



UNIVERSITÀ
DEGLI STUDI
DI PADOVA

Reddit Network Analysis

subreddits interactions over the period
2016-2017

Anthony Palmieri, 2038503

Network Science,
A.A. 2022/2023



Overview

In this presentation you'll find

1 The Network

2 Degree Distribution

3 Centrality measures

4 Community detection

5 Robustness

6 Link prediction

7 Conclusions



- Inferring the distribution
- Clustering Coeffs
- Assortativity

- Degree
- Pagerank
- HITS

- Louvain
- Authority Shifting

The Network

The network describes interactions among subreddits during the years 2016/2017.

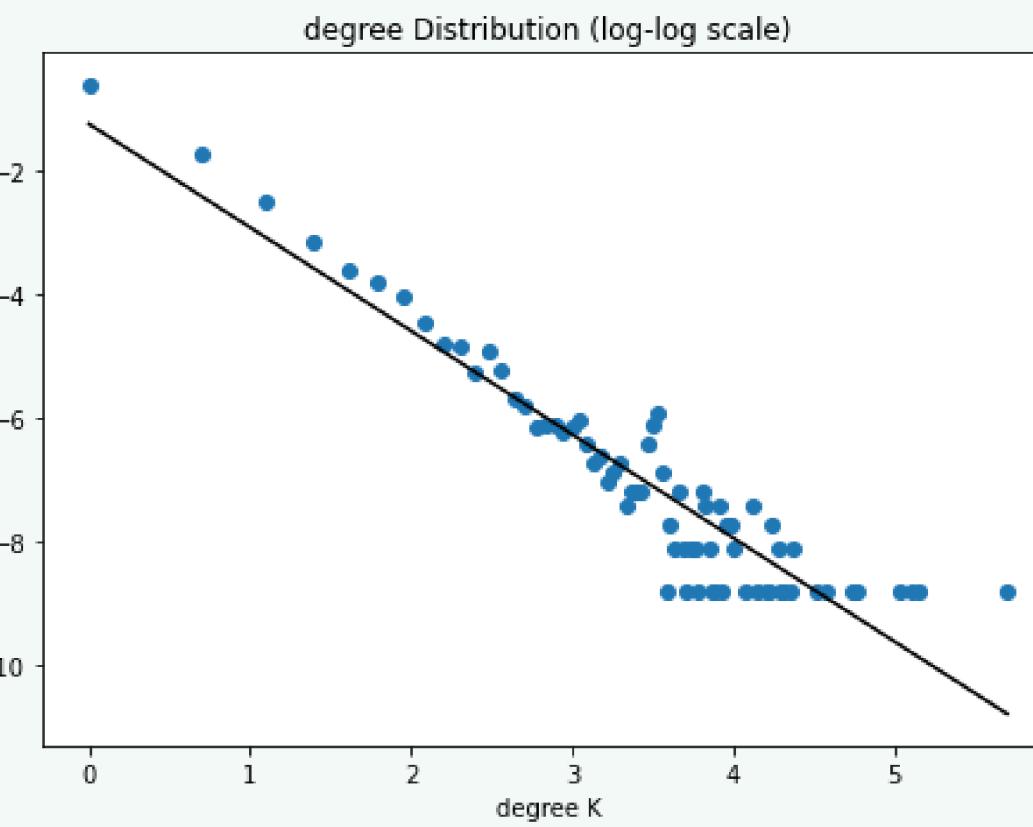
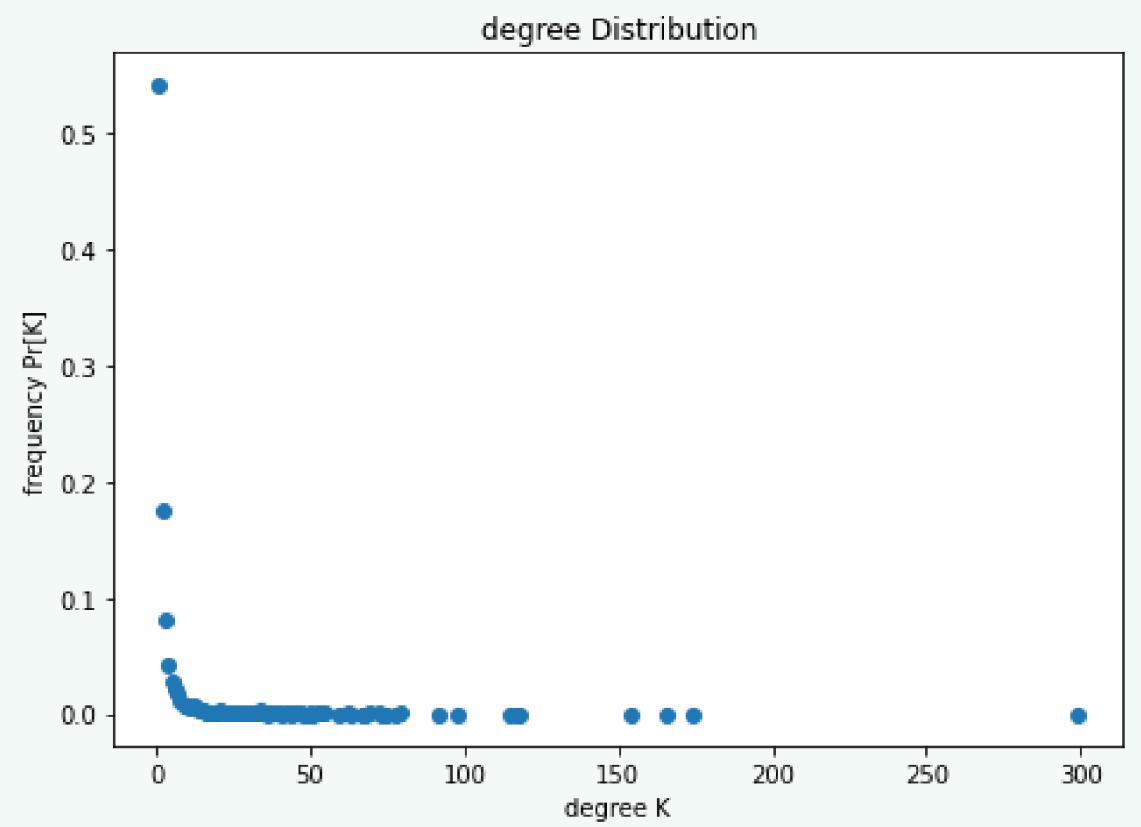
Nodes are subreddits and edges are hyperlinks between subreddits.

