

FUNDAMENTALS OF ONLINE ((SOCIAL) MEDIA)

NETWORK ANALYSIS

Lecture 1

WHO IS THIS GUY?

WHY ARE YOU HERE?

Who are you?

Why did you choose this course?

What are your expectations for this day?

THE PLAN

1. Online (Social) Media Network Fundamentals
2. Network fundamentals
3. [Pause]
4. Data Mining Possibilities and Difficulties
5. Network Analysis Methods

Afterwards:

Practical on Data Collection and Exploratory Analysis with Descriptive Statistics in Python

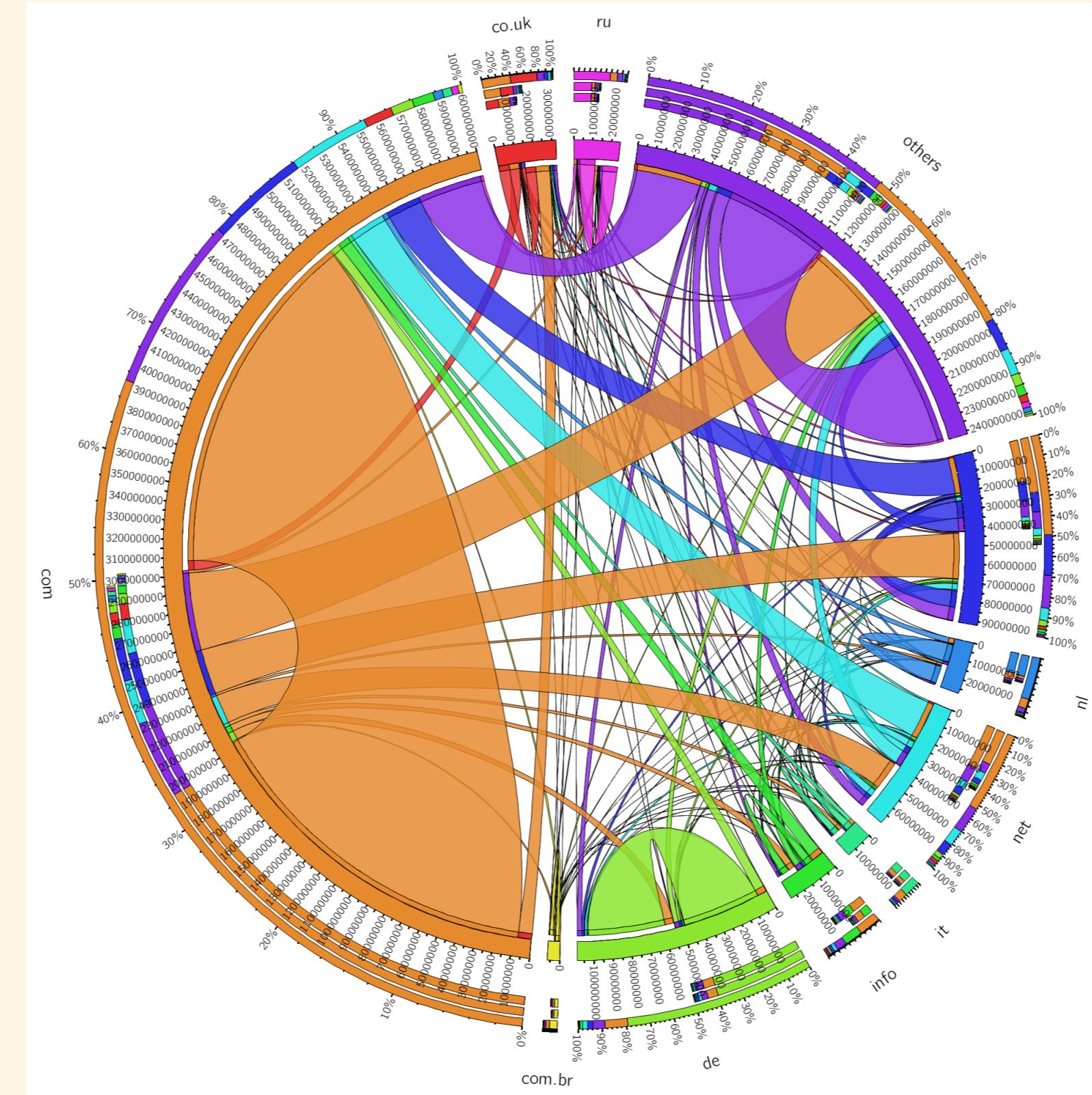
ONLINE ((SOCIAL) MEDIA) NETWORKS

NOT TECHNICAL INFRASTRUCTURE NETWORKS

The Internet: 1997 - 2021 Contextual



MOSTLY NOT HYPERLINK NETWORKS



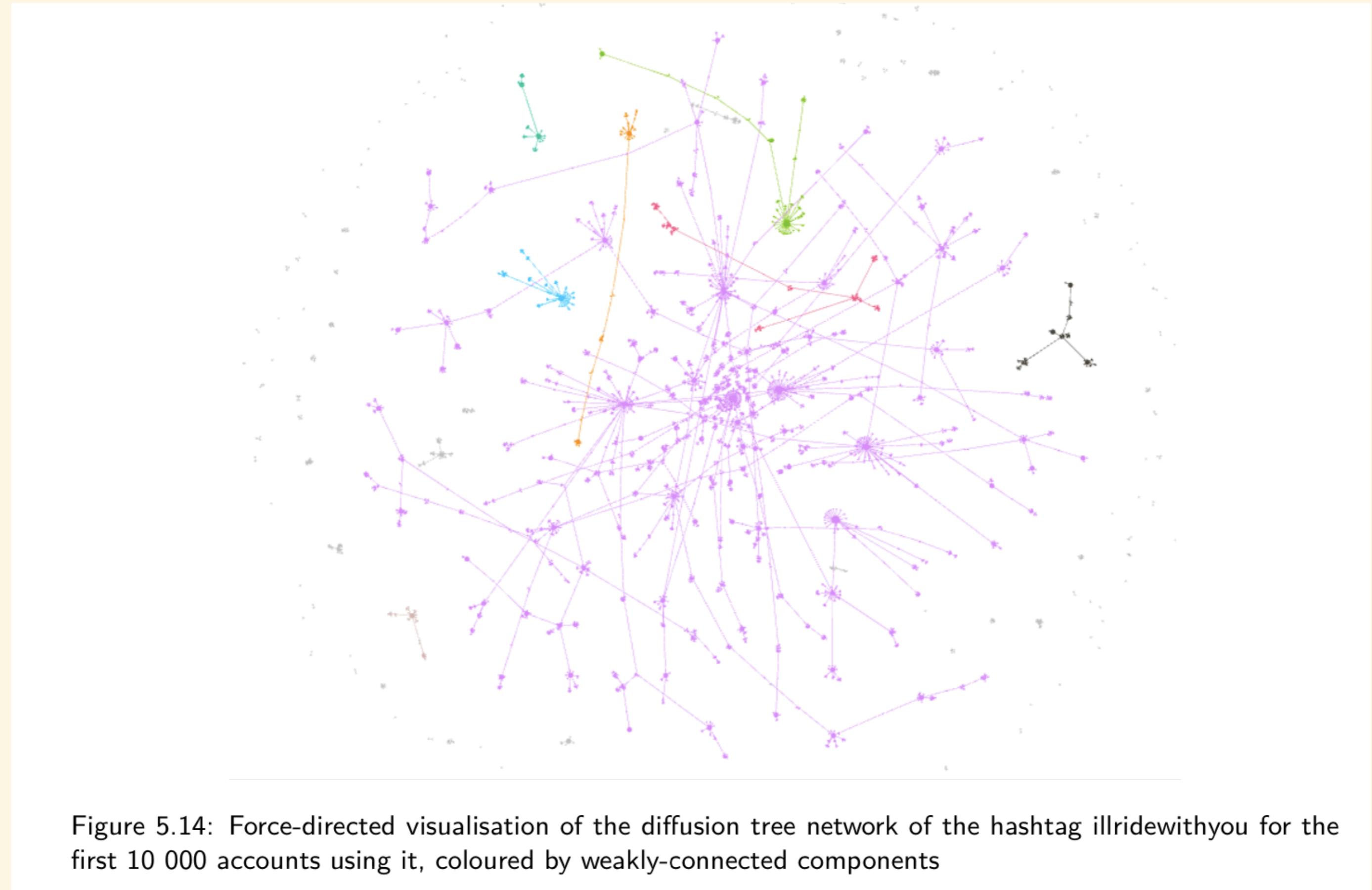
Links between top level domains in 2012 ("Topology of the WDC Hyperlink Graph",
<http://km.aifb.kit.edu/sites/webdatacommons/hyperlinkgraph/topology.html>)

