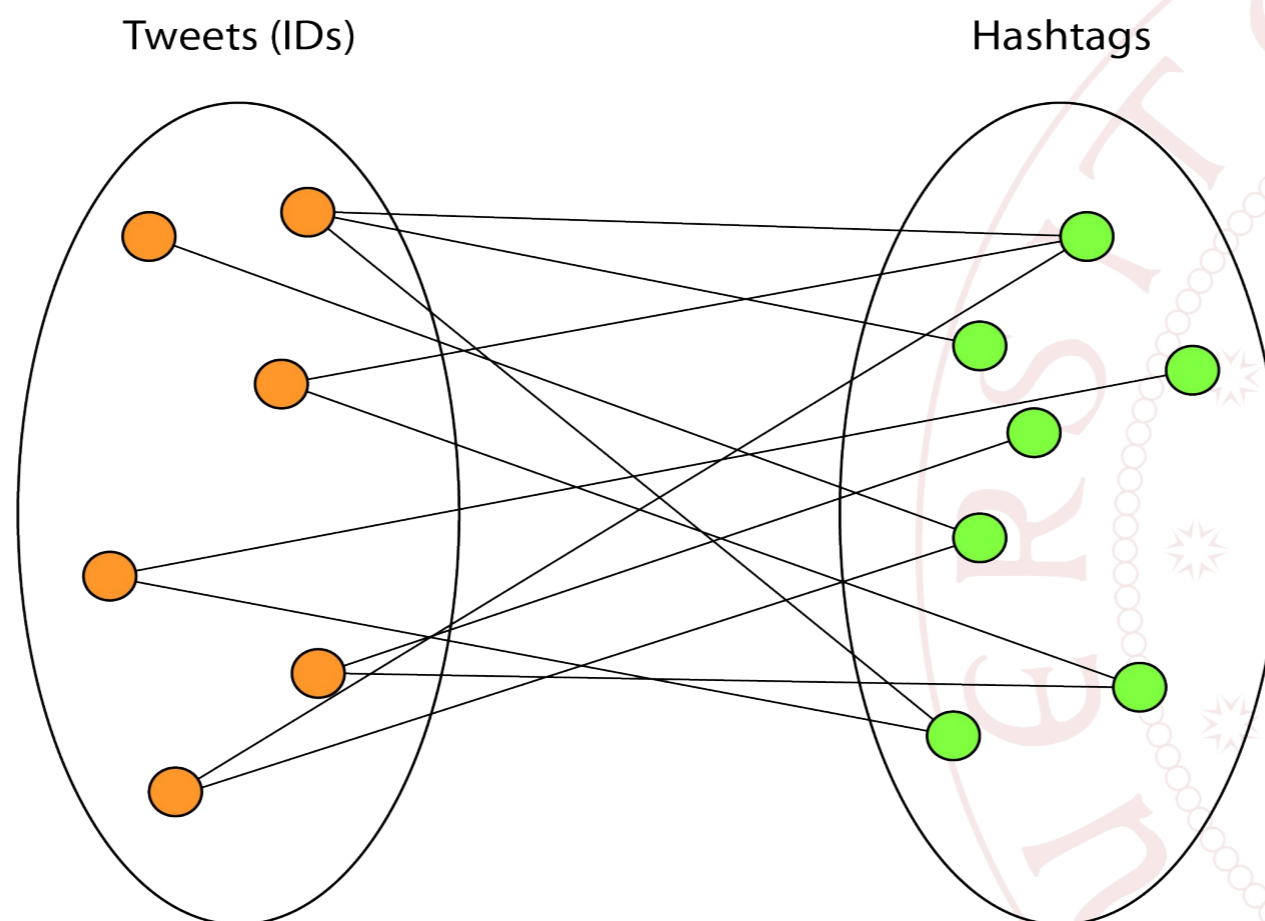
Data Acquisition

- Tweets downloaded through the Twitter API using a Developer Account 
- The two periods are kept at a **sufficient distance** to avoid noise in the transition phase in which Greta captured public attention
- 22nd April, 2018 - 31st July, 2018 and 1st March, 2019 - 9th June, 2019 (**100 days per period**)
- **10K tweets** for each period (100 tweets/day)
- Tweets captured starting from a **random hour** during the day (to get a uniform sampling)
- Search on tweets with at least one hashtag in our predefined seed hashtag list:
#climatechange, #gretathunberg, #climatestrike, #fridays4future, #climatecrisis, #parisagreement

Network Definition

Once tweets are available, the structure of the graph must be defined.

- Idea: hashtags in a tweet summarize the topic the writer is talking about
- Creation of a **bipartite graph** having tweets (identified by IDs) and hashtags as nodes



Network Definition

- A tweet is linked to all hashtags it contains
- Identifying the groups of hashtags that are observed to be in correspondence can be an effective way to identify the most general ideas spread across the net of Twitter user
- Definition of a similarity matrix using **SimRank**