An edge exists between u and v if there is at least a post in u that links to v and vice versa.

The weight of the edge is the minimum between the number of mutually exchanged links.

- undirected Simple Graph
- weighted
- # nodes: 7969
- # edges: 13179
- # C.C: 511
- # nodes G.C: 6666
- # edges G.C: 12355

Degree Distribution

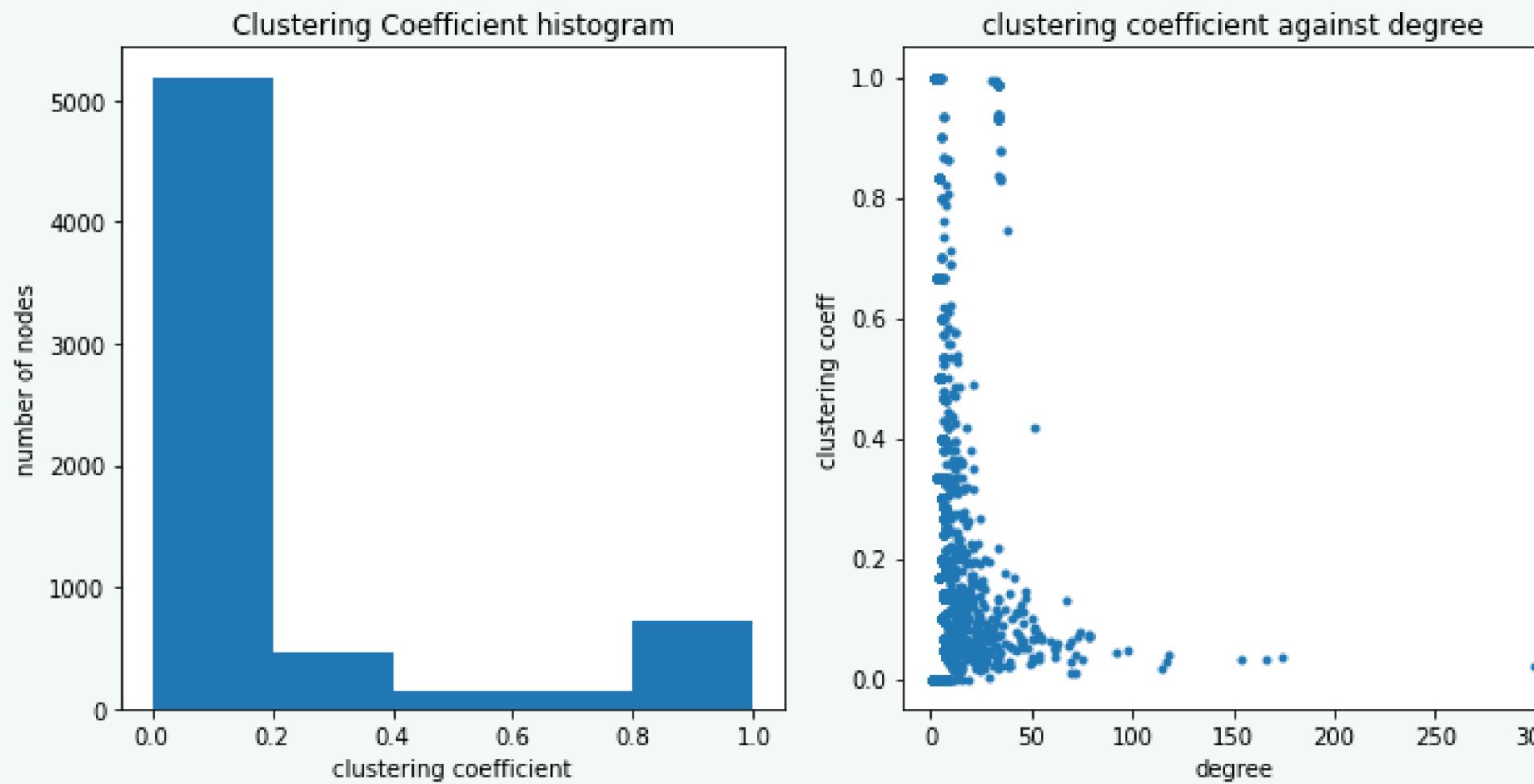


- Power-Law distribution!
- Using CCDF we estimate $\gamma \sim 2.52$
- Scale-Free regime
- $\langle K \rangle = 5.55$ (theoretical)
- $\langle K \rangle = 3.71$ (empirical)

Caution:

Degree 1 nodes make up 54% of the entire network! Problematic for Link prediction...