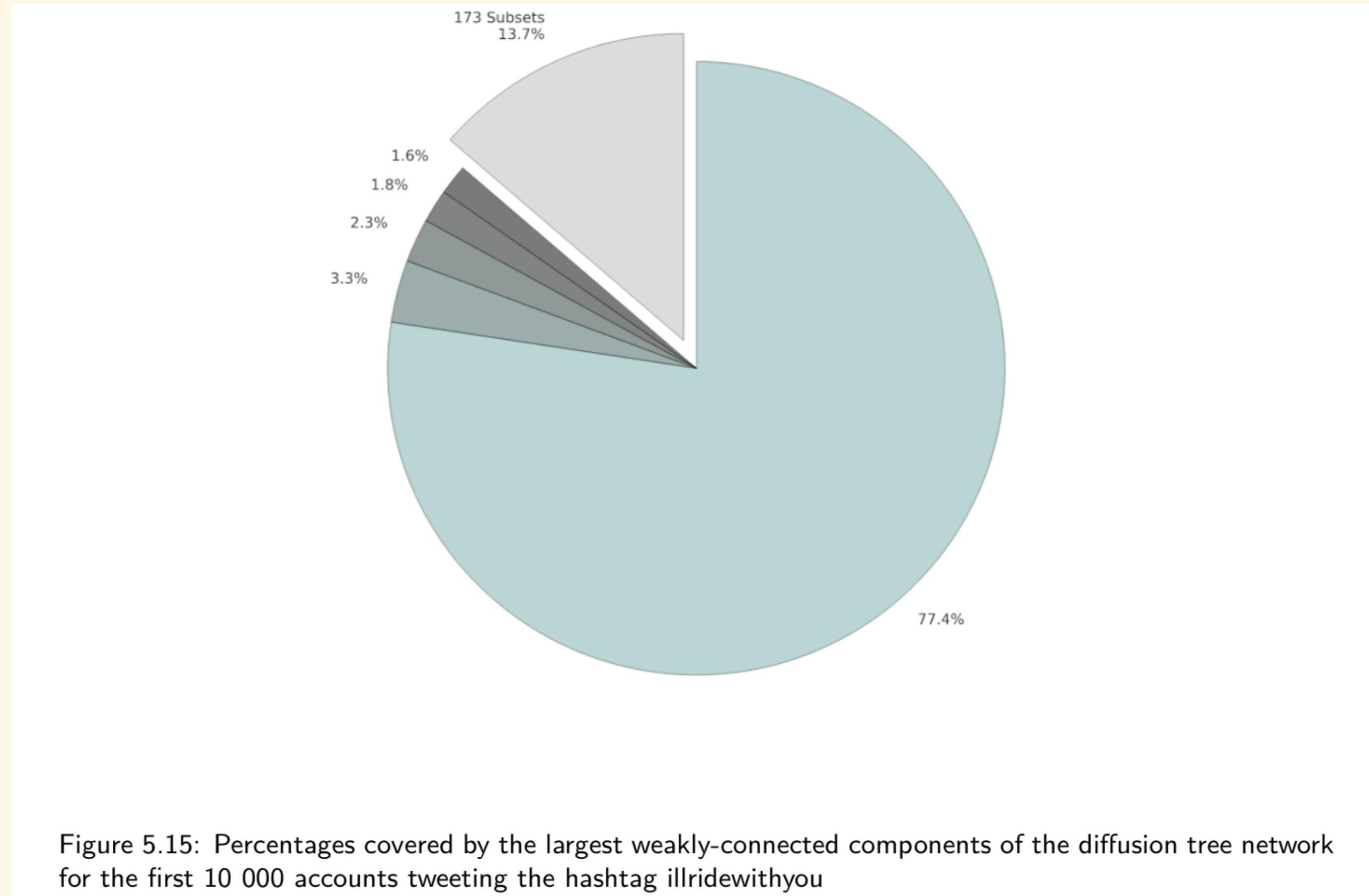
INFLUENCE (SOME/MOST?) INFORMATION DIFFUSION

#SYDNEYSIEGE VS #ILLRIDEWITHYOU VS \BREXIT PETITION

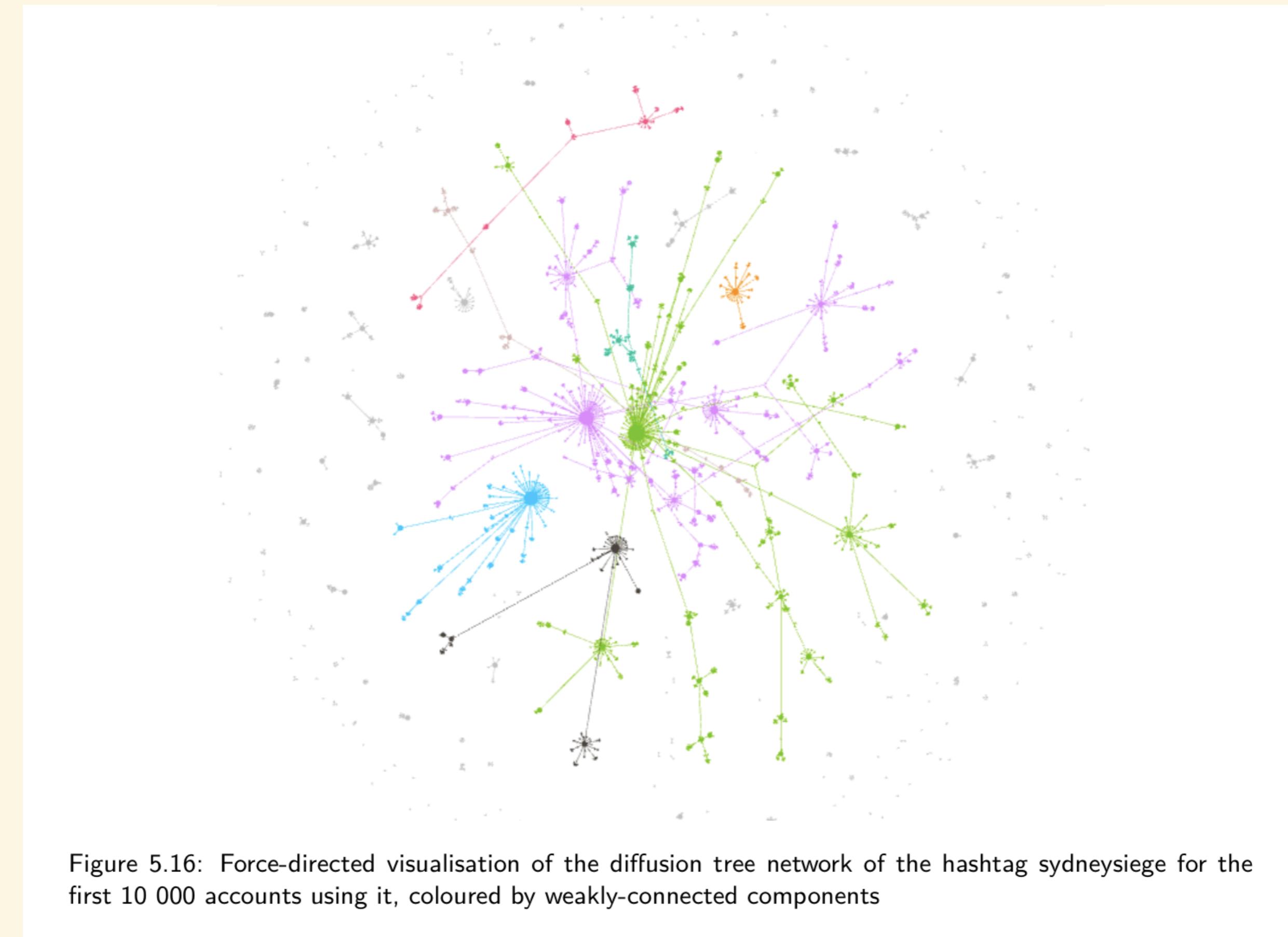
Münch, F. V. (2019). *Measuring the Networked Public – Exploring Network Science Methods for Large Scale Online Media Studies* [PhD thesis, Queensland University of Technology].
<https://doi.org/10.5204/thesis.eprints.125543>