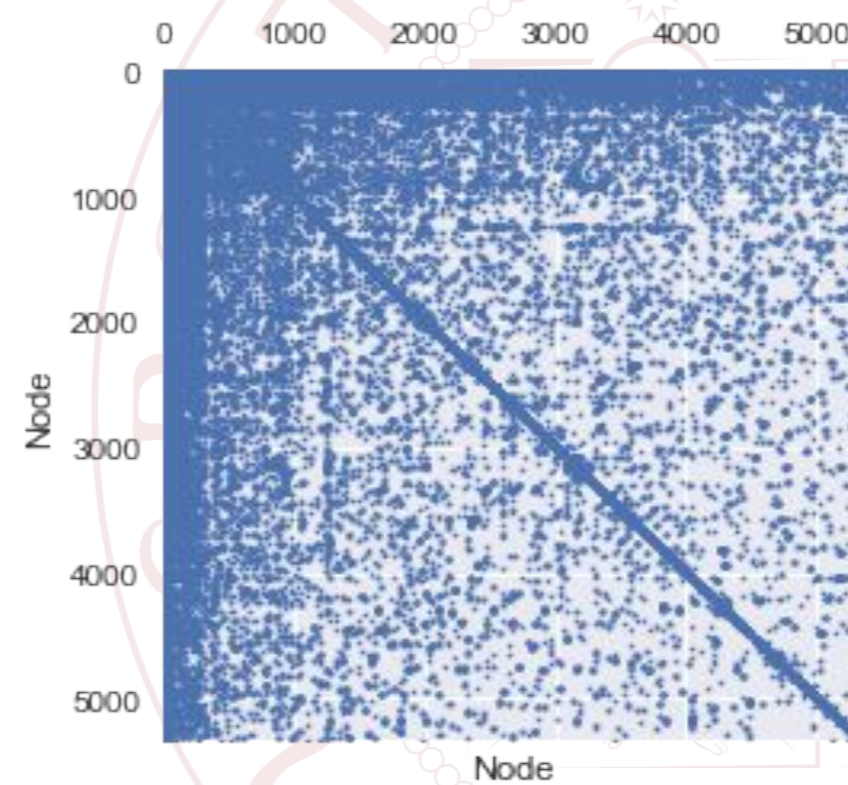
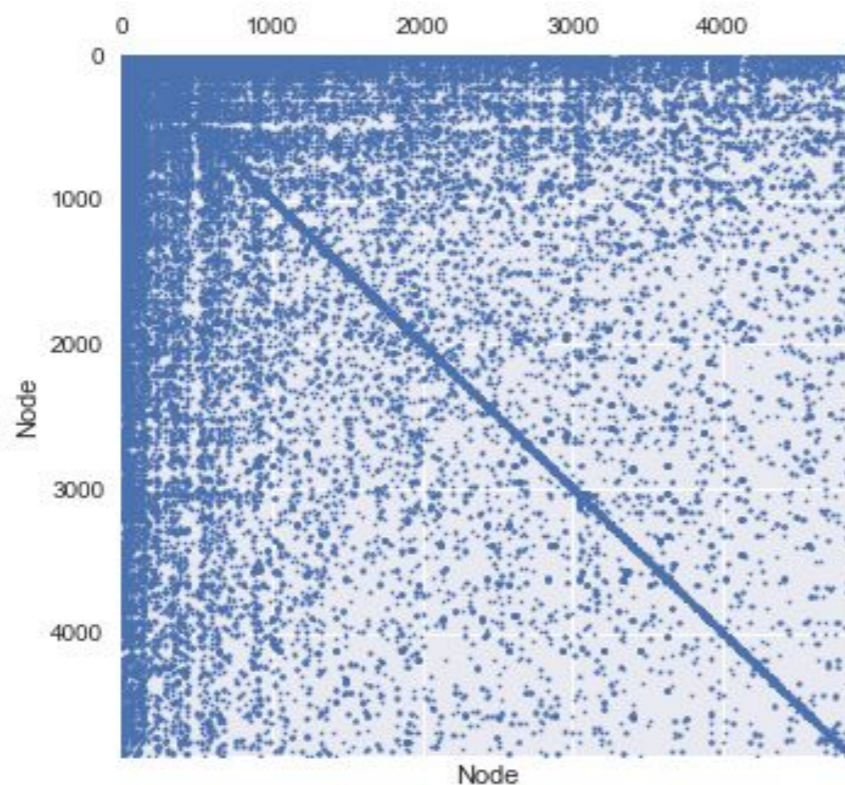
Network Definition

- Only rows and columns related to hashtags are kept (our focus!)
- Create **sparsity** setting to zero all the connections between nodes that never jointly appear in the same tweet
- Make the matrix symmetric considering arithmetic mean between the two directed connections (*SimRank* output)
- Keep only the greatest connected component
- Elimination of hashtag **#climatechange**: it present connections with the vast majority of nodes, reducing distance between almost every couple of hashtag and making distinction in communities particularly difficult.

Network Structure

In the adjacent matrix the density of rows and columns is higher at the beginning