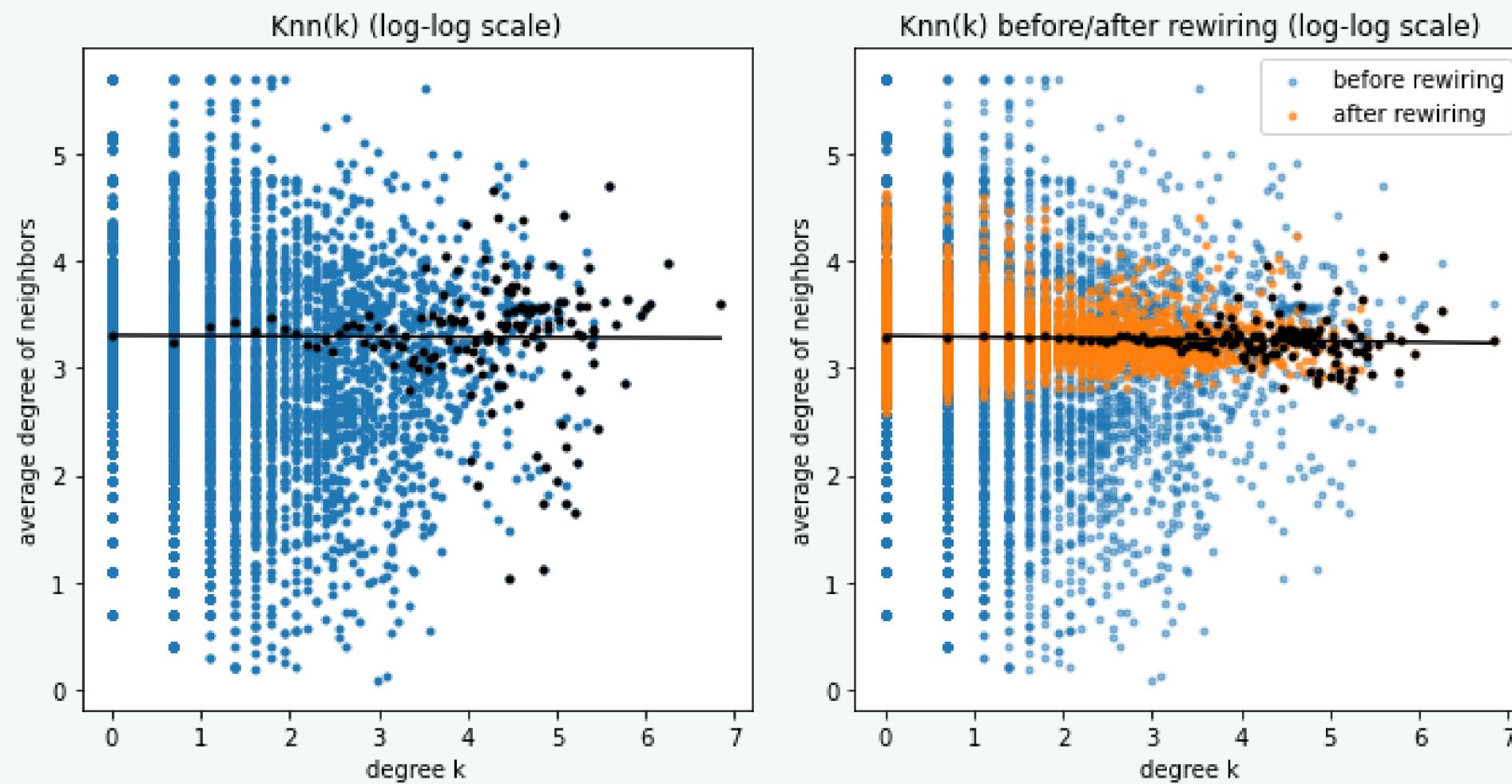
Clustering Coeffs



- **Global Average Clustering Coefficient ~ 0.16**
- **Graph sparse also at local level!**
- **However 9.8% of nodes has a clustering coeff. = 1**

The nodes with clustering coeff. =1 span a variety of topics, ranging from sports, games, trading, politics and more

Assortativity



- Global degree assortativity ~ -0.03
- Ties tend to be random
- Also confirmed by Average Neighbour Degree function
- Knn behaviour remains stable after 100 random edges rewiring

The plots show the behaviour of $K_{nn}(k)$ before and after random edge rewiring