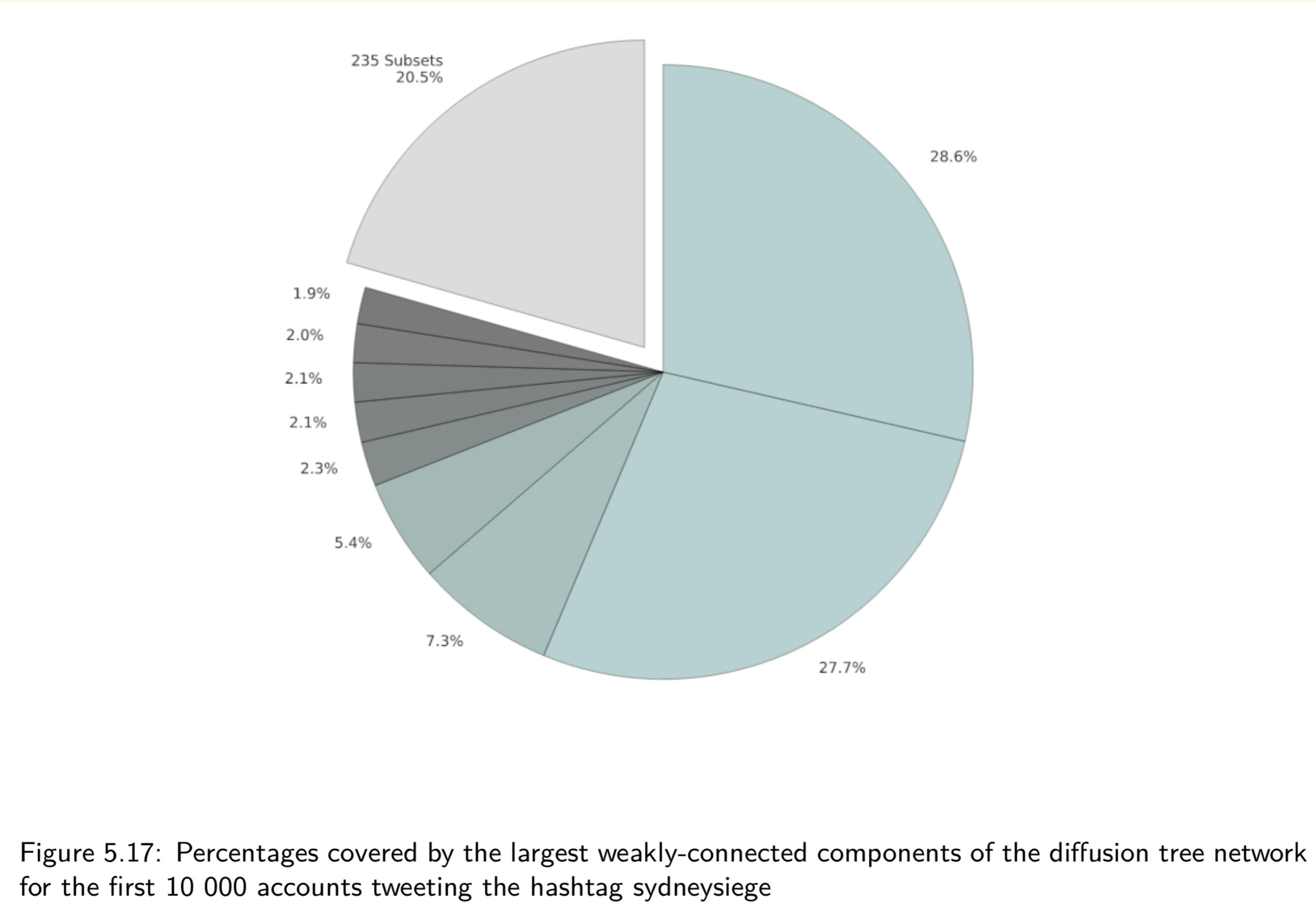
#ILLRIDEWITHYOU



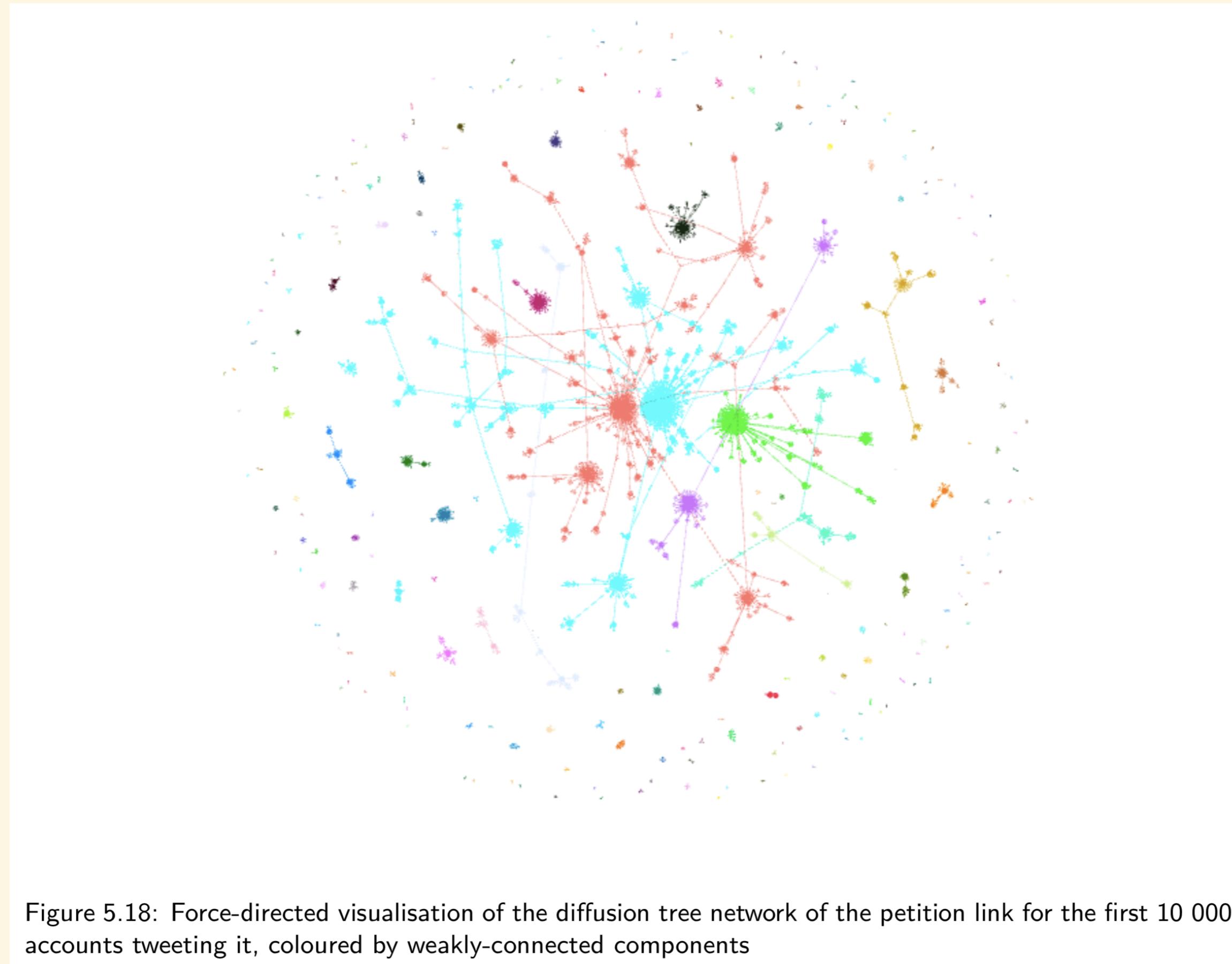


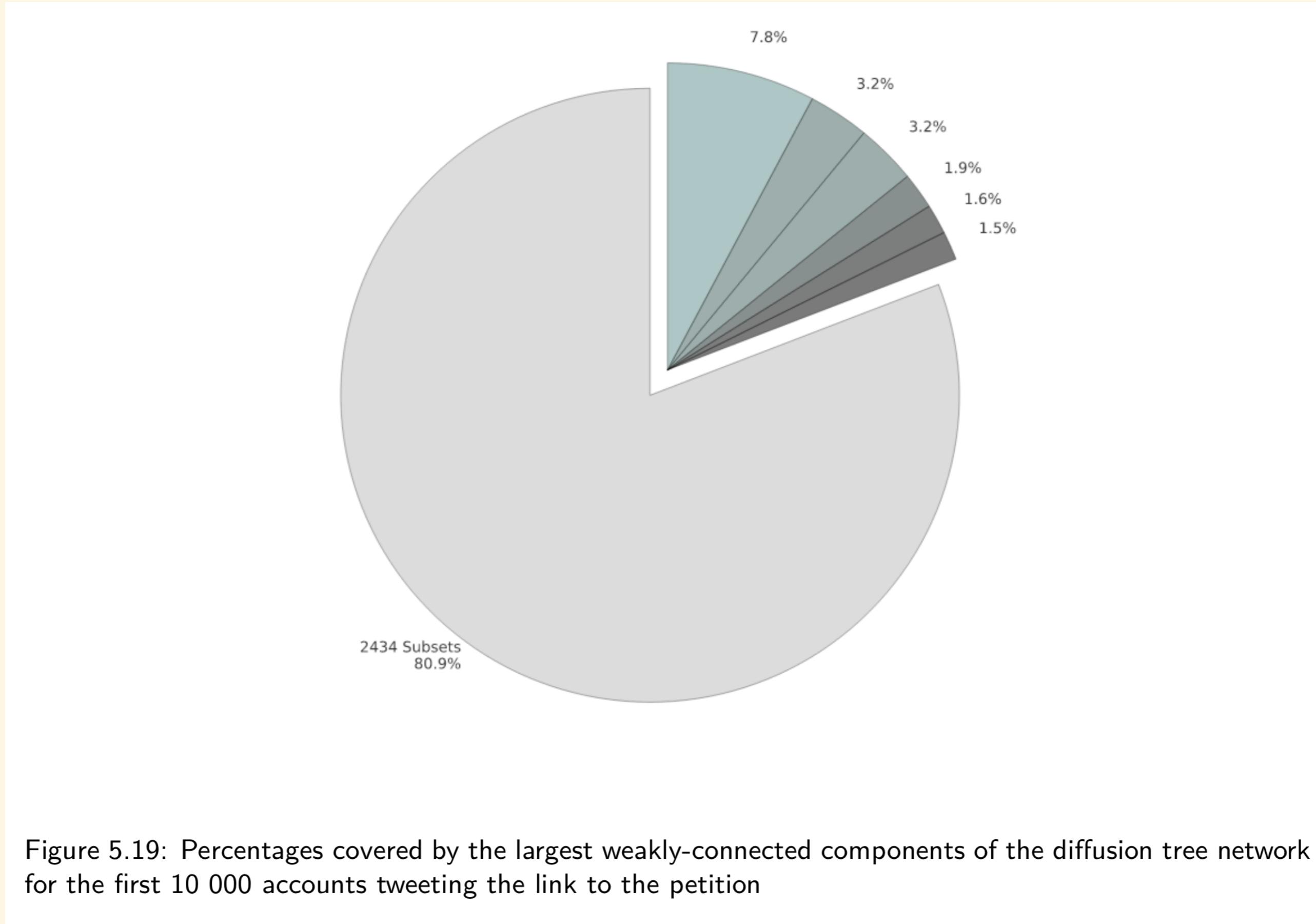
#SYDNEYSIEGE





ANTI-BREXIT PETITION





**THE NUMBER AND SIZE OF CONNECTED COMPONENTS INDICATES THE INFLUENCE OF THE NETWORK COMPARED TO
OUTSIDE SOURCES**

LEAD TO CLASSIFIABLE COMMUNICATION PATTERNS

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). *Classifying Twitter topic-networks using social network analysis*. 1–38. <https://doi.org/10.1177/2056305117691545>

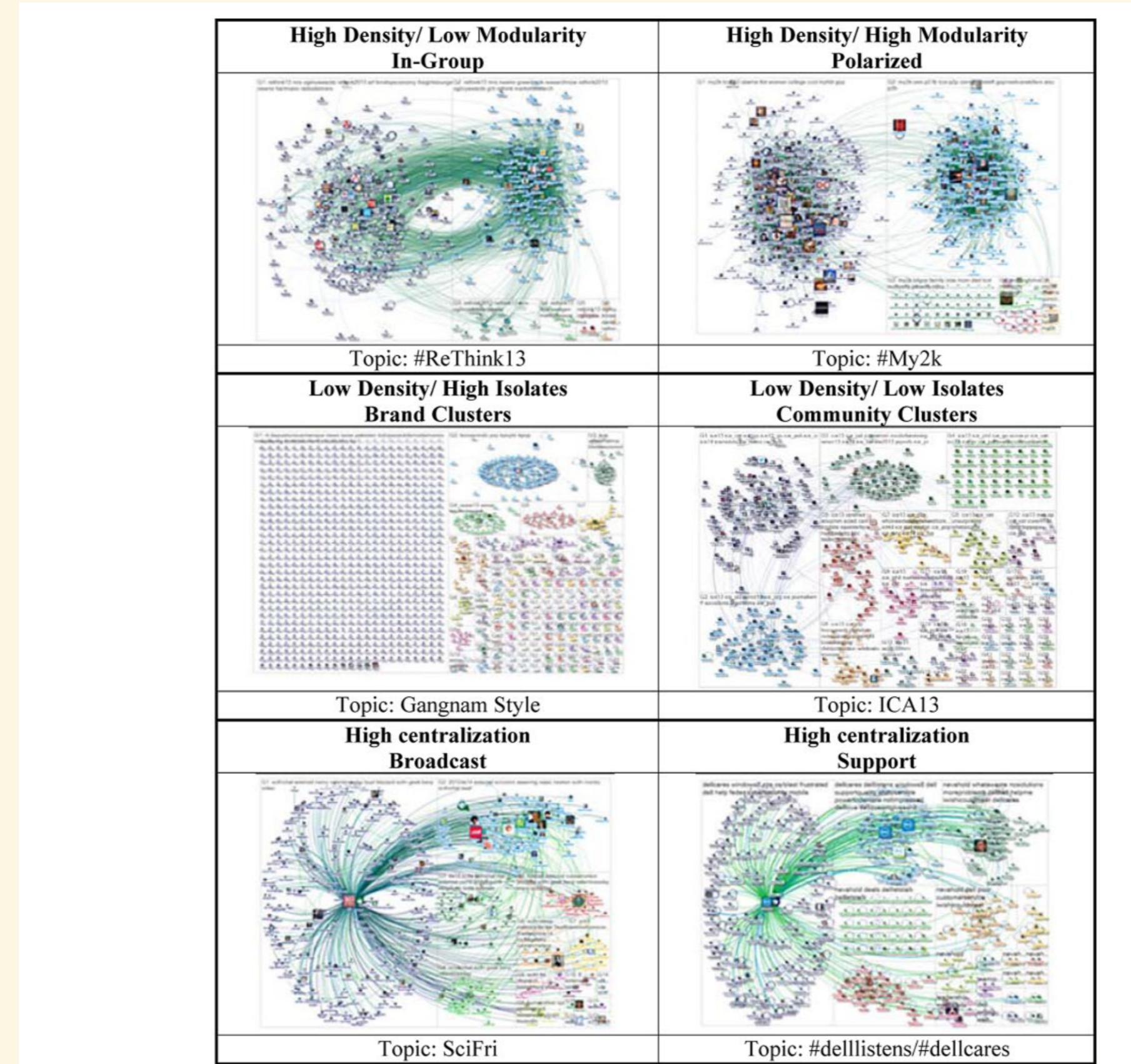
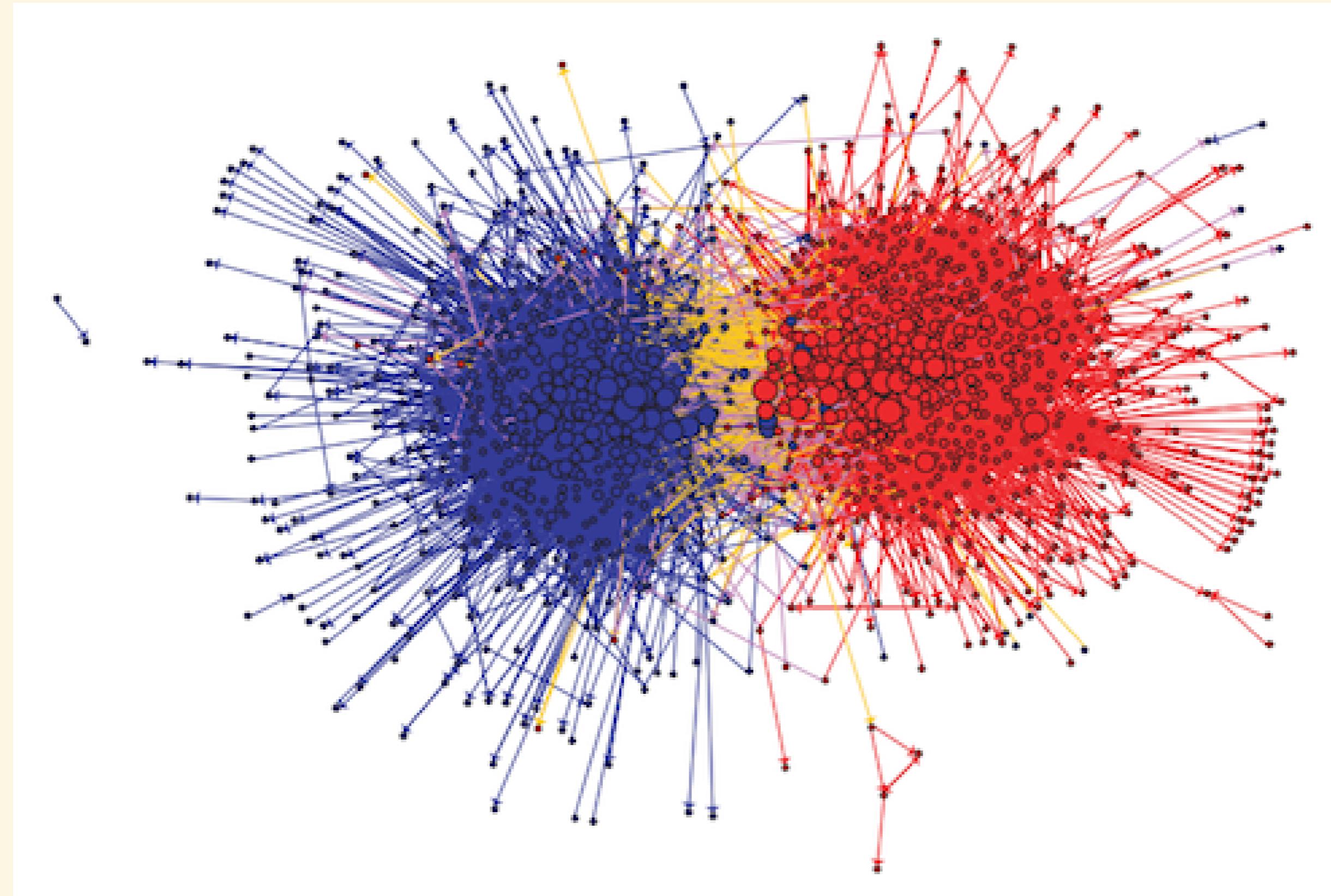


Figure 3. Network visualization by topic-network category.