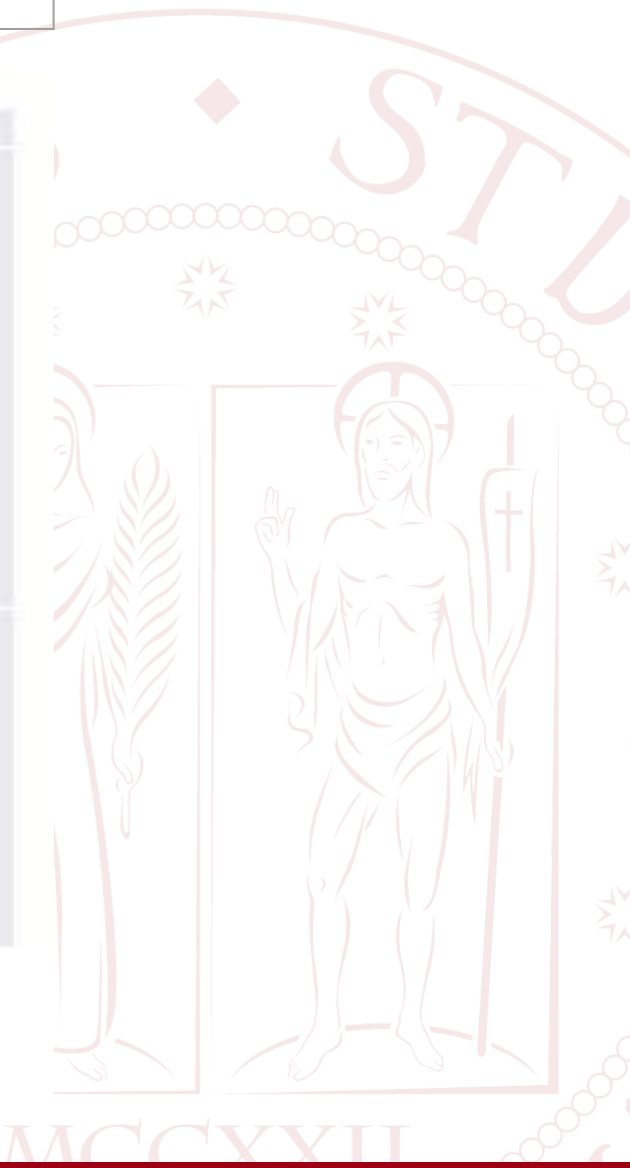
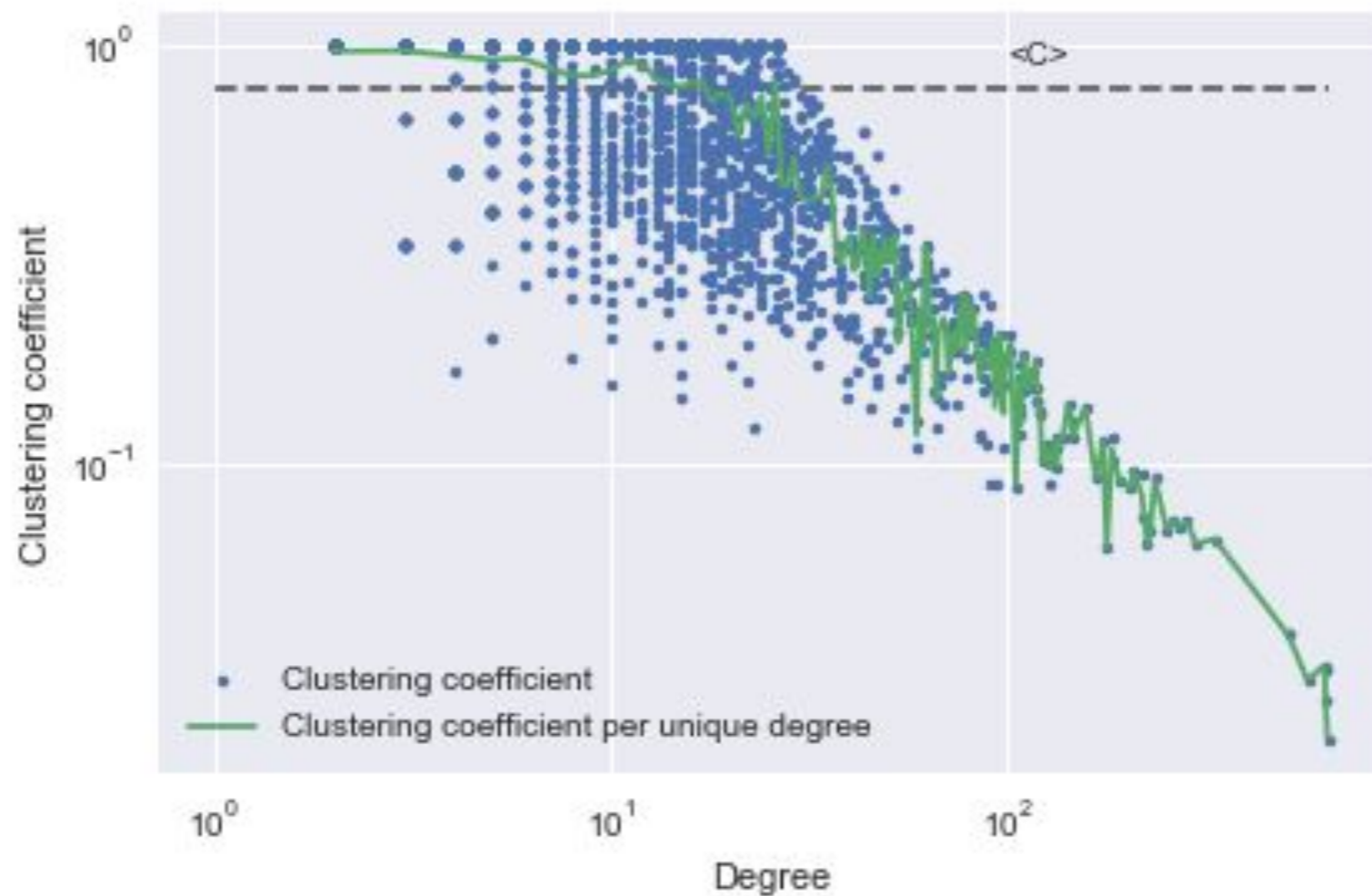
	Pre-Greta	Post-Greta
Nodes	4847	5329
Links	32246	33021
Avg Degree	13.306	12.392
Shortest Path Distance	3.222	3.177
Diameter	9	8