Degree Centrality



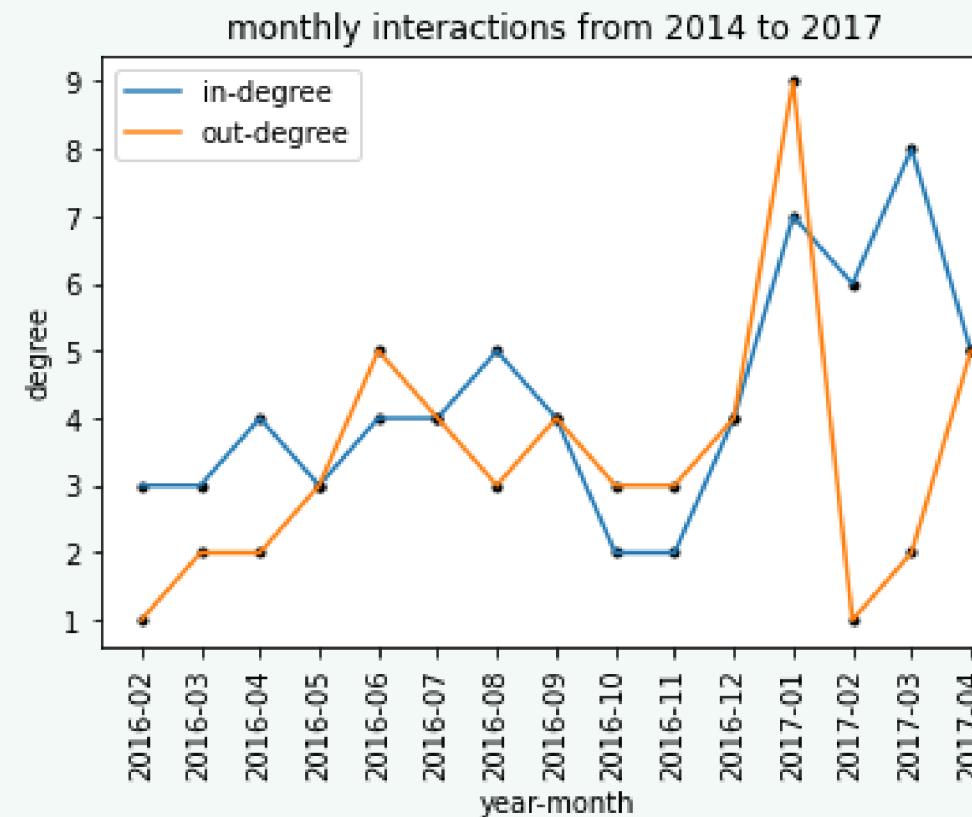
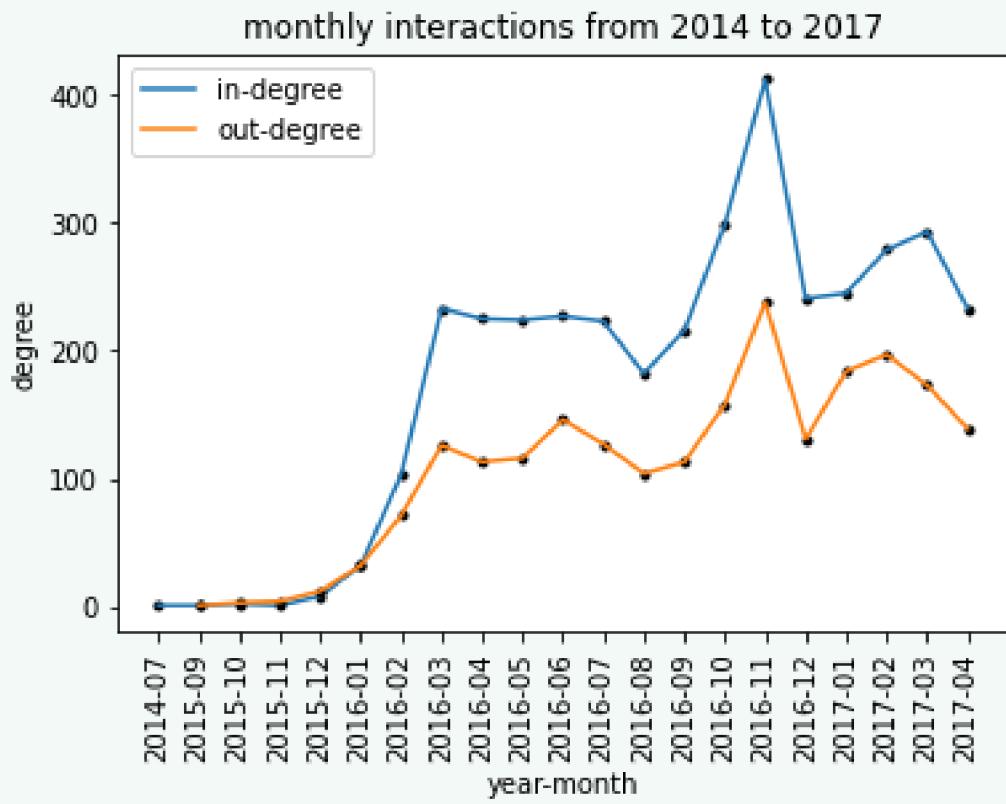
2016 U.S. presidential election seems to play a role here...

Is there any correlation between *the_donald* and *conspiracy*?

Highest ranked nodes:

- ***the_donald* (299)**
- ***conspiracy* (174)**
- ***gaming* (166)**
- ***subredditdrama* (154)**
- ***SandersForPresident* (118)**
- ***iama* (117)**
- ***LeagueOfLegends* (115)**