EXAMPLE: POLARISATION

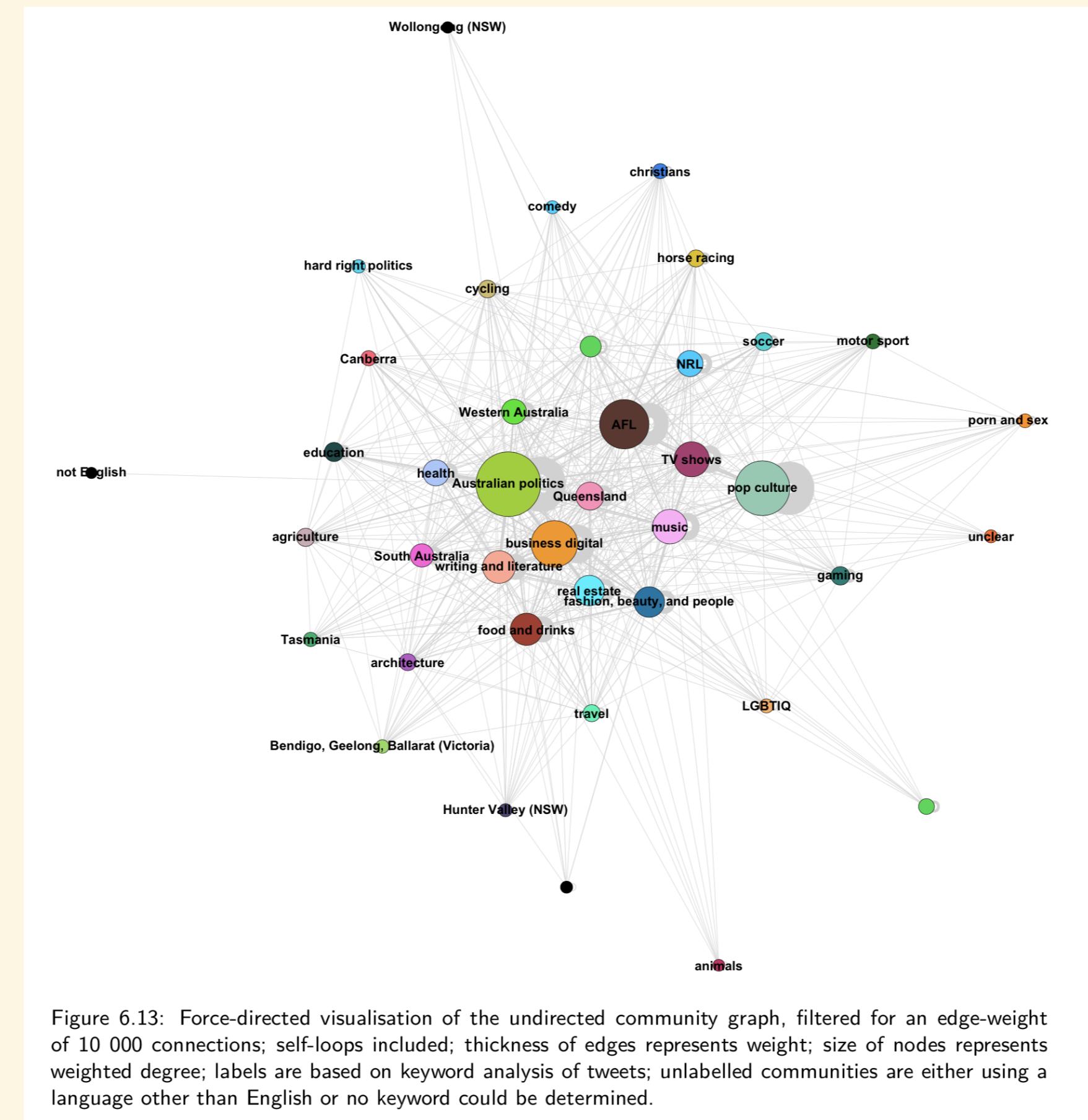


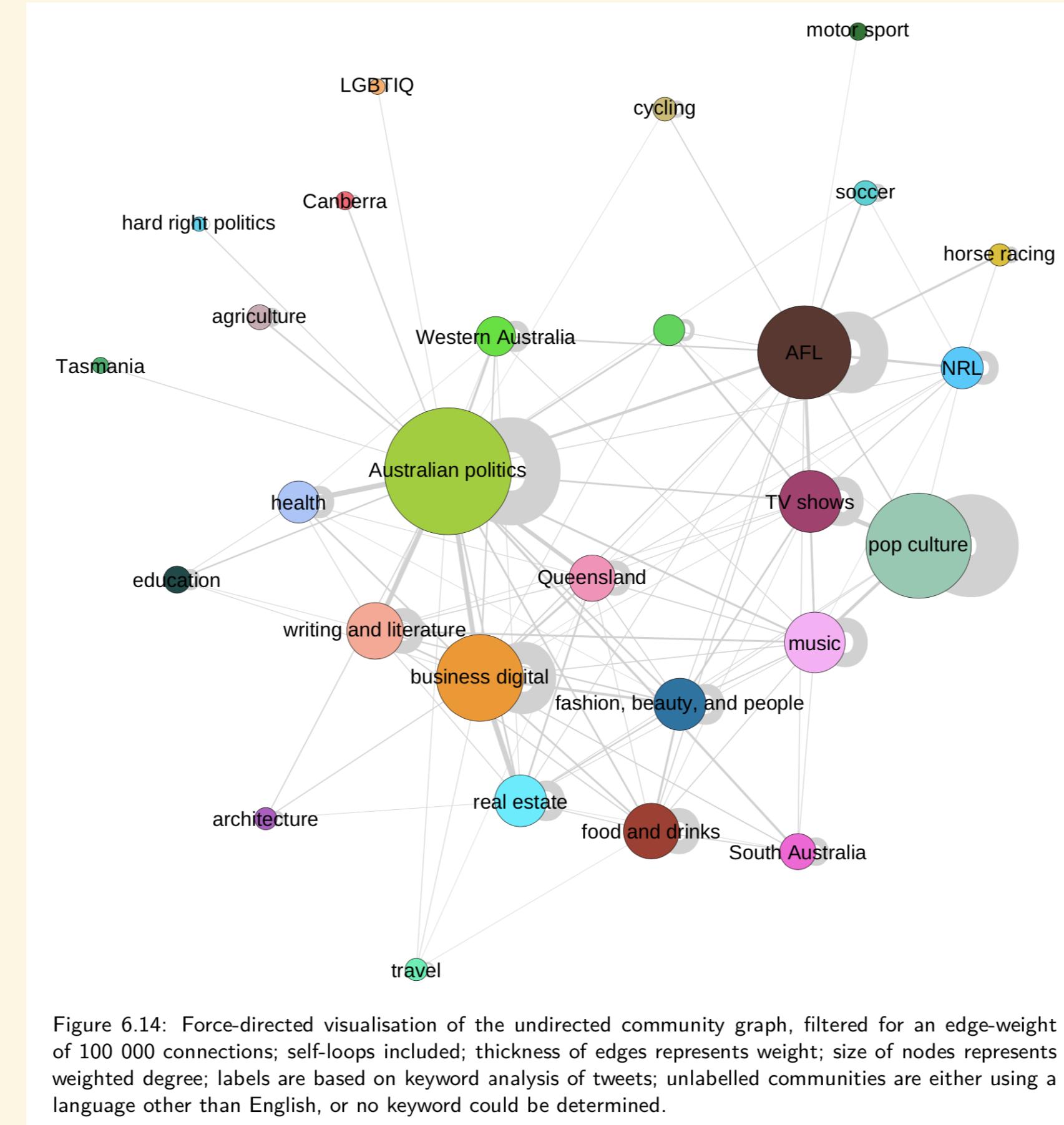
Lada A. Adamic and Natalie Glance. 2005. The political blogosphere and the 2004 U.S. election: divided they blog. In Proceedings of the 3rd international workshop on Link discovery (LinkKDD '05). Association for Computing Machinery, New York, NY, USA, 36–43. <https://doi.org/10.1145/1134271.1134277>

REFLECT LONG-TERM STRUCTURED SYSTEMS OF (PARTS OF) SOCIETY

- Bruns, A., Moon, B., Münch, F. V., & Sadkowsky, T. (2017). The Australian Twittersphere in 2016: Mapping the follower/followee network. *Social Media + Society*, 3(4).
<https://doi.org/10.1177/2056305117748162>
- Münch, F. V. (2019). *Measuring the Networked Public – Exploring Network Science Methods for Large Scale Online Media Studies* [PhD thesis, Queensland University of Technology].
<https://doi.org/10.5204/thesis.eprints.125543>
- Münch, F. V., & Rossi, L. (2020, October 5). A Tale of Two Twitters? Identifying Bridges Between Language Based Twitterspheres. *AoIR Selected Papers of Internet Research*.
<https://doi.org/10.5210/spir.v2020i0.11283>
- Münch, F. V., & Rossi, L. (2020). *Bootstrapping Follow Networks of Influential Twitter Accounts*. IC2S2. <https://vimeo.com/431470176>
- Münch, F. V., Thies, B., Puschmann, C., & Bruns, A. (2021). Walking Through Twitter: Sampling a Language-Based Follow Network of Influential Twitter Accounts. *Social Media + Society*.
<https://doi.org/10.1177/2056305120984475>

AUSTRALIAN TWITTERSPHERE





GERMAN TWITTERSPHERE

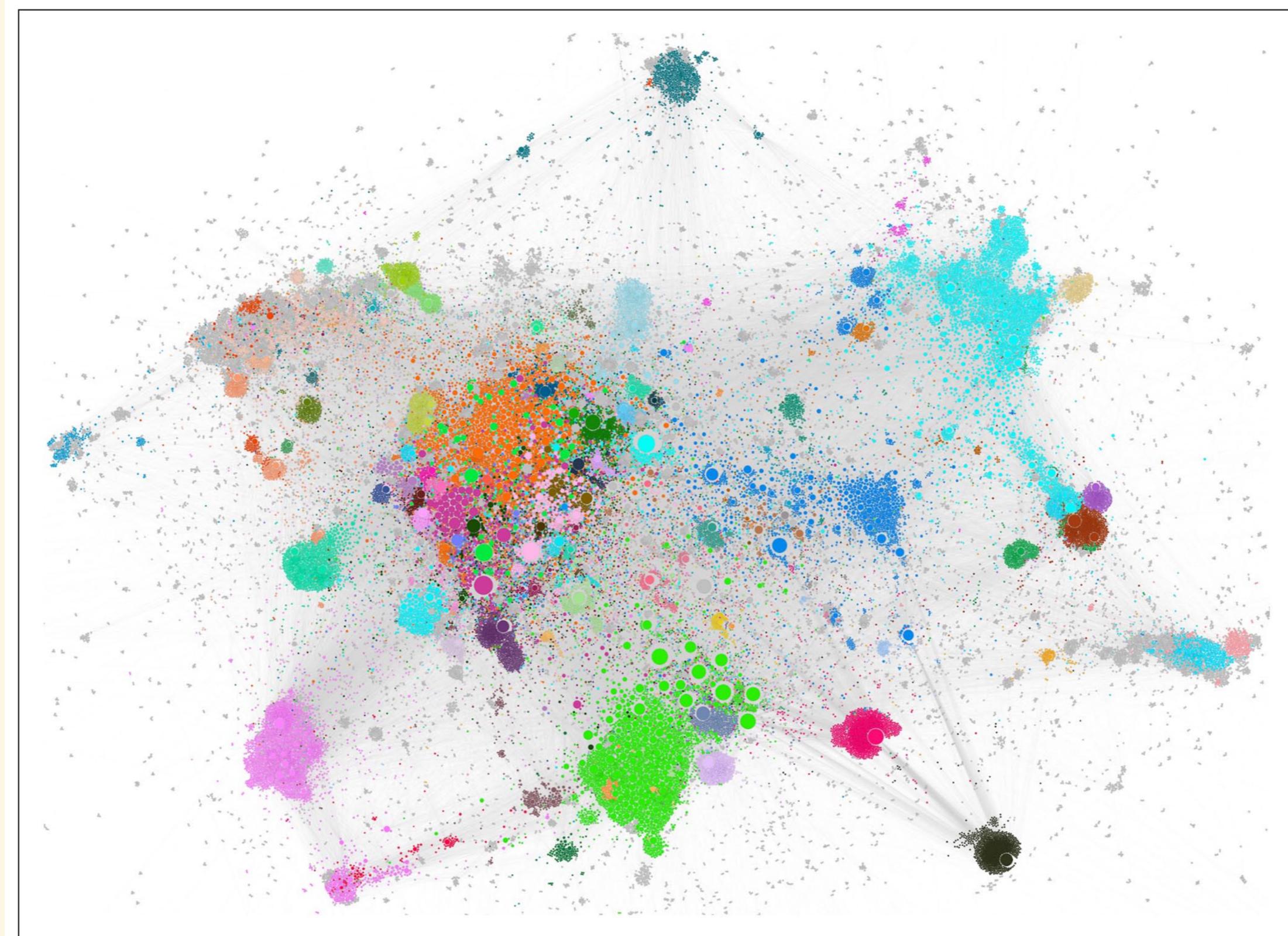


Figure 12. Central communities in the 3-core of our sample network; colored by largest communities detected with the Infomap community detection algorithm (Rosvall & Bergstrom, 2008; Rosvall et al., 2009); node size represents Page Rank (Brin & Page, 1998); layout done with Force Atlas 2 in Gephi (Bastian et al., 2009); (colored version available online).