Clustering Coefficient

High value of the clustering coefficient -> high number of share edges

	Pre-Greta	Post-Greta
Clustering Coefficient	0.787	0.798



Robustness

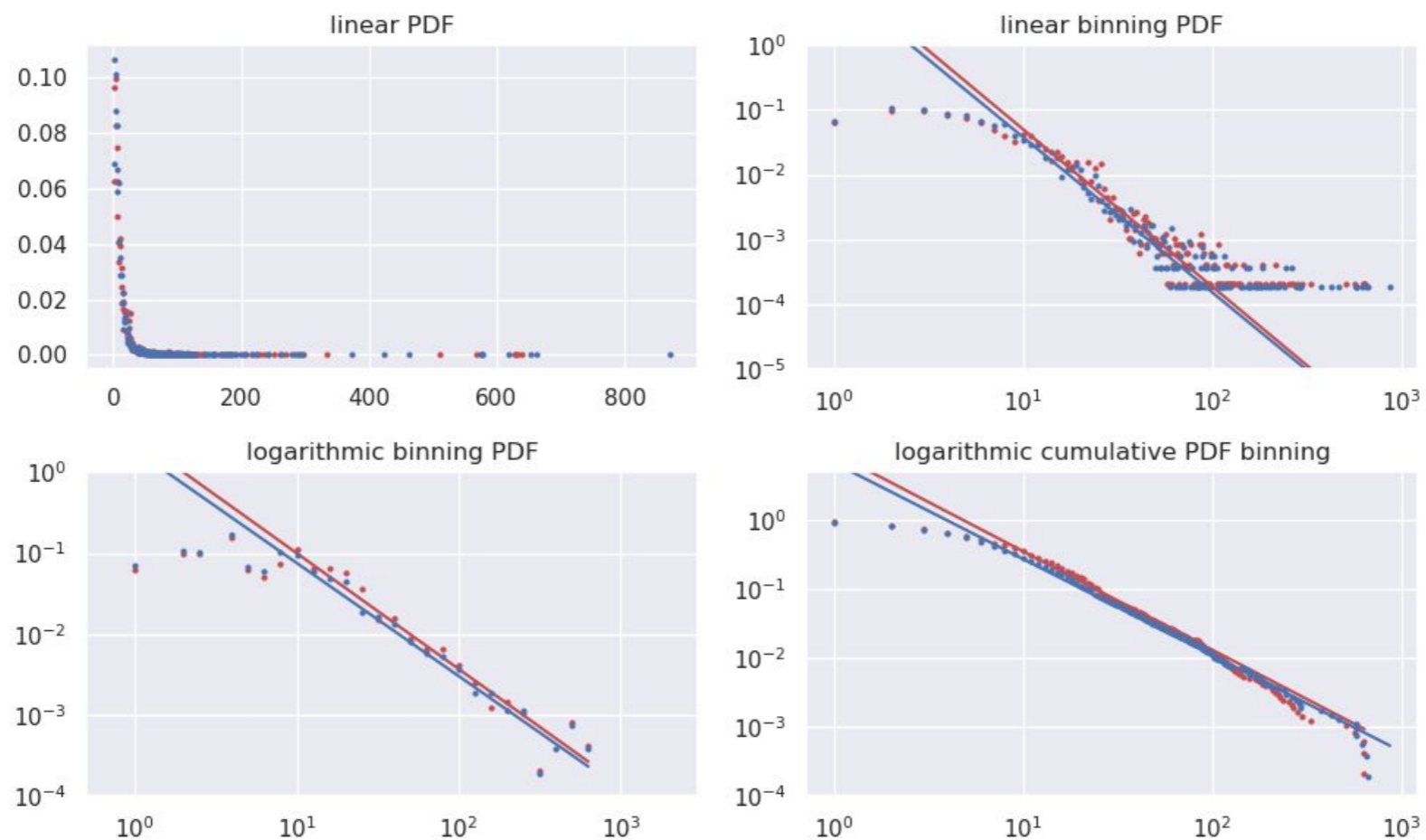
- The value of the robustness suggests that there are giant components.
- The value of the breaking point suggests that the network is hard to break.
 - About 90% of the nodes are necessary to be removed to break it apart

