FUNDAMENTALS OF (ONLINE ((SOCIAL) MEDIA)) NETWORK ANALYSIS

LECTURE 2

Epistemological/methodological challenges of network science in computational social science, network science as a way to handle organised complexity, research examples

WHO IS THIS GUY?

Dr. Felix Victor Münch

Media Research Methods Lab & Social Media Observatory of the Research Institute Social Cohesion

@ Leibniz-Institute for Media Research | Hans-Bredow-Institut

THE PLAN

1. Computational Social Science
2. Today's Example: The Networked Public Sphere
3. Organised Complexity and Network Science
4. Challenges in Network Science for Computational Communication Science
5. A Possible Methodological Framework
6. Example 1: Measuring Virality of Communication
7. Example 2: Mapping a Public Sphere
8. What is it good for?

Afterwards:

Practical for students from last time

1. COMPUTATIONAL SOCIAL SCIENCES

HOW TO BECOME A "COMPUTATIONAL SOCIAL/COMMUNICATION SCIENTIST"?

C.V. tl;dr:

Physics B.Sc. -> Journalism M.A. -> Online Brand Communication Agency

-> PhD @ Queensland University of Technology, Digital Media Research Centre

🤷‍♂️

WHAT IS COMPUTATIONAL SOCIAL SCIENCE?

Social Sciences by computational means, thereby

- Interdisciplinary
- Multi-methods, sometimes mixed methods
- **neither quantitative nor qualitative**

WHAT IS COMPUTATIONAL SOCIAL SCIENCE?

INTERDISCIPLINARY

- Social Sciences
- Communication Studies
- Computer Sciences
- Mathematics
- ...

WHAT IS COMPUTATIONAL SOCIAL SCIENCE?

MIXED/MULTI-METHODS

- Content Analysis
- Natural Language Processing
- Machine Learning
- Network Analysis
- Surveys
- Questionnaires
- Data Donations
- ...

WHAT IS COMPUTATIONAL SOCIAL SCIENCE?