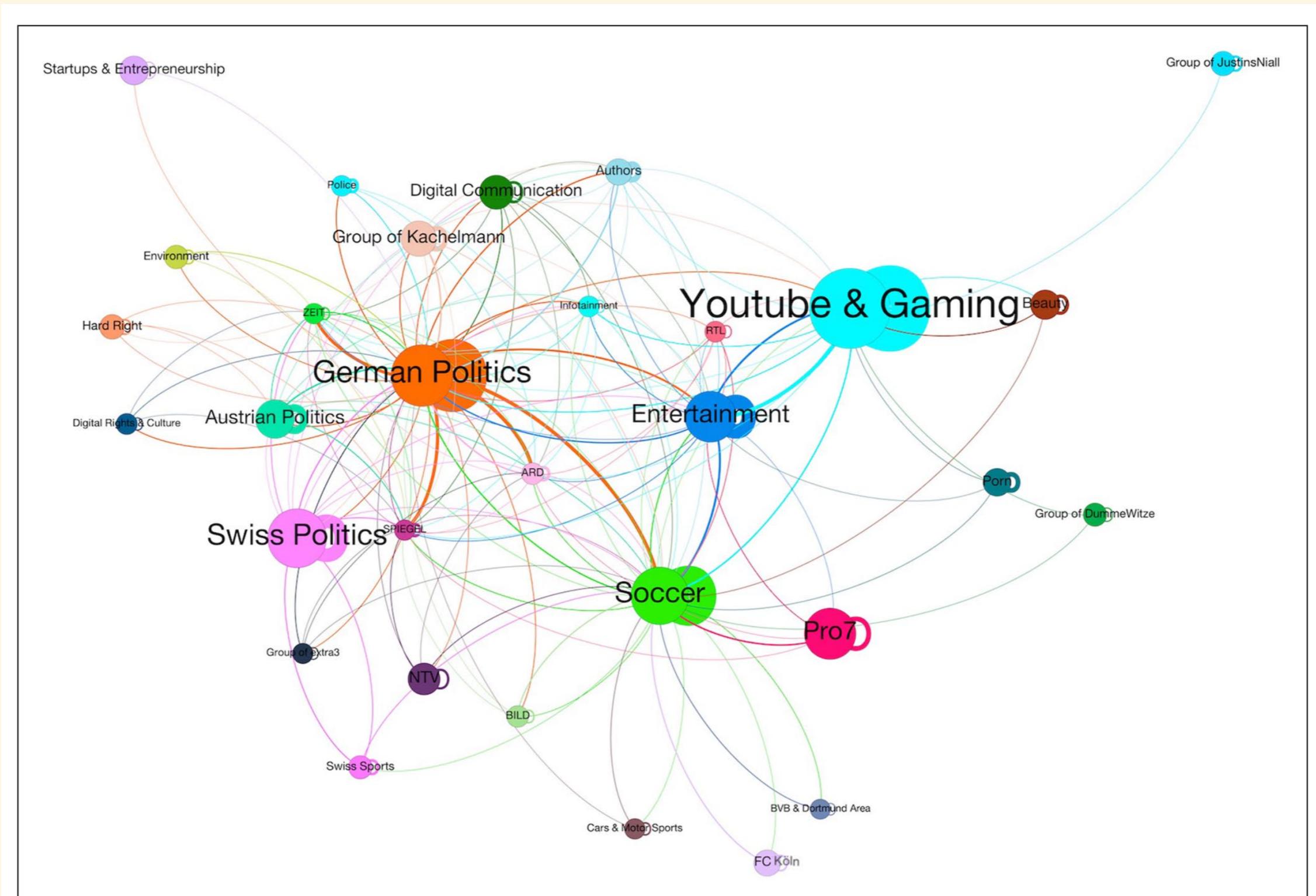


Figure 13. Community graph of communities in the 3-core of our sample with over 300 accounts, at least 80 active accounts during the examined timeframe, and edges with a weight of at least 150; edge width represents weight; edge direction follows clockwise curvature; edges colored by source node; node size represents the number of accounts in each community; node colors correspond with Figure 12; node labels based on interpretation of keywords and top accounts (see Supplemental Material); (colored version available online).

GERMAN-ITALIAN TWITTERSPHERE

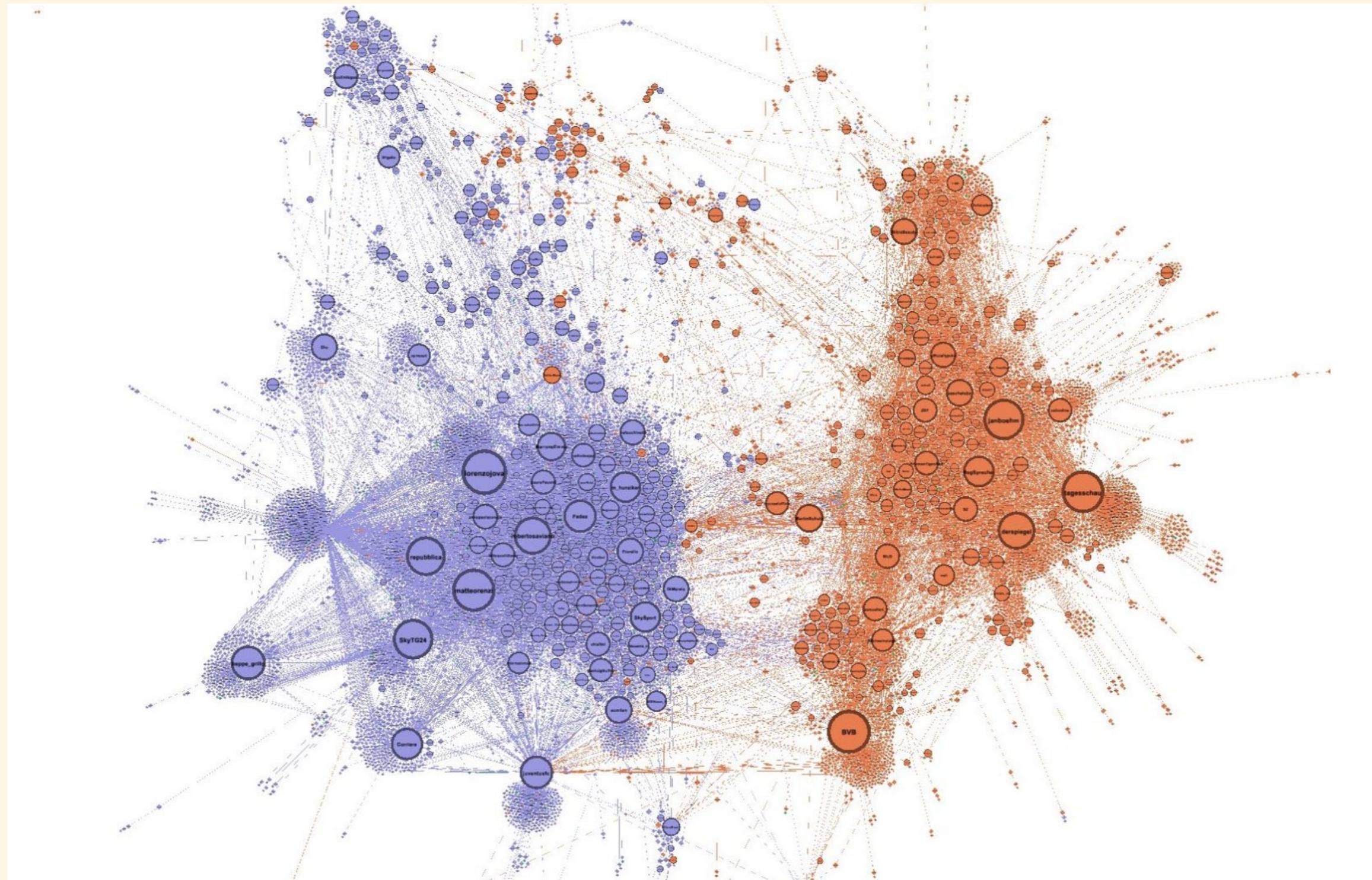


Figure 1: Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (purple) - German (orange) follow network sample. Nodes sized by betweenness centrality (Brandes, 2001).

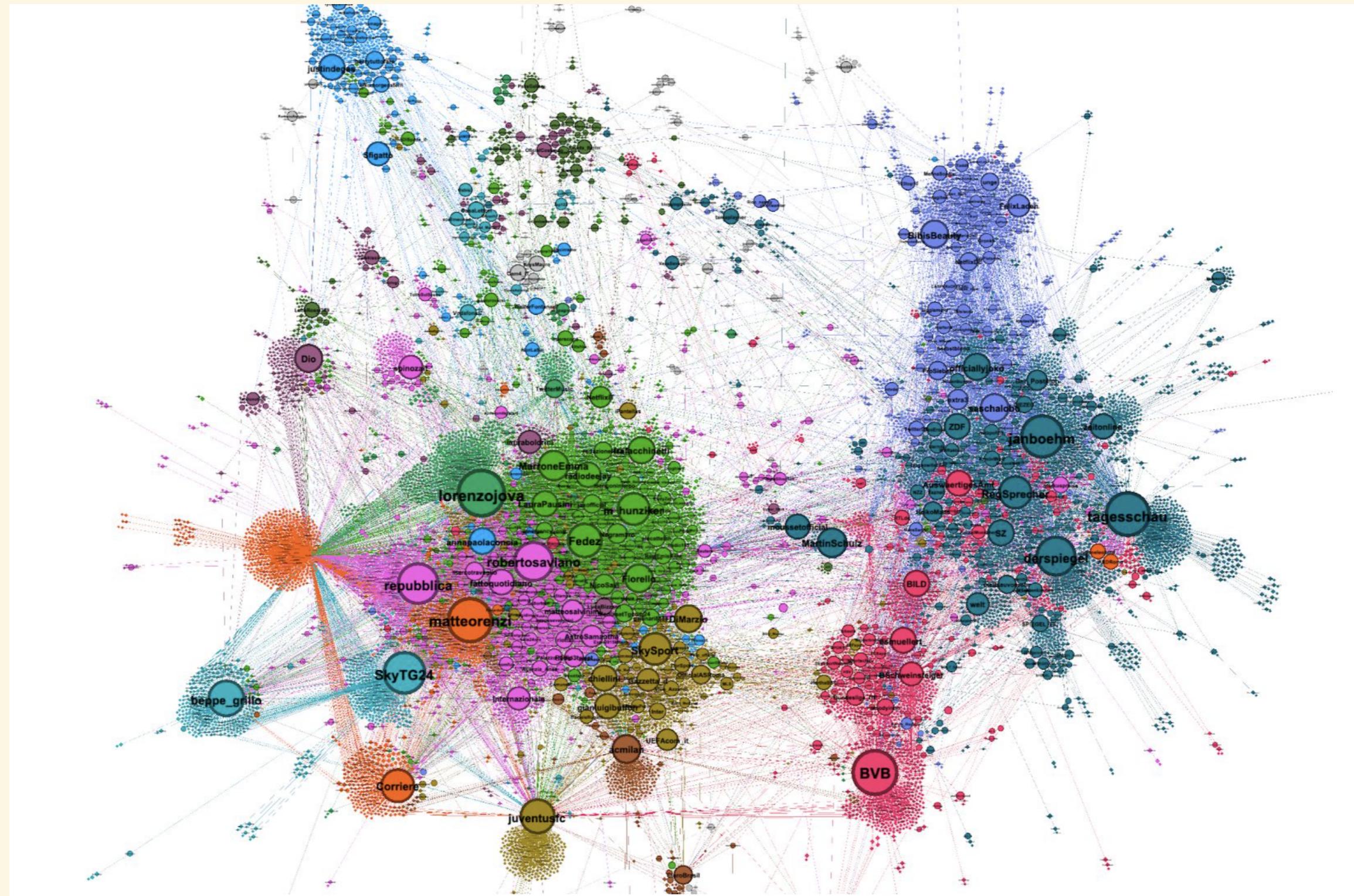


Figure 3: Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (left) - German (right) follow network sample. Nodes sized by betweenness centrality (Brandes, 2001). Coloured by modularity maximising clusters.

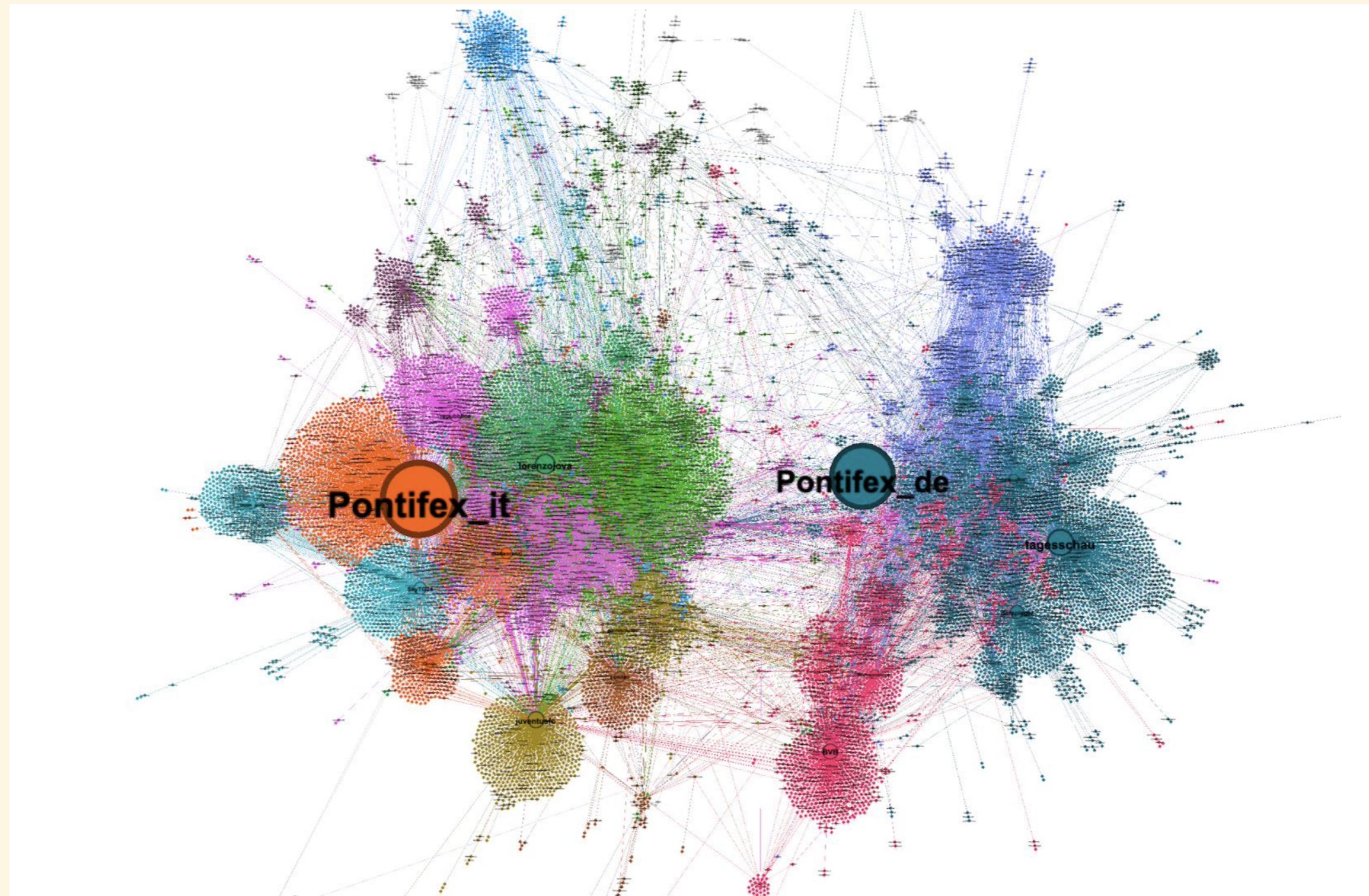
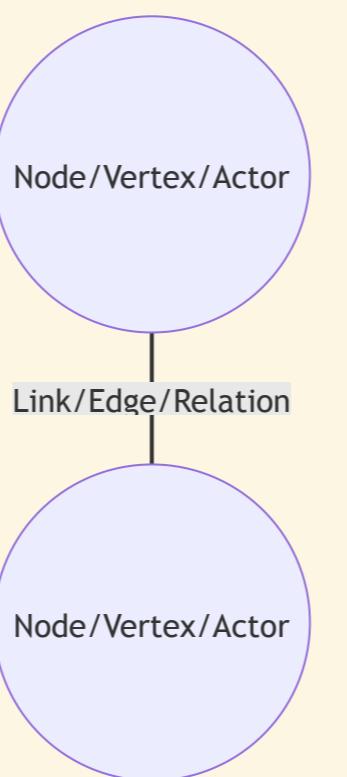