	Pre-Greta	Post-Greta
Robustness	13.340	12.418
Breaking Point	0.919	0.912

Scale Free Property

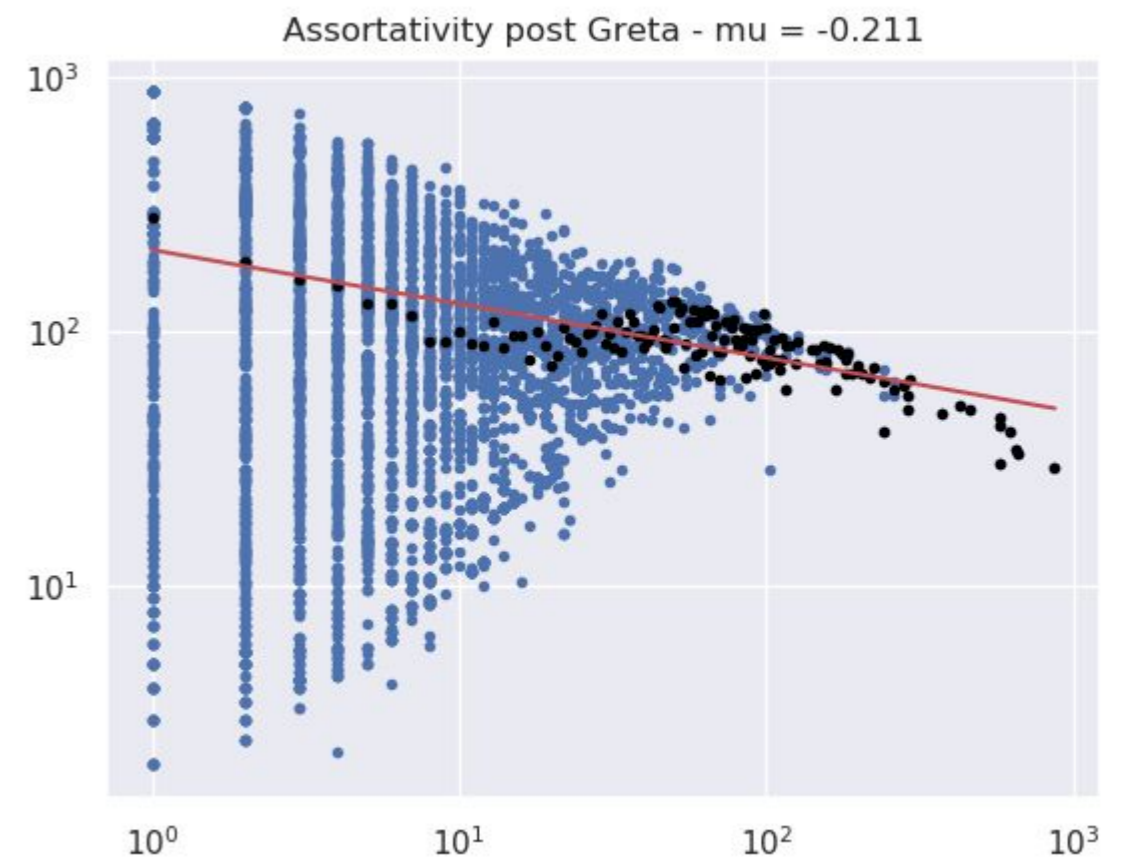
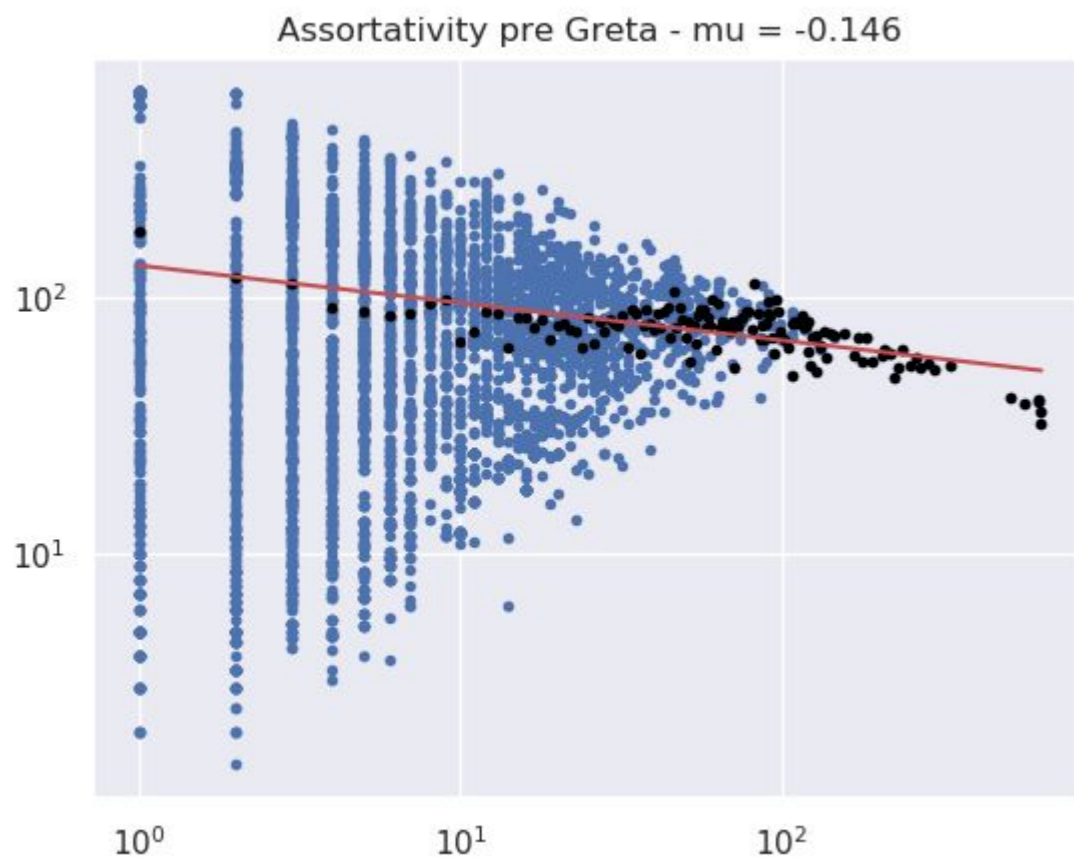
Twitter hashtag network is reasonably generated by a preferential attachment growth. We expect our subnet to be self-consistent.

The net structure is the one expected. Both pre (red) and post (blue) nets are weakly scale free with γ respectively equal to **2.42** and **2.39**.