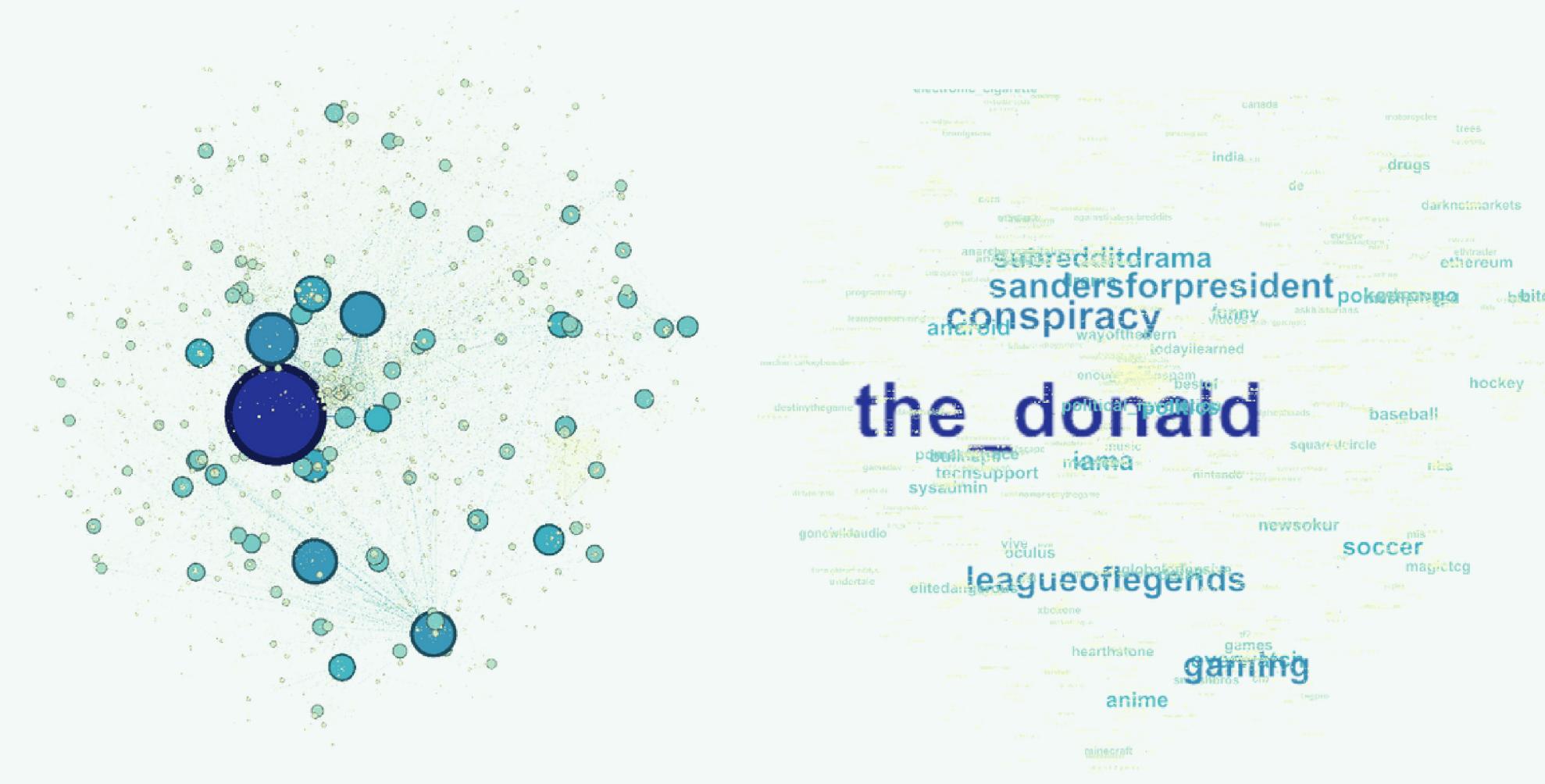
Degree Centrality



- On the left: *the_donald* monthly in/out-degree from 2014 to 2017
- On the right: monthly exchanged hyperlinks between *the_donald* and *conspiracy* from 2014 to 2017
- monthly exchanges begin to spike in January 2017, when Trump officially became POTUS

Trump has a large base of supporters known for their conservative political views and support to various conspiracy theories. The network seems to reflect the climate of unrest that surrounded Trump's election and presidency