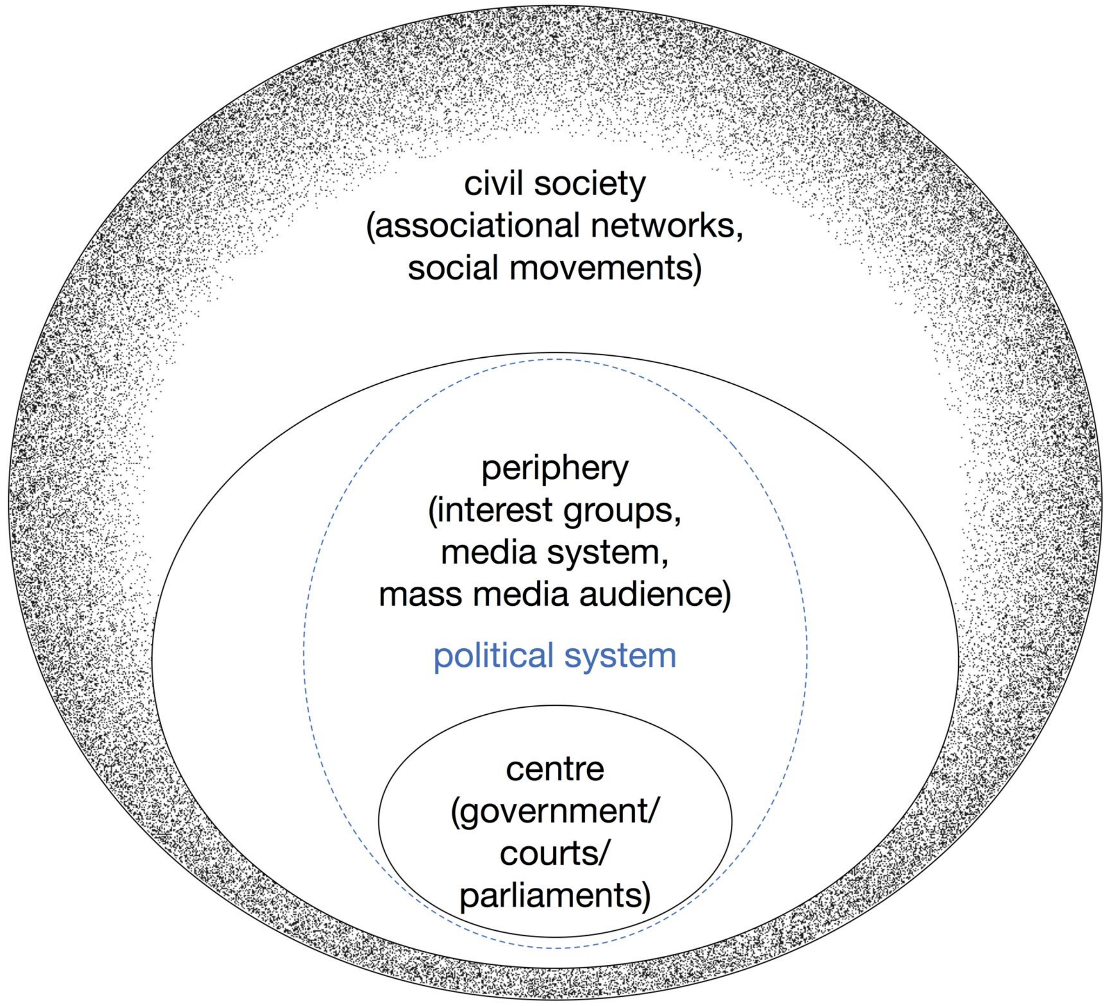
NEITHER QUANT NOR QUAL

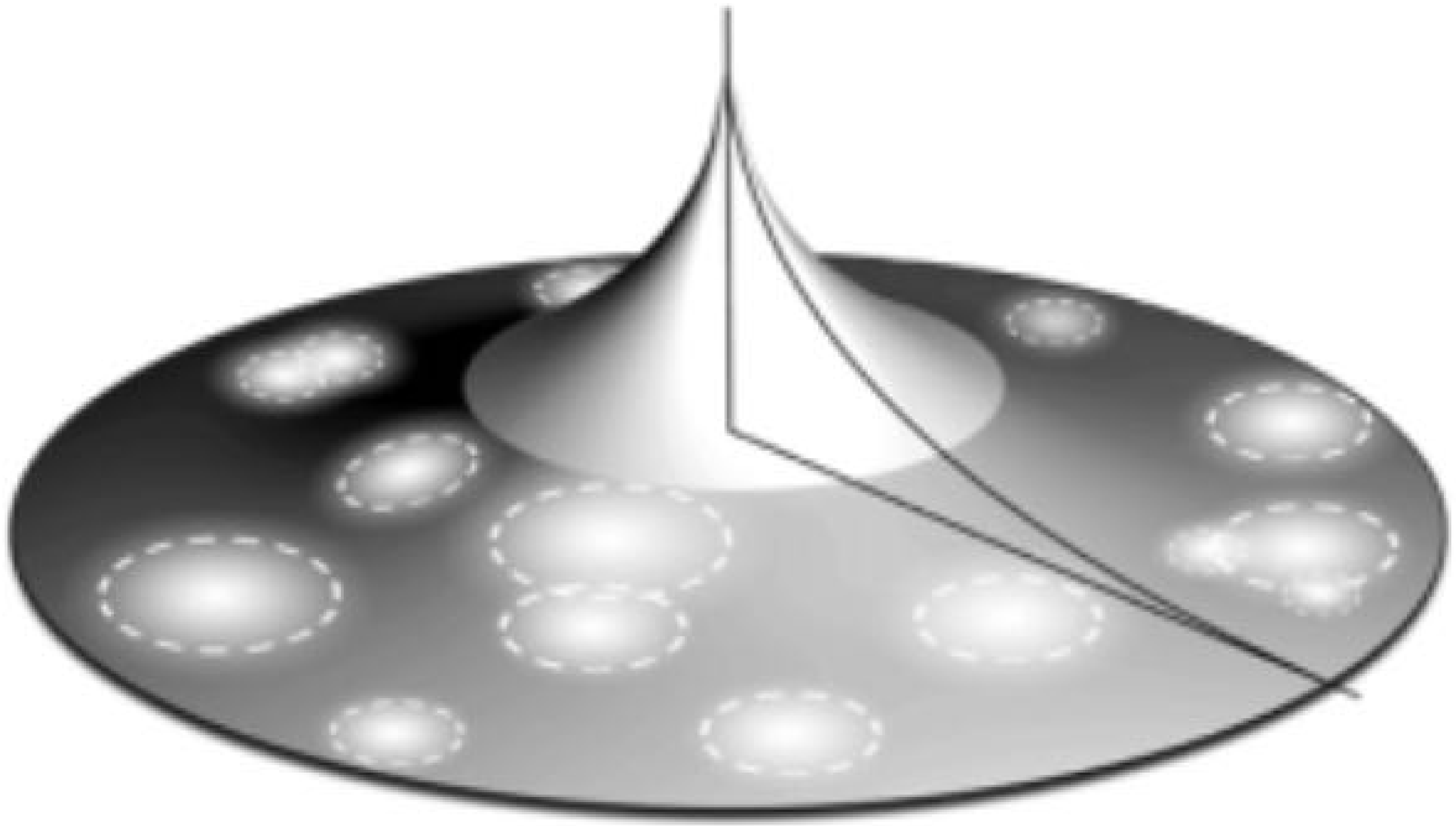