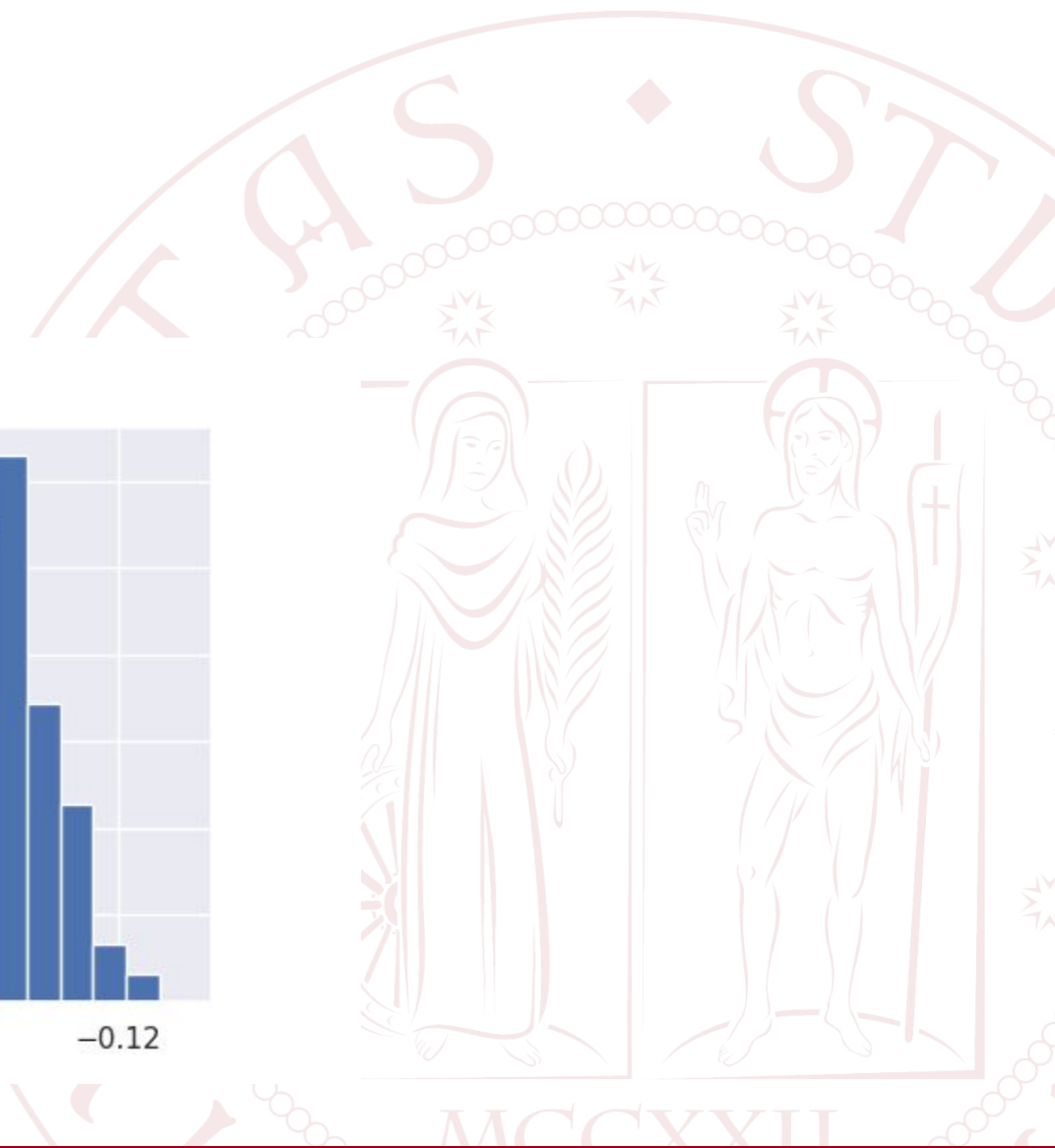
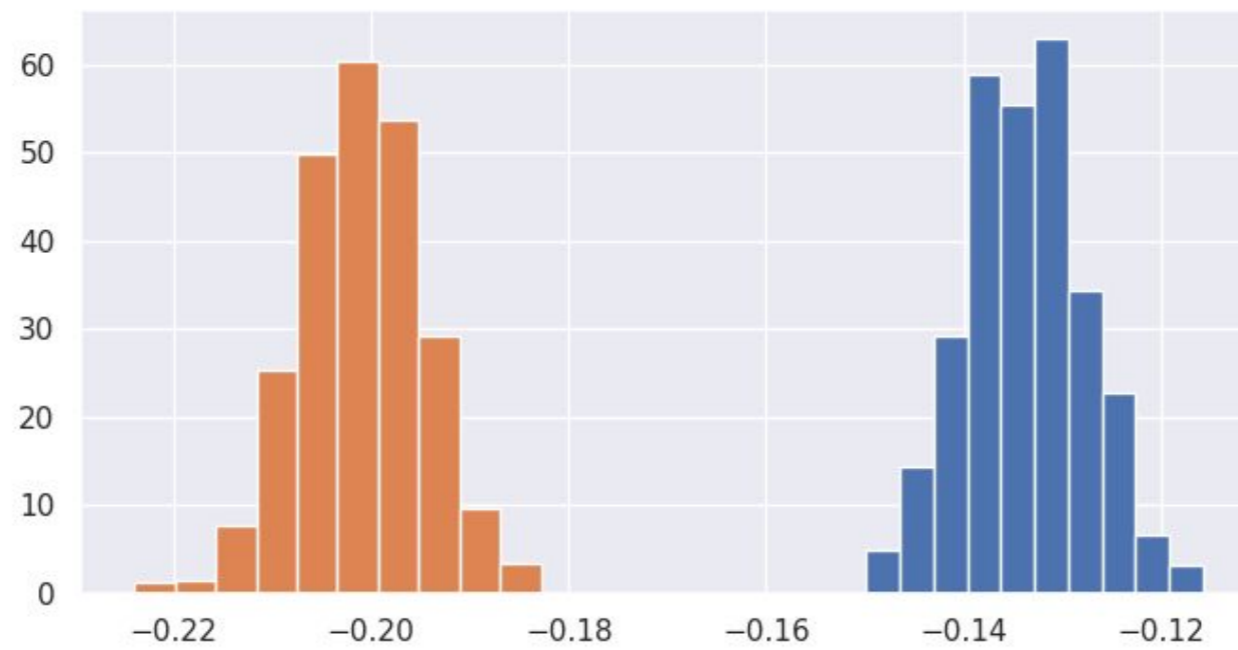
Assortativity

Assortativity coefficient is negative but it is not very informative. The value of k_s is much lower than k_{max} .