PageRank Centrality

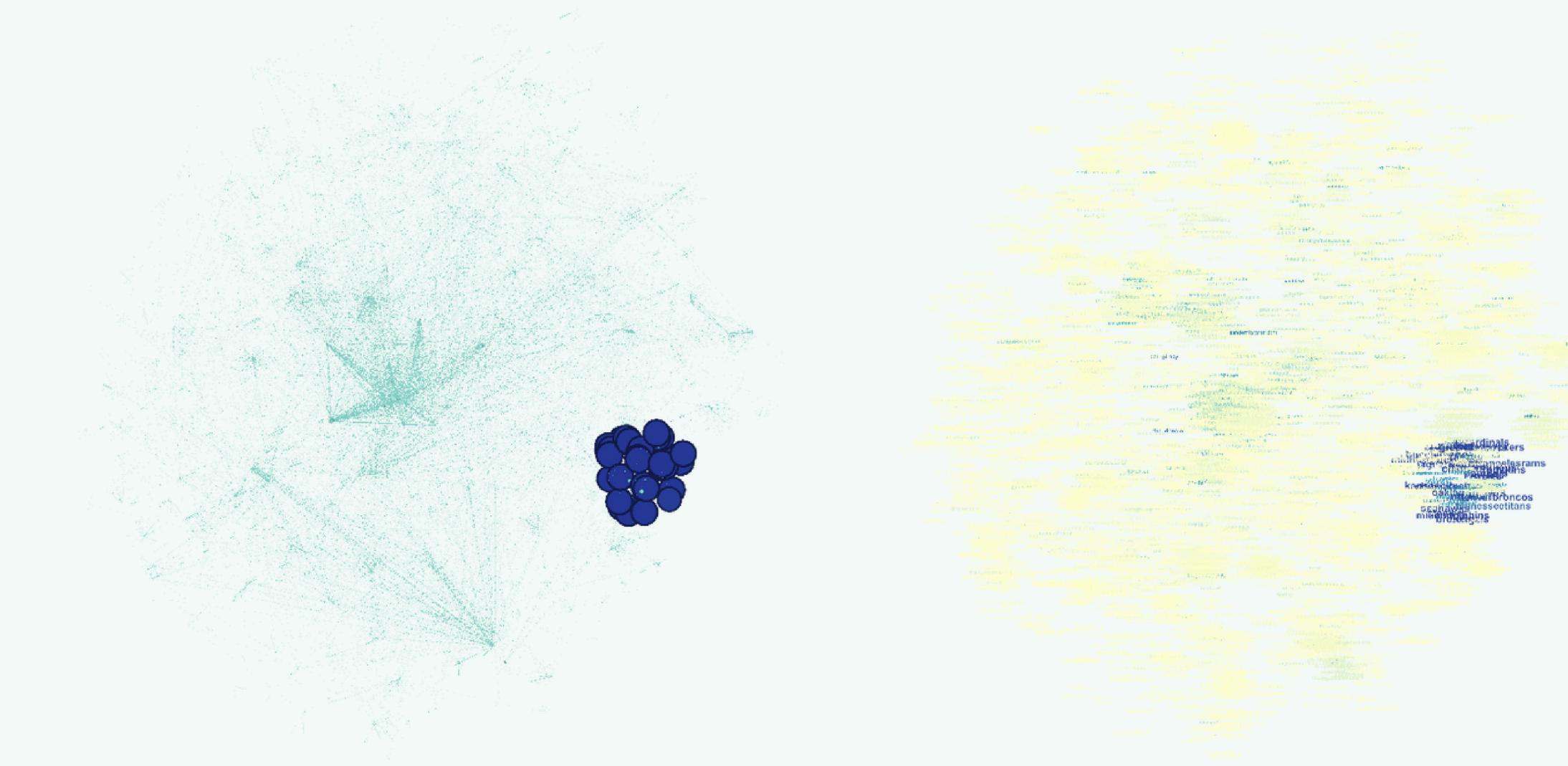


Highest ranked nodes:

- *the_donald*
 - *conspiracy*
 - *gaming*
 - *subredditdrama*
 - *SandersForPresident*
 - *iama*
 - *LeagueOfLegends*

Pagerank results are very similar to degree centrality, but we already expected this!