🤪

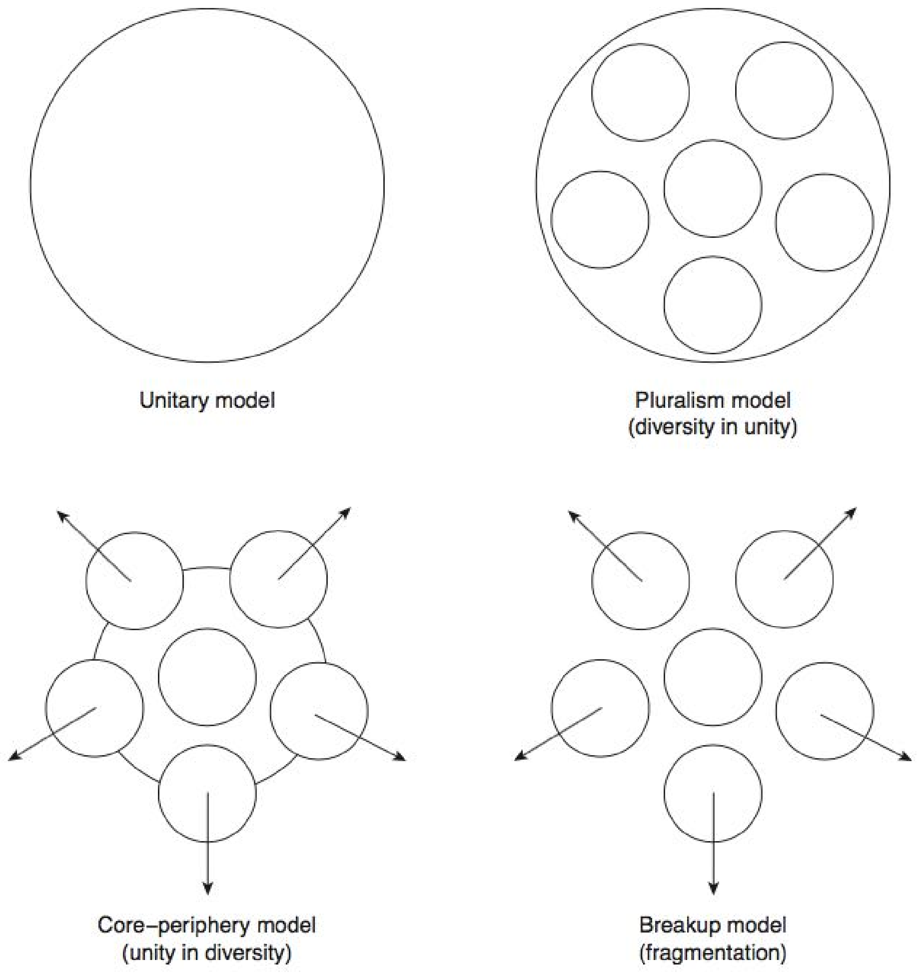
Well, this needs a whole lecture … this lecture