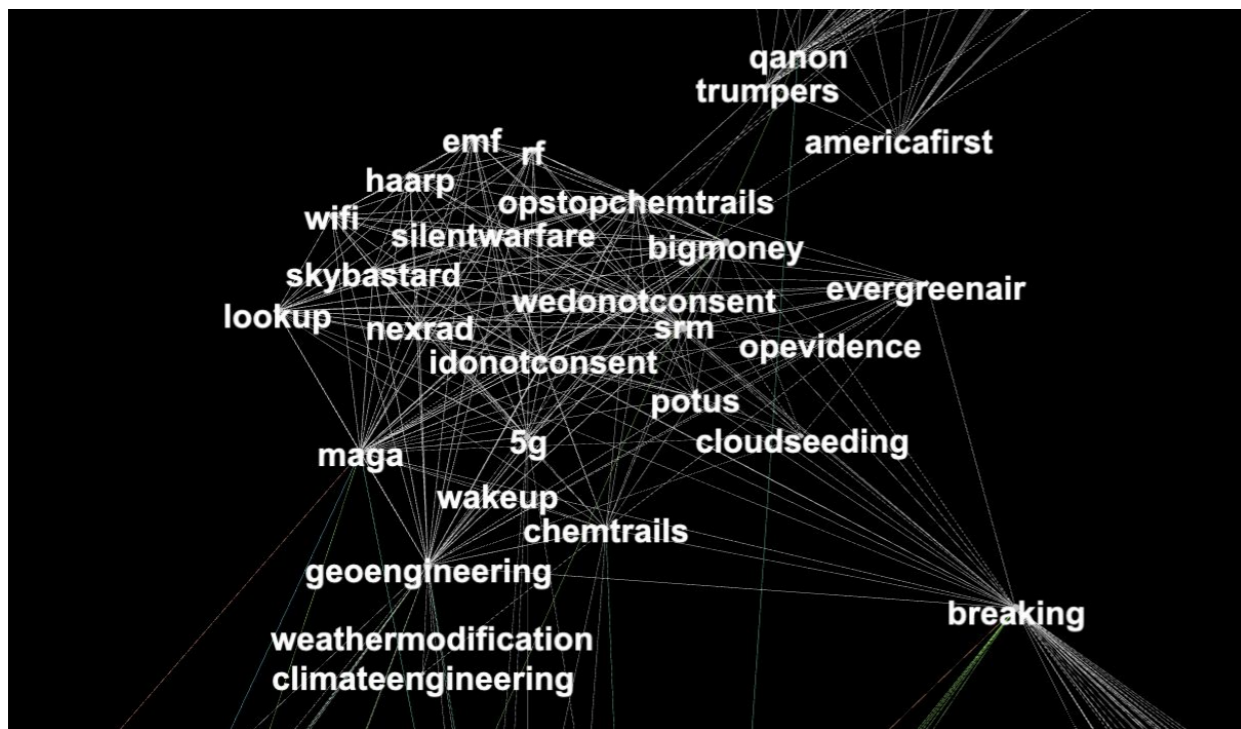
Assortativity

text



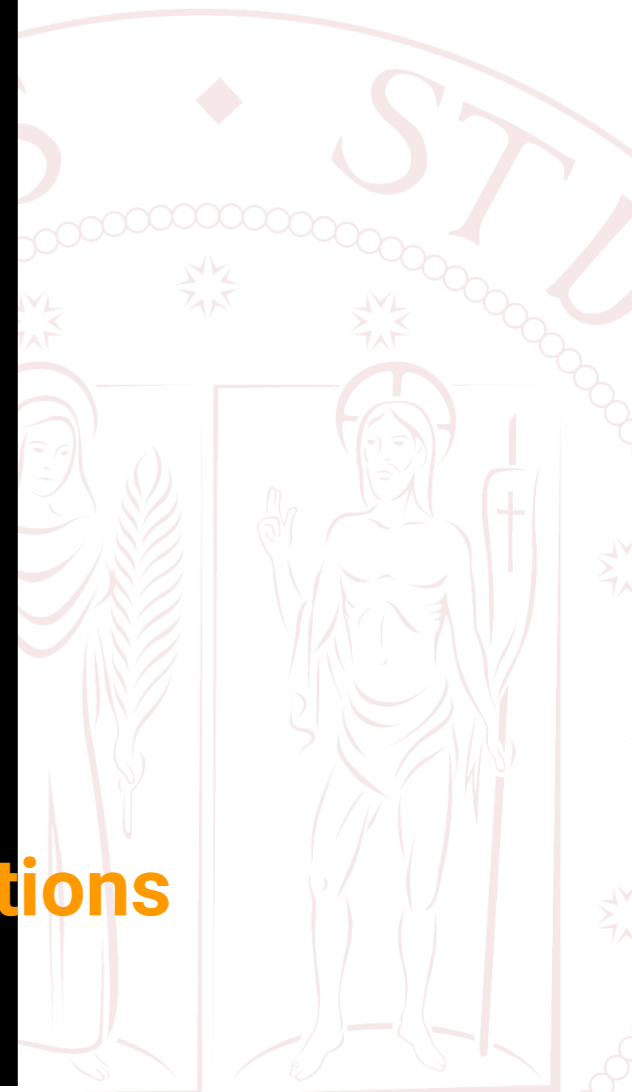
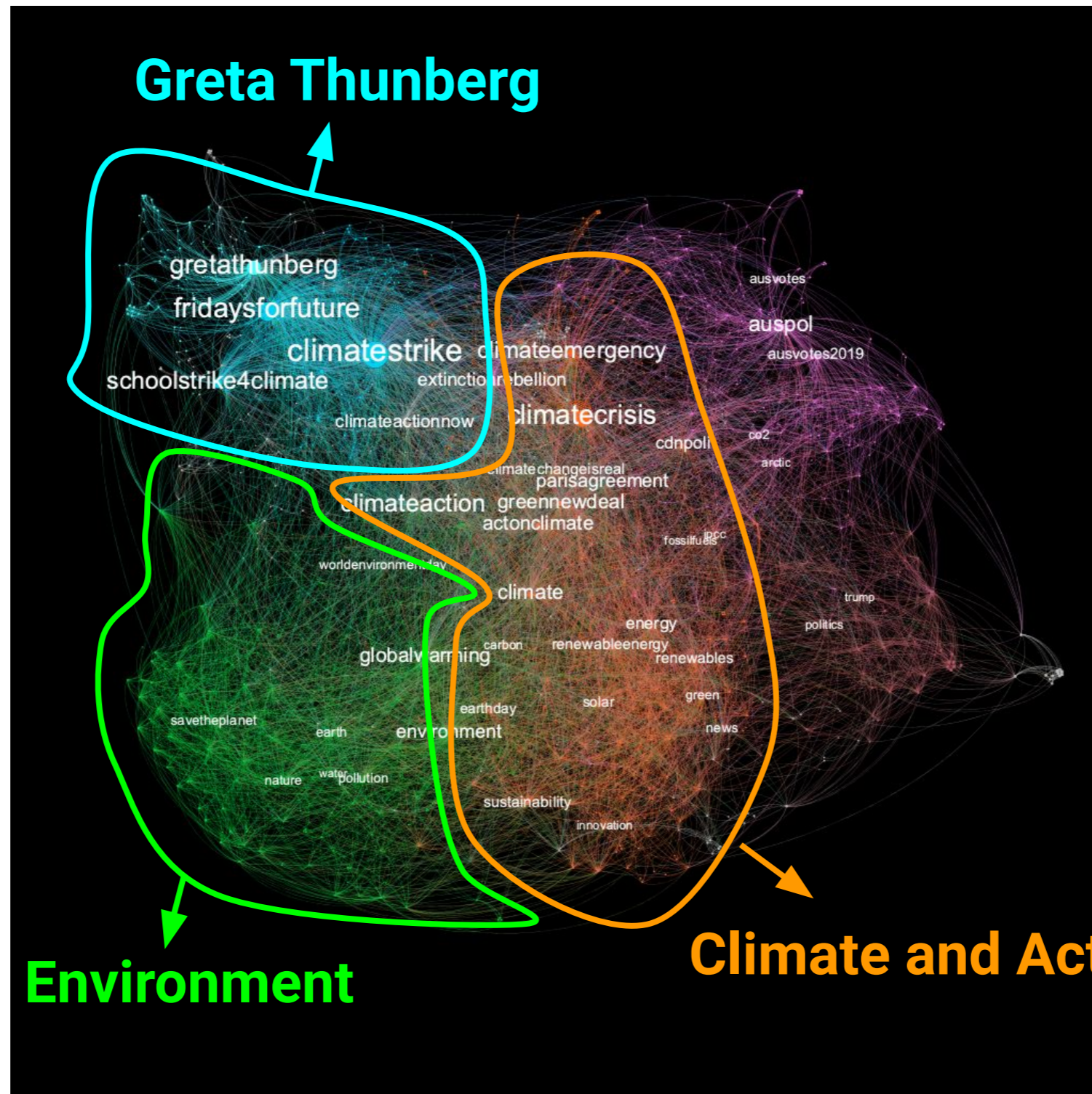
Community detection

- Louvain algorithm
- Found between 20 and 30 communities
 - Most of them are small and very specific communities



- Focus on **larger communities** and analyze the most central hashtags
- Identified the three most important communities for each network
- Visualization:
 - 5 communities displayed
 - Removed cliques and nodes with small degree

Community detection: Post-Greta



Results & Conclusions

GRETA THUNBERG COMMUNITY

The general features of this new community are:

- High-impact communication
- A change in people's behavior

Greta Thunberg use a communication based on new, short and strong sentences that become **incisive hashtags** (e.g. #climatestrike, #fridaysforfutue, etc.).

Results & Conclusions

GRETA THUNBERG COMMUNITY

The use of new strong words like “strike”, “emergency”, “act now” and “crisis” means a change in people’s behavior.

Now people are **more involved** in climate change debate. They start to protest in accord to Greta in an effort to ask to people with power to take measures in time.

So:

Yes, Greta Thunberg **has changed** the climate change debate on Twitter.

Results & Conclusions

Few more considerations:

- #climatechangeisreal hashtag dropped in terms of popularity.
- In both networks there is a **small Australian community** debating the topic in relation to their country.
- In both networks there is the permanence, even if minimal, of **Donald Trump**, the climate change denier for excellence.

Roles

aggiungete il vostro nome affianco al punto su cui avete lavorato

- *Tweet collection*: Matteo, Martino
- *Network creation*: Matteo, Martino
- *Network structure*
 - *Initial analysis of the network* (avg degree, shortest path distance, diameter, ...): Carlotta
 - *Clustering*: Carlotta
 - *Robustness*: Carlotta
 - *Degree Correlation*: Domenico, Carlotta
 - sublist di vari topic per carlotta e domenico?
- *Community detection*: Matteo