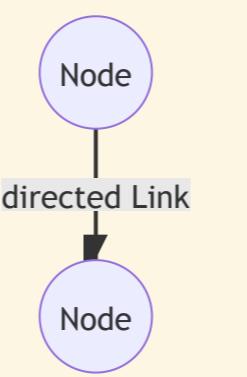
Figure 2: Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (left) - German (right) follow network sample. Nodes sized by Page Rank (Brin & Page, 1998). Coloured by modularity maximising clusters.

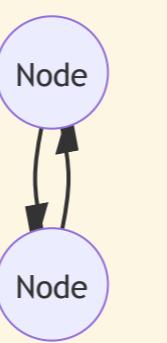
NETWORK ANALYSIS FUNDAMENTALS

ELEMENTS AND PROPERTIES OF NETWORKS

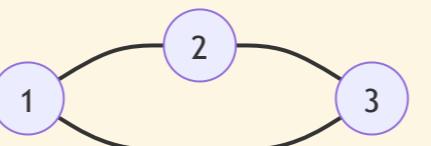
DYADS



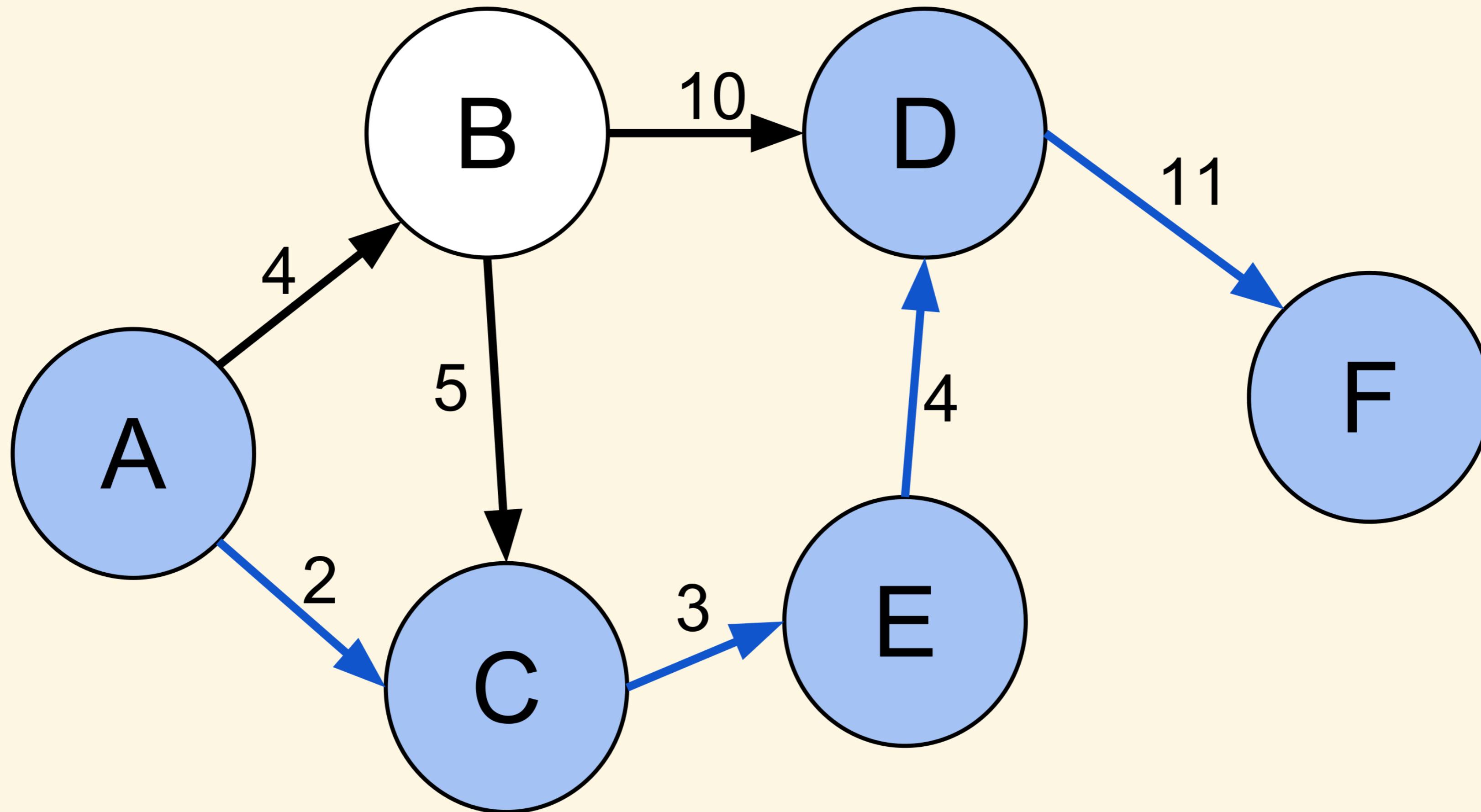




TRIADS



WEIGHTED LINKS & (SHORTEST) PATHS



MEASUREMENTS OF NETWORKS/GRAPHS AND THEIR ELEMENTS

NODE MEASURES

IMPORTANT CENTRALITY MEASURES

(IN/OUT-)DEGREE CENTRALITY

CLOSENESS CENTRALITY

EIGENVECTOR CENTRALITY

PAGE RANK

K-CORENESS

BETWEENNESS-CENTRALITY

LOCAL CLUSTERING COEFFICIENT

NETWORK MEASURES

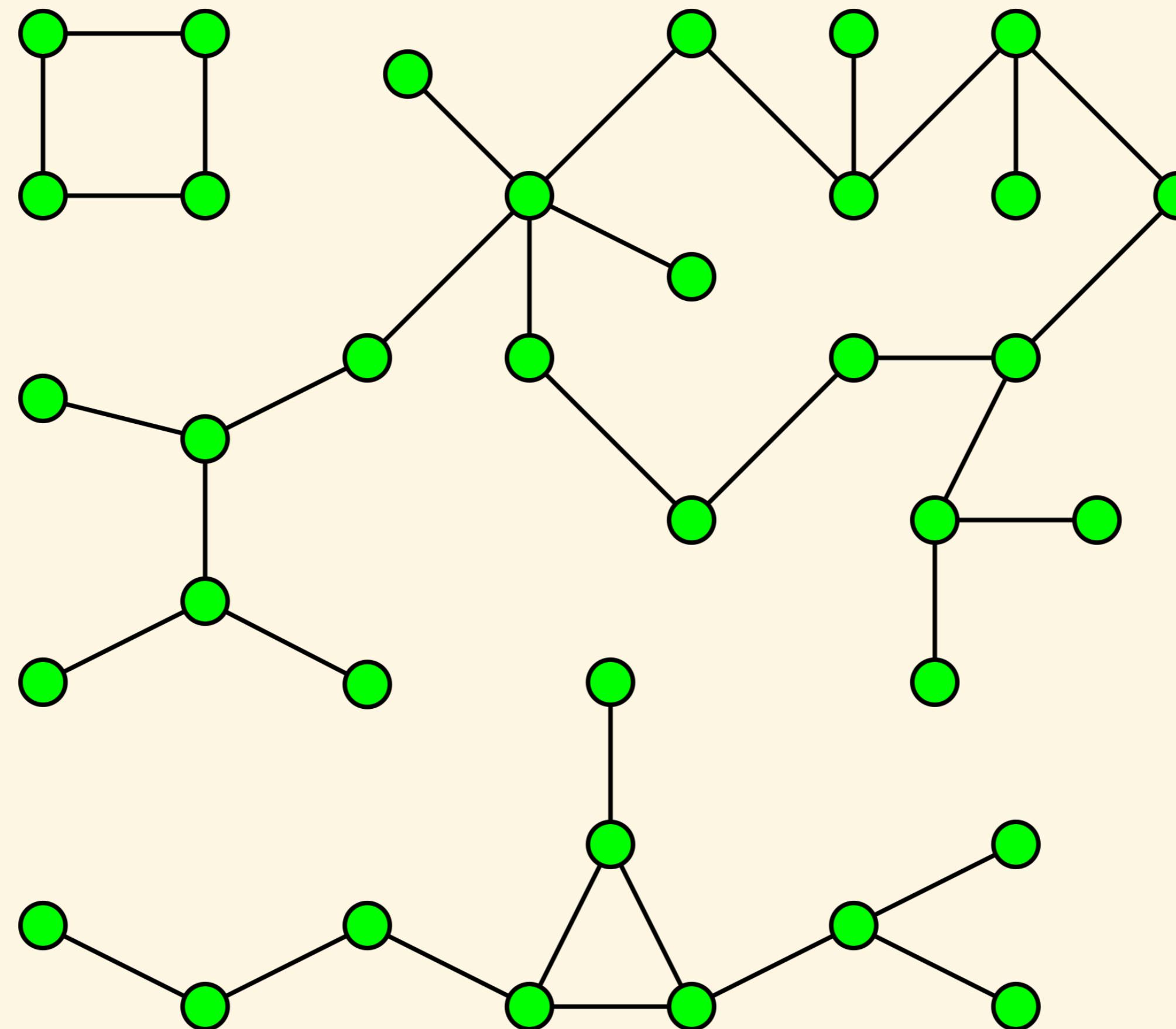
GLOBAL CLUSTERING COEFFICIENT

DIAMETER

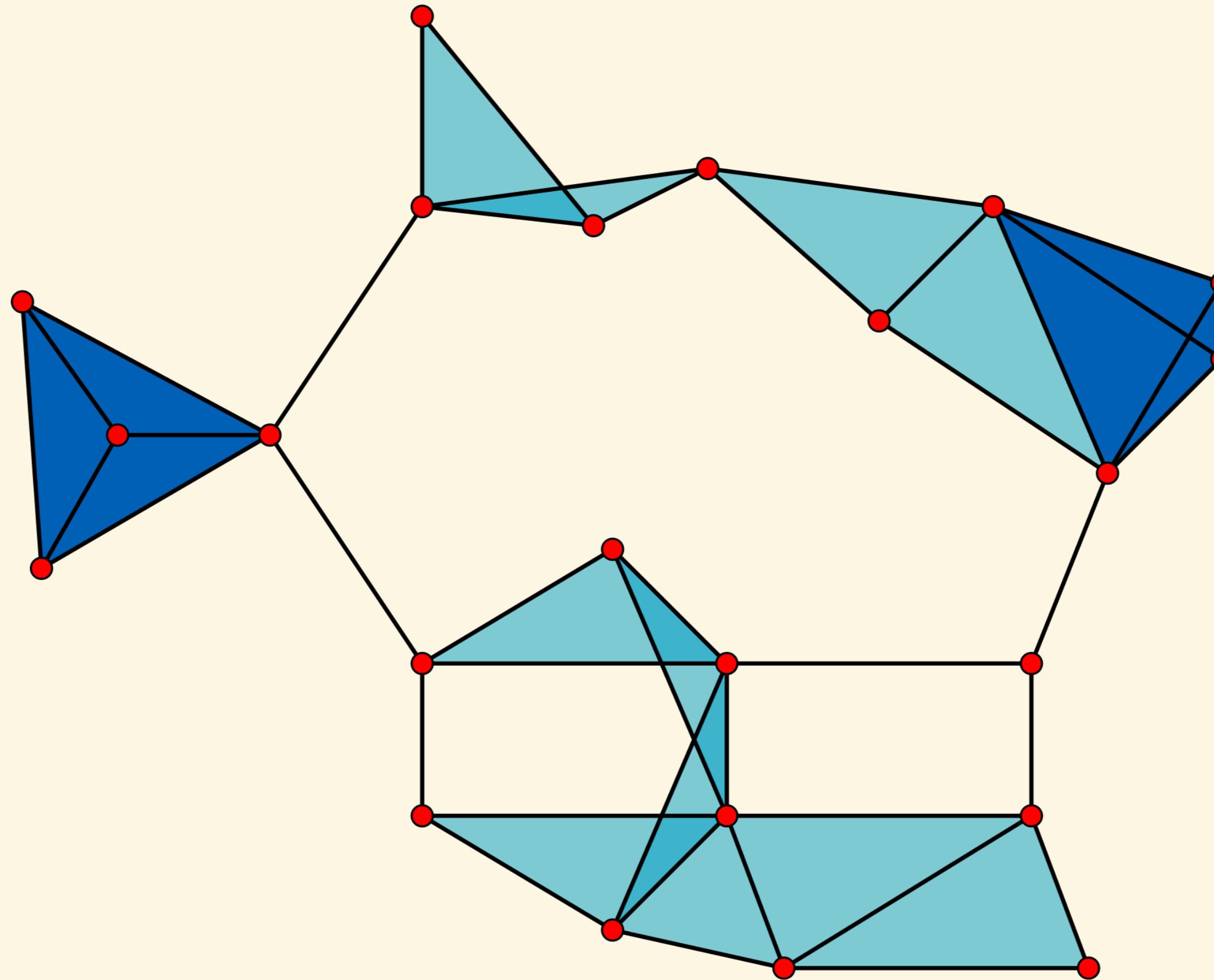
DENSITY

NETWORKS WITHIN NETWORKS

(WEAKLY) CONNECTED COMPONENTS

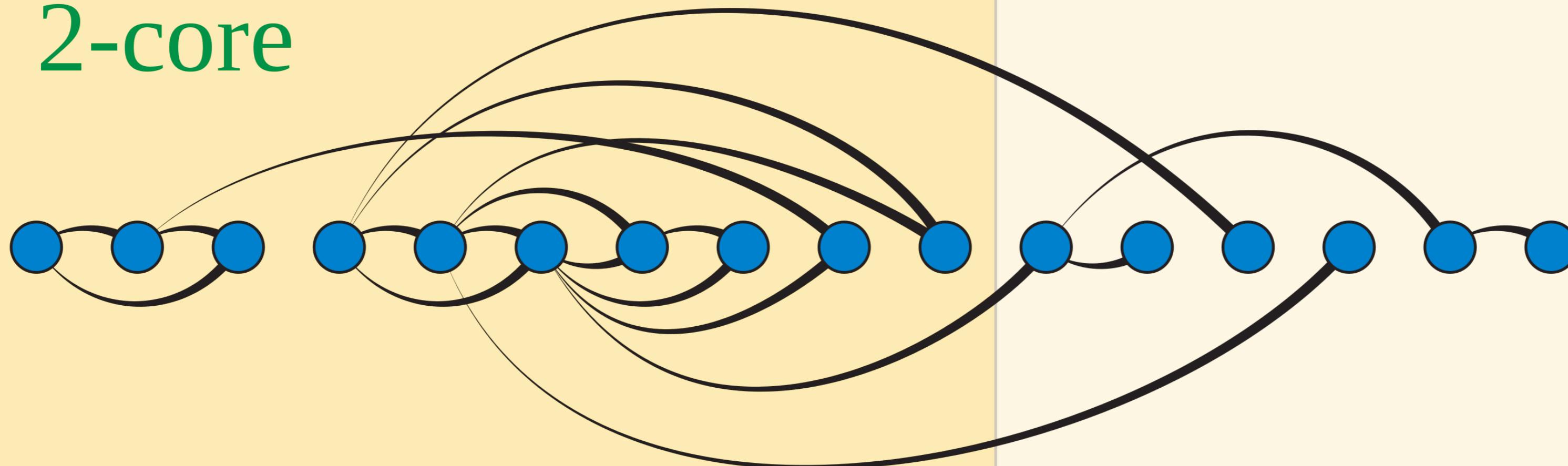


CLIQUE