2. TODAY'S EXAMPLE: THE NETWORKED PUBLIC SPHERE

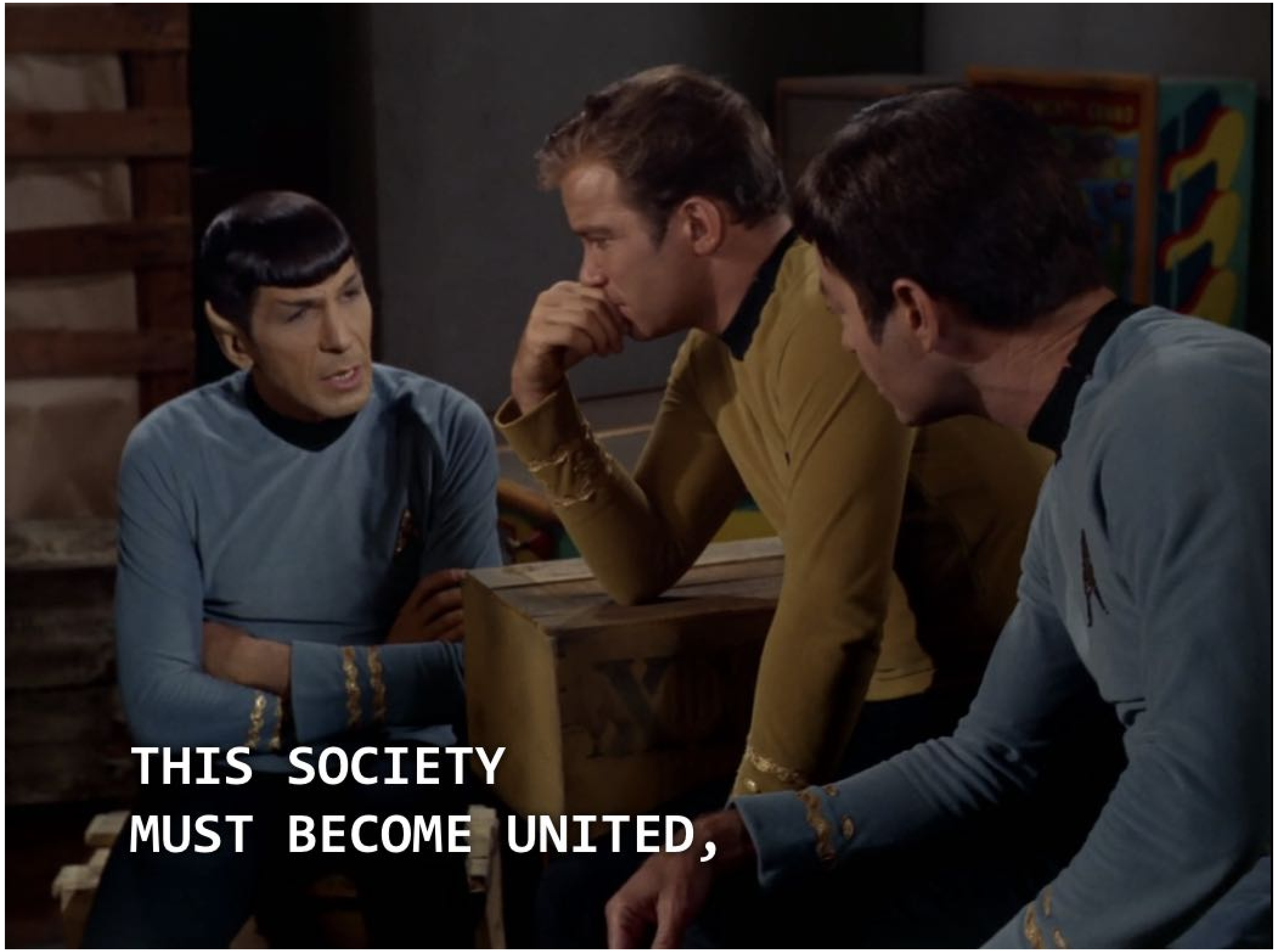
THE PUBLIC SPHERE