HITS Centrality

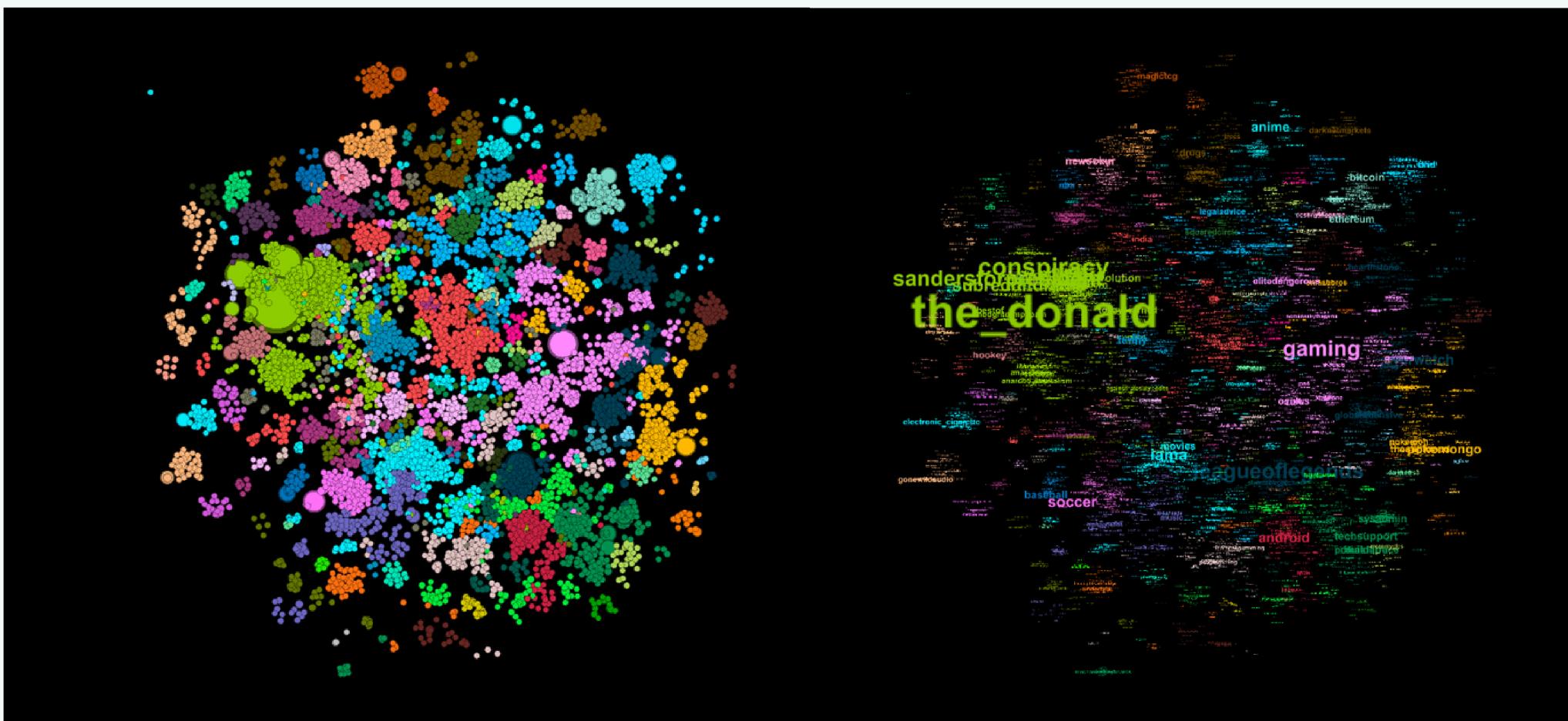


- results very different from degree and pagerank
- HITS assigns small score to all nodes except for a small group
- This small group revolves around NFL teams

HITS scores are interesting and unexpected...

Could they be due to the high average clustering coefficient and average degree of the group?

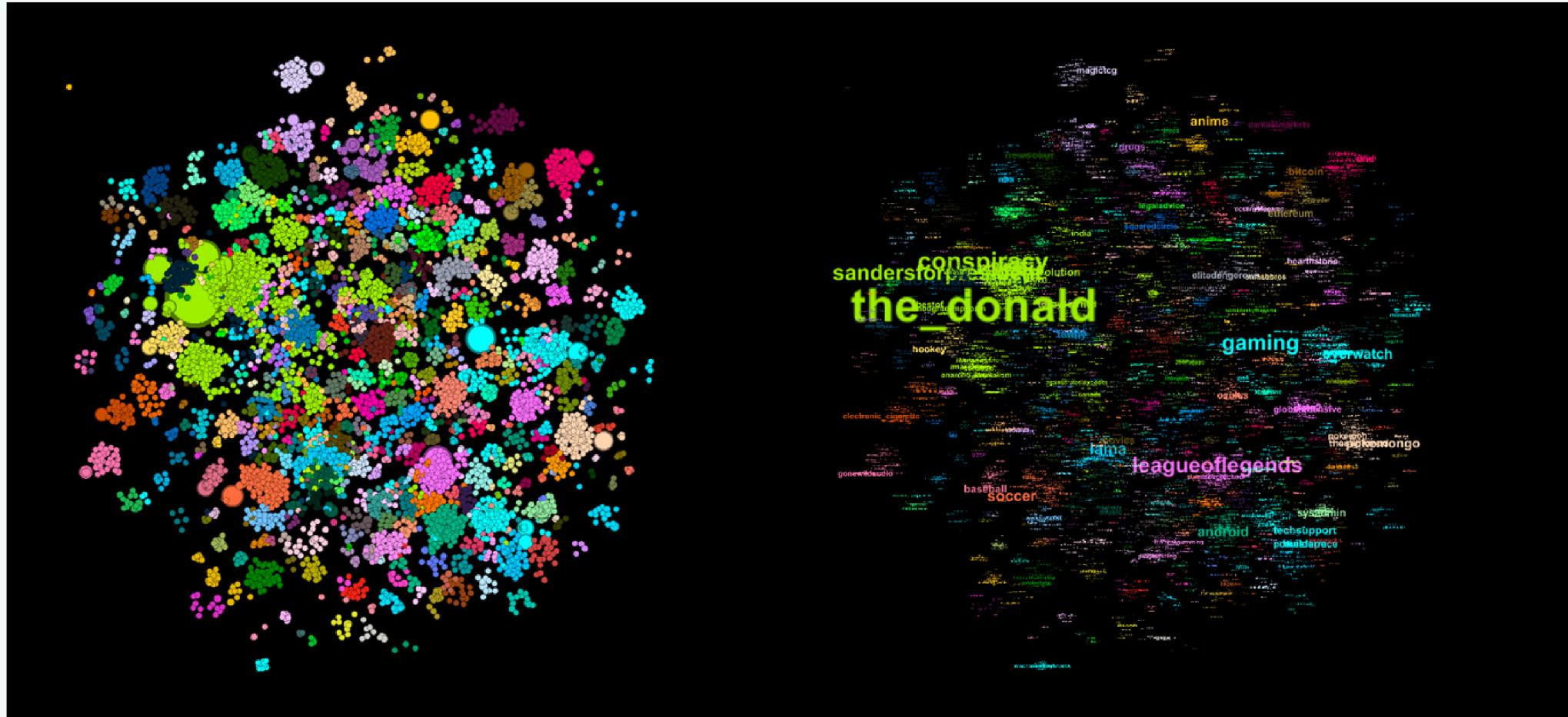
Louvain



- 57 communities
- modularity ~ 0.84
- largest community 856 nodes
- *the_donald* dominates largest community
- other major communities: gaming, soccer, smartphones, tech support, cryptocurrency, anime and so on

The communities seem to be consistent and coherent! The largest community, although containing many political subreddits, also spans several other topics, including some of the most popular subreddits in general