K-CORES

2-core



COMMUNITIES/CLUSTERS

Depend on the detection algorithms used. Two of the most popular are

- **Modularity Maximisation (mostly in the Louvain implementation)** based on the relative density of in-/out-group edges
 - and
- **Map Equation (infomap)** based on the length of stay of random walks in certain regions of the network (technically the minimization of the description length of its path)

FLAT COMMUNITIES

- simplify complex systems
- easy to understand
- can oversimplify
- inherently have a resolution limit/arbitrary resolution

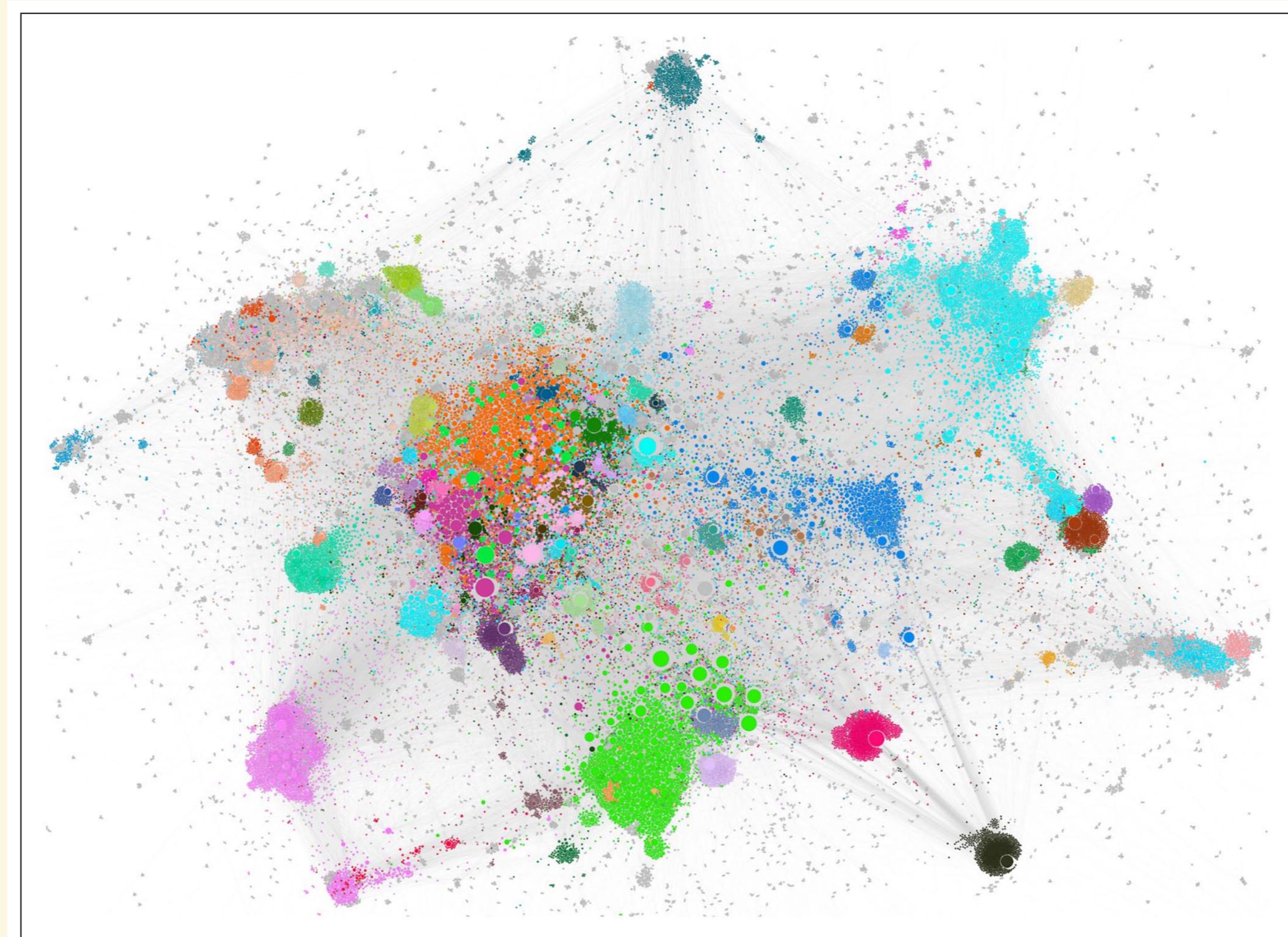
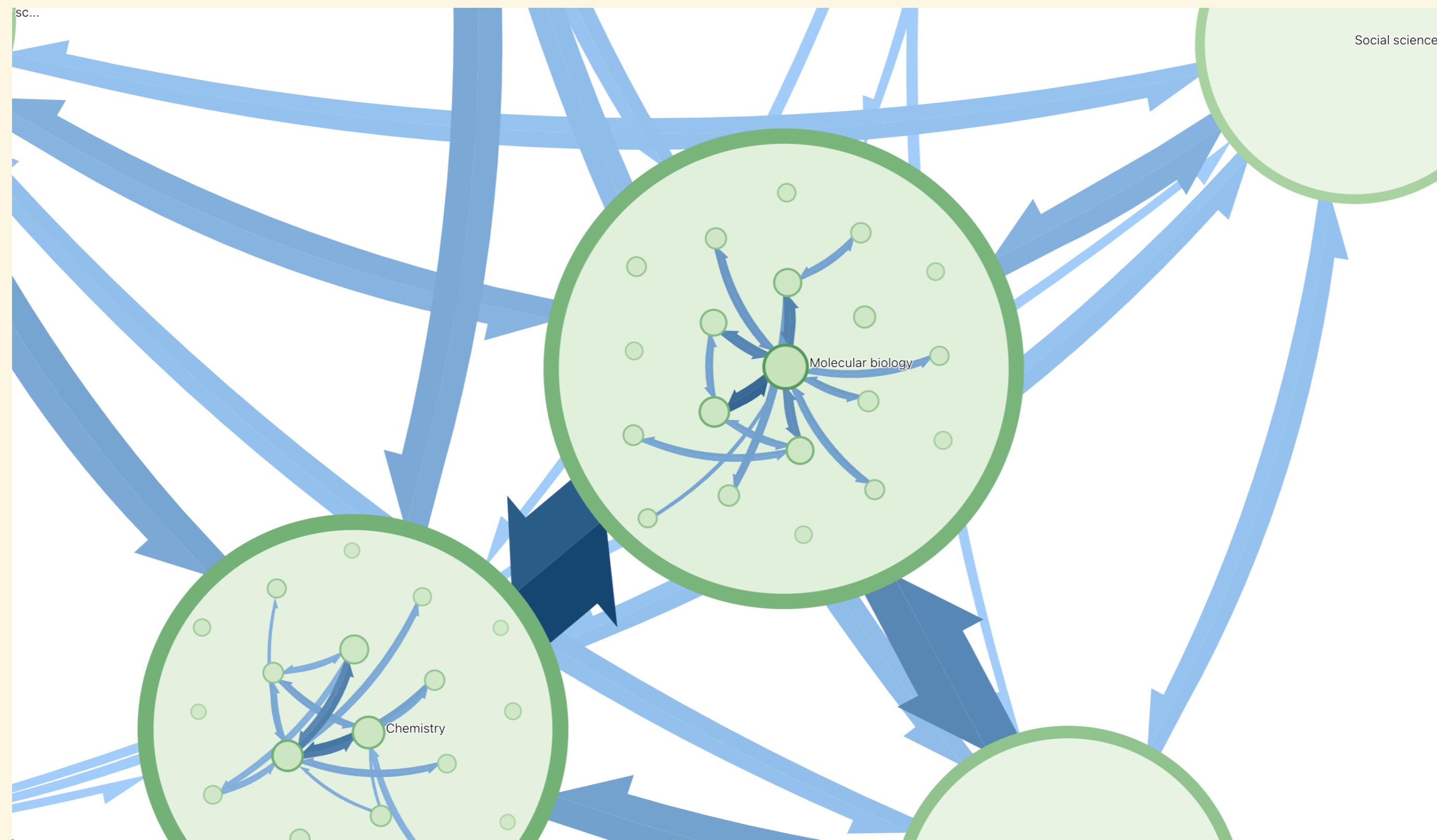


Figure 12. Central communities in the 3-core of our sample network; colored by largest communities detected with the Infomap community detection algorithm (Rosvall & Bergstrom, 2008; Rosvall et al., 2009); node size represents Page Rank (Brin & Page, 1998); layout done with Force Atlas 2 in Gephi (Bastian et al., 2009); (colored version available online).