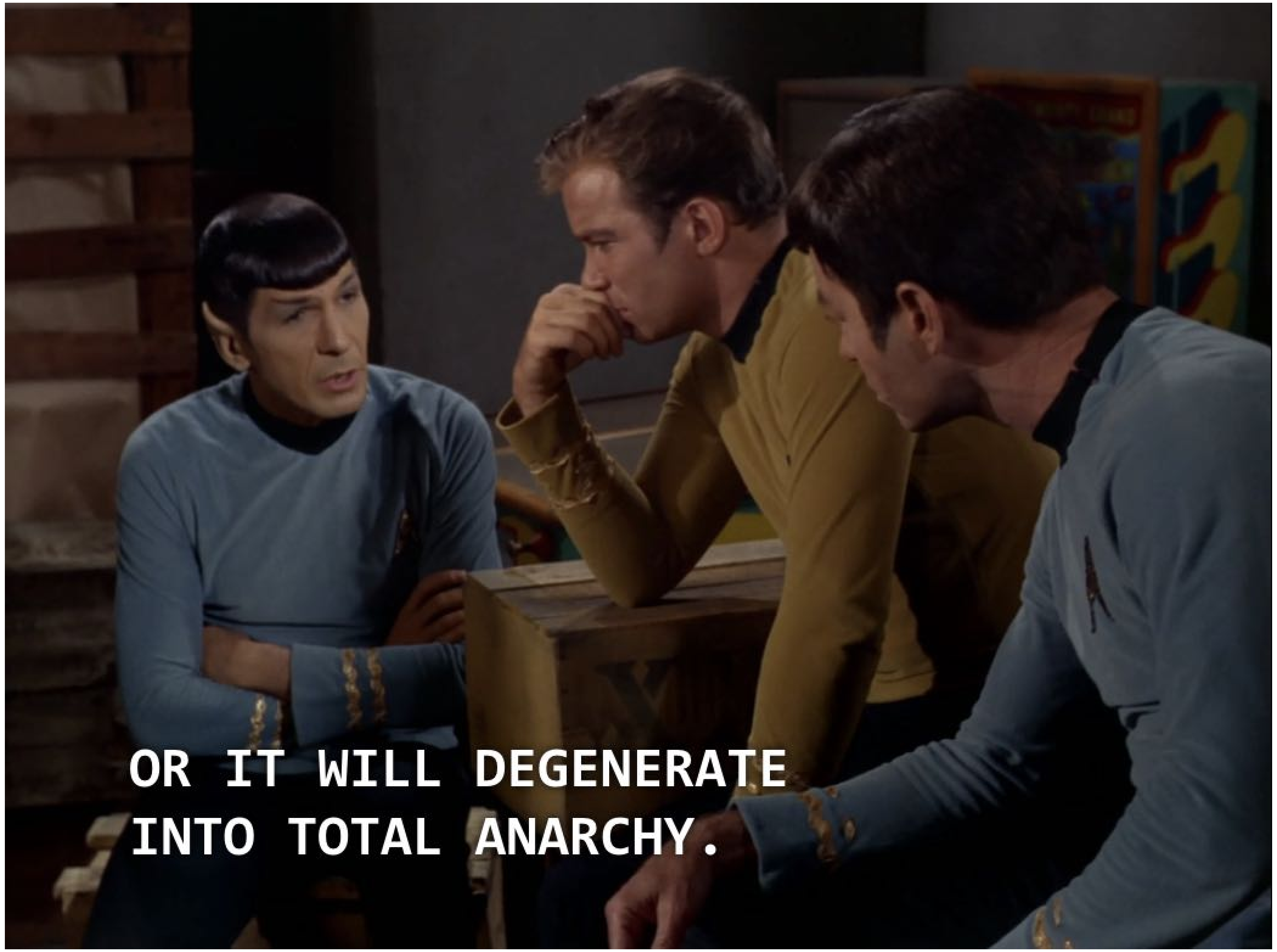




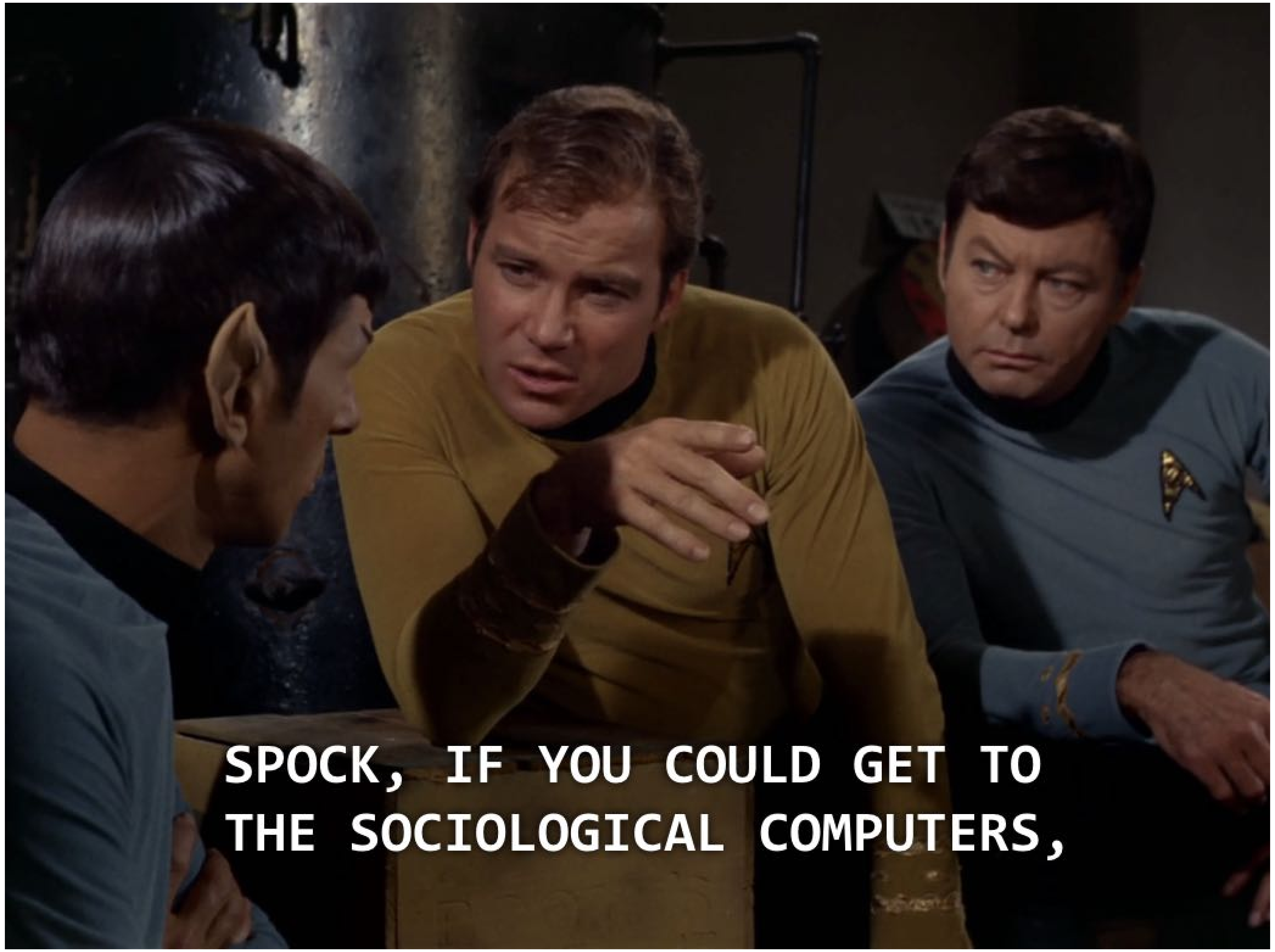


after Habermas (2006), from Bruns (2008), from McQuail (2010) respectively


THIS SOCIETY
MUST BECOME UNITED,


OR IT WILL DEGENERATE
INTO TOTAL ANARCHY.