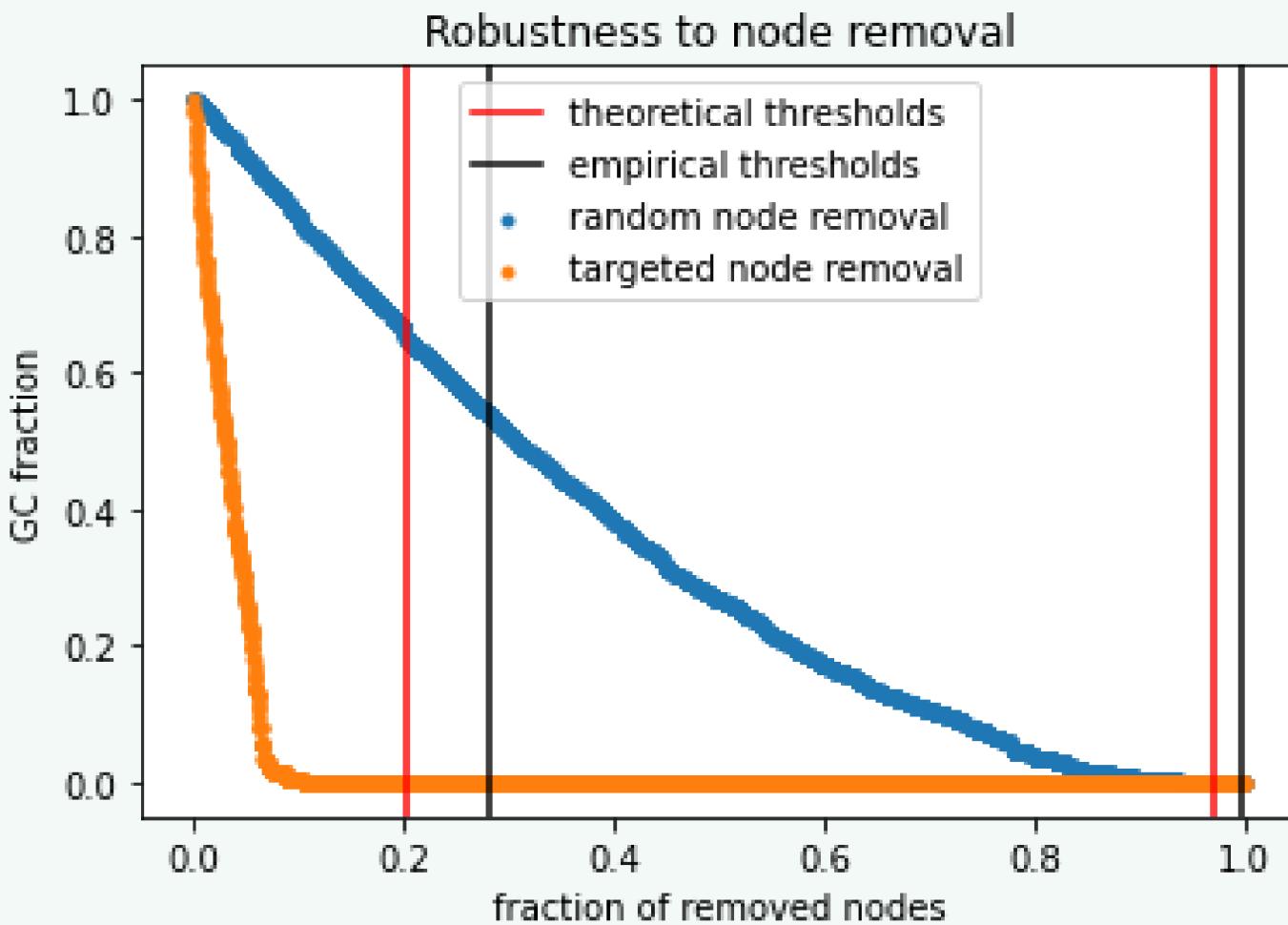
Authority Shifting



- 422 communities
- modularity ~ 0.79
- largest community 840 nodes
- *the_donald* dominates largest community
- other major communities: gaming, soccer, smartphones, tech support, cryptocurrency, anime and so on, just as before
- communities are now extremely topic-centric

The communities seem again to be consistent and coherent. The biggest difference with Louvain is that now many communities are extremely topic-centric, and even highly related topics produce different communities

Robustness



Using *Molloy-Reed criterion* we can compute critical thresholds both for random and targeted node removal

| | theoretical | empirical |
|---------------------|-------------|-----------|
| fc random removal | 0.97 | 0.996 |
| fc targeted removal | 0.20 | 0.28 |

Network is robust against random node removal, but significantly vulnerable to targeted attacks, consistent with the underlying power-law distribution

Link Prediction

Algorithms:

- Jaccard Common Neighbours
- Resource Allocation
- Adamic Adar
- Local Paths
- Random Walk with Restart

Activation threshold selection:

Train-Validation-Test splitting of the edges to select optimal threshold for link activation

Best model:

- RA
- TPR ~ 71% on Test graph
- FPR ~ 19% on Test graph
- best model even with different seeds
- extremely low precision

All the methods suffer of very low precision, probably due to sparsity of the graph and abundance of degree 1 nodes

Conclusions

To summarize, we found:

- power-law degree distribution, pointing to a scale-free regime
- network locally sparse and randomly tied (neutral)
- measures of centrality highlight the impact of 2016 U.S. presidential election, with *the_donald* and *conspiracy* strongly correlated
- community detection yields coherent communities spanning very different topics, from politics to gaming, sports and so on
- network robust to random removal but weak to targeted attacks
- link prediction results not so satisfactory, probably due to sparsity and high number of nodes with just one link

Thanks for the attention!