HIERARCHICAL AND OVERLAPPING COMMUNITIES

- Possible, e.g. with infomap
- often closer to reality
- often easier to interpret
- harder to analyse and visualise



<https://www.mapequation.org/navigator/>

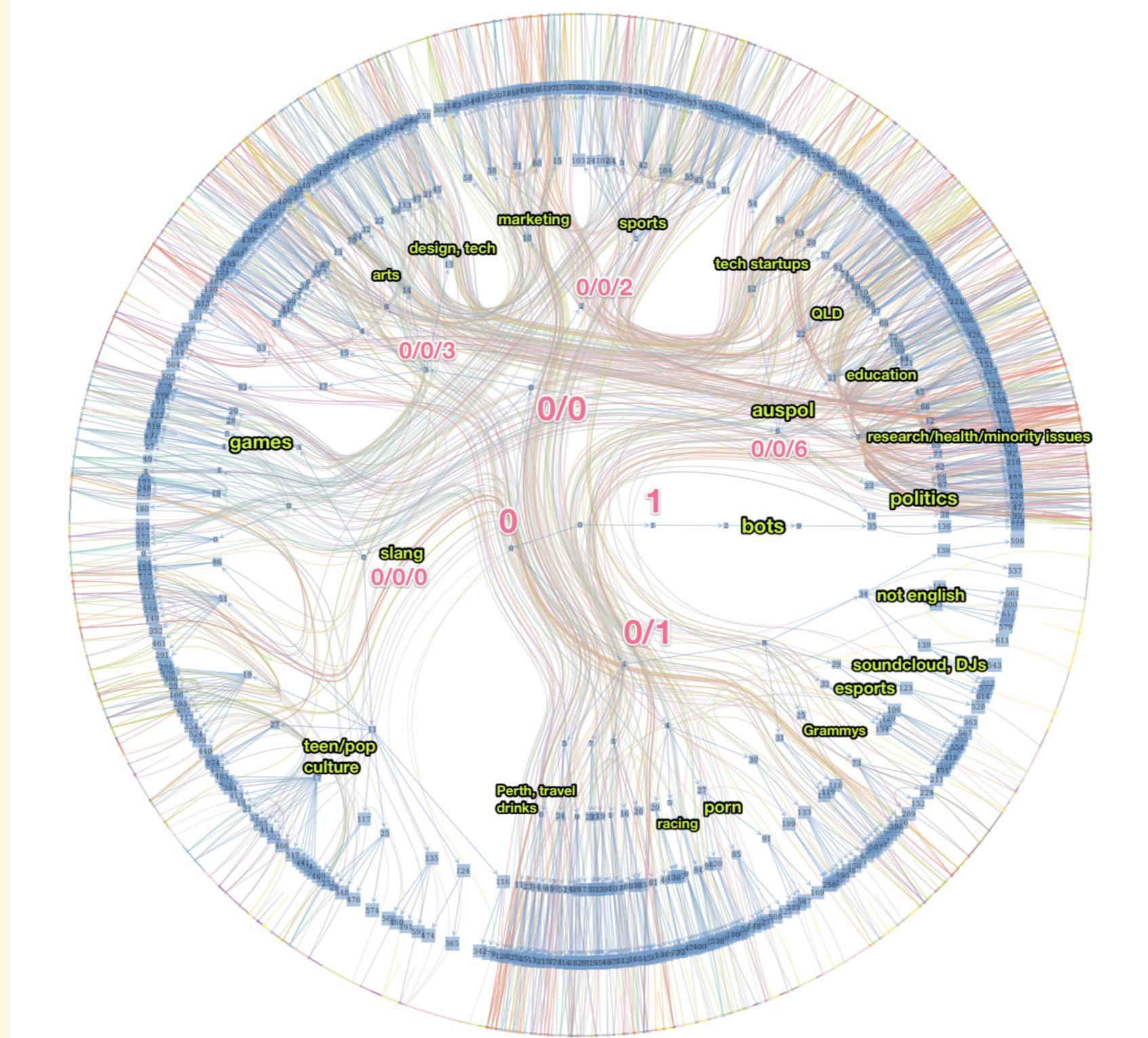


Figure 6.22: Annotated visualisation of the nested stochastic block model inferred for the filtered Australian follow network, edges sampled down to 1000 edges; nodes on outer circle represent accounts; labels based on clearly interpretable results from keyword extraction for blocks detected on level 2; labelling not complete and for overview only; pink labels used for references in text. (High resolution version available online.)

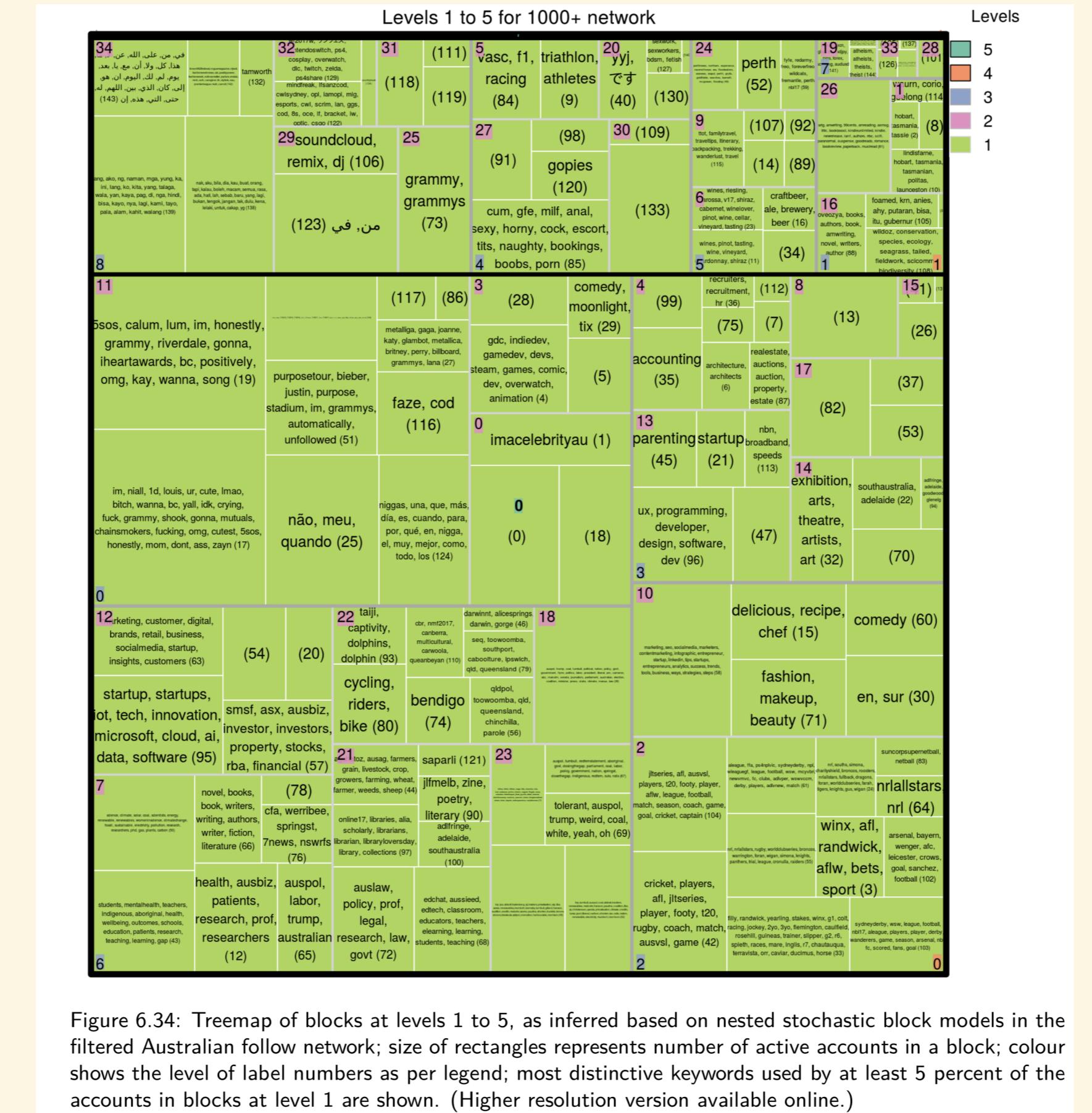


Figure 6.34: Treemap of blocks at levels 1 to 5, as inferred based on nested stochastic block models in the filtered Australian follow network; size of rectangles represents number of active accounts in a block; colour shows the level of label numbers as per legend; most distinctive keywords used by at least 5 percent of the accounts in blocks at level 1 are shown. (Higher resolution version available online.)

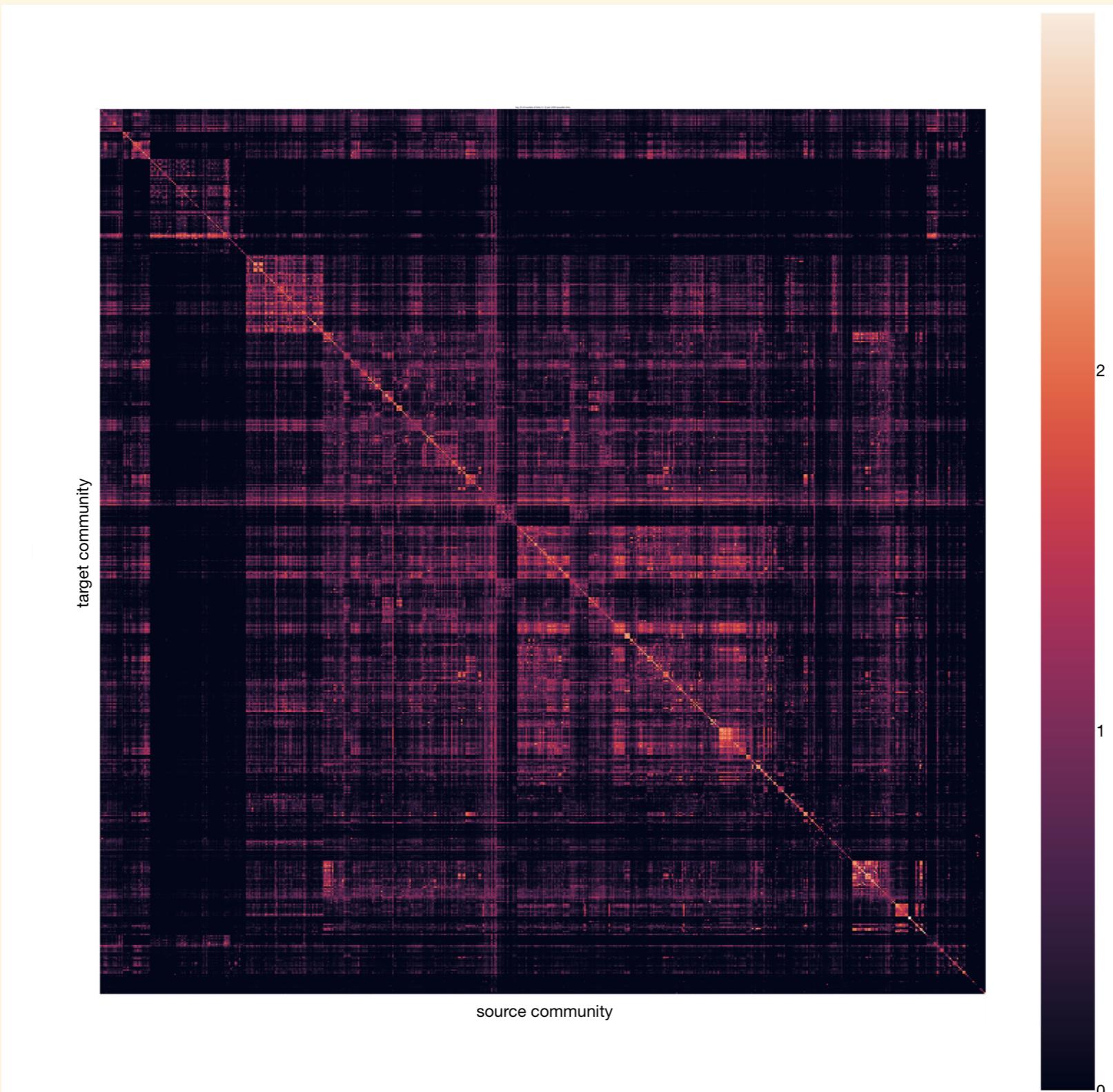


Figure 6.40: Visualisation of the adjacency matrix of the directed block graph of the filtered Australian follow network at level 0, as inferred based on nested stochastic block models; colour shows the logarithm to base 10 of the number of connections + 1 per 1000 possible connections between the source block on the horizontal, and the target block on the vertical axis.

DATA SOURCES FOR (ONLINE ((SOCIAL) MEDIA)) NETWORKS

Repositories

APIs

Scraping