AND WE'VE GOT
TO DO SOMETHING TO
STRAIGHTEN THIS MESS OUT.


SPOCK, IF YOU COULD GET TO
THE SOCIOLOGICAL COMPUTERS,


DO YOU THINK YOU
COULD FIND A SOLUTION?

3. ORGANISED COMPLEXITY AND NETWORK SCIENCE

ORGANISED COMPLEXITY

Three eras in the history of science (Warren Weaver, 1948):

- simplicity: two variables (e.g., length and force of a leaver) -> simple calculus
- disorganised complexity: averages, distributions (e.g., audience size and ratings) -> statistics
- organised complexity:

> *This was a new science focused on problems where the identity of the elements involved in a system, and their patterns of interactions, could no longer be ignored. … a new math needed to emerge. (Hidalgo, 2016, p. 2)*

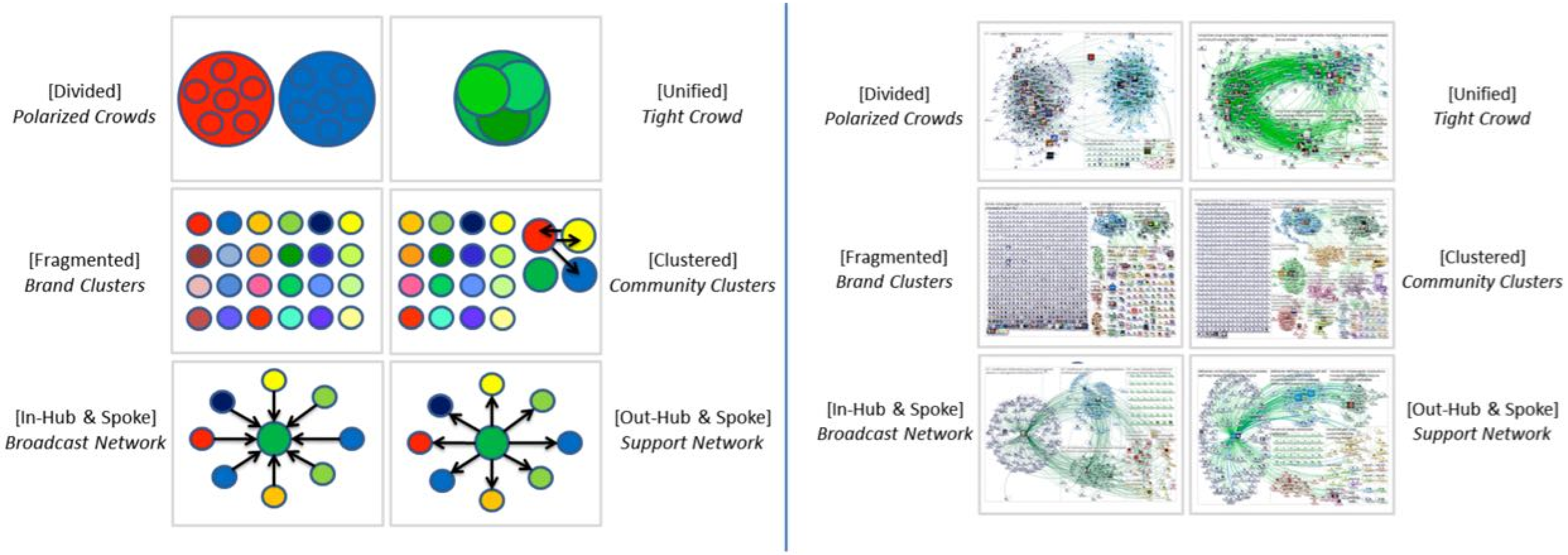
=> **Network Science**, Agent Based Modeling, Machine Learning, …

NETWORK SCIENCE

Networks:

- “mathematical objects that help us keep track of the identity of the elements involved in a system and their patterns of interaction”
- “ideal structures to describe problems of organized complexity” (Hidalgo, 2016, p. 2)

NETWORK SCIENCE



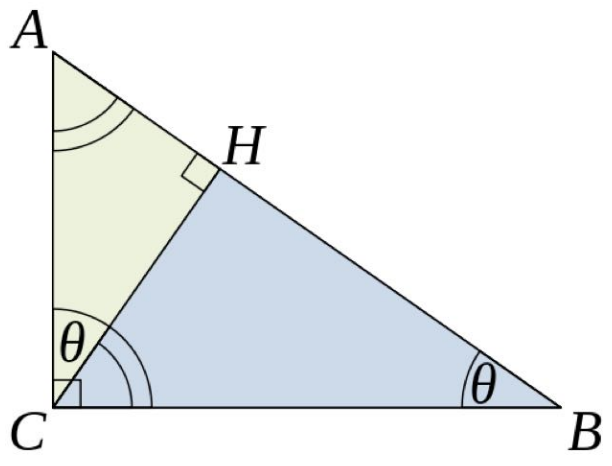
(Smith et al., 2014) (source: http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/)

PROBLEM: DISCIPLINARY DIVIDES

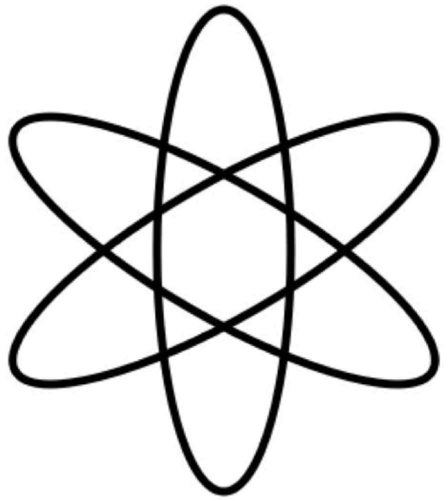
NETWORK SCIENCE

MEDIA STUDIES

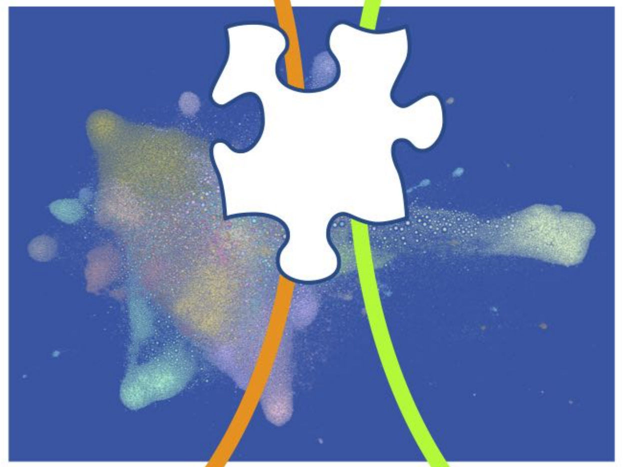
Mathematics



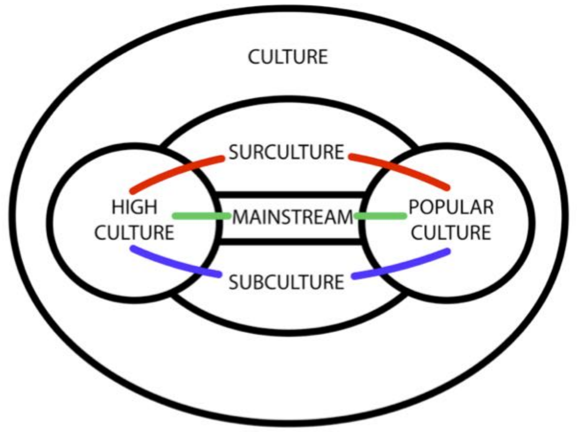
Natural Science



Social Sciences



Cultural Studies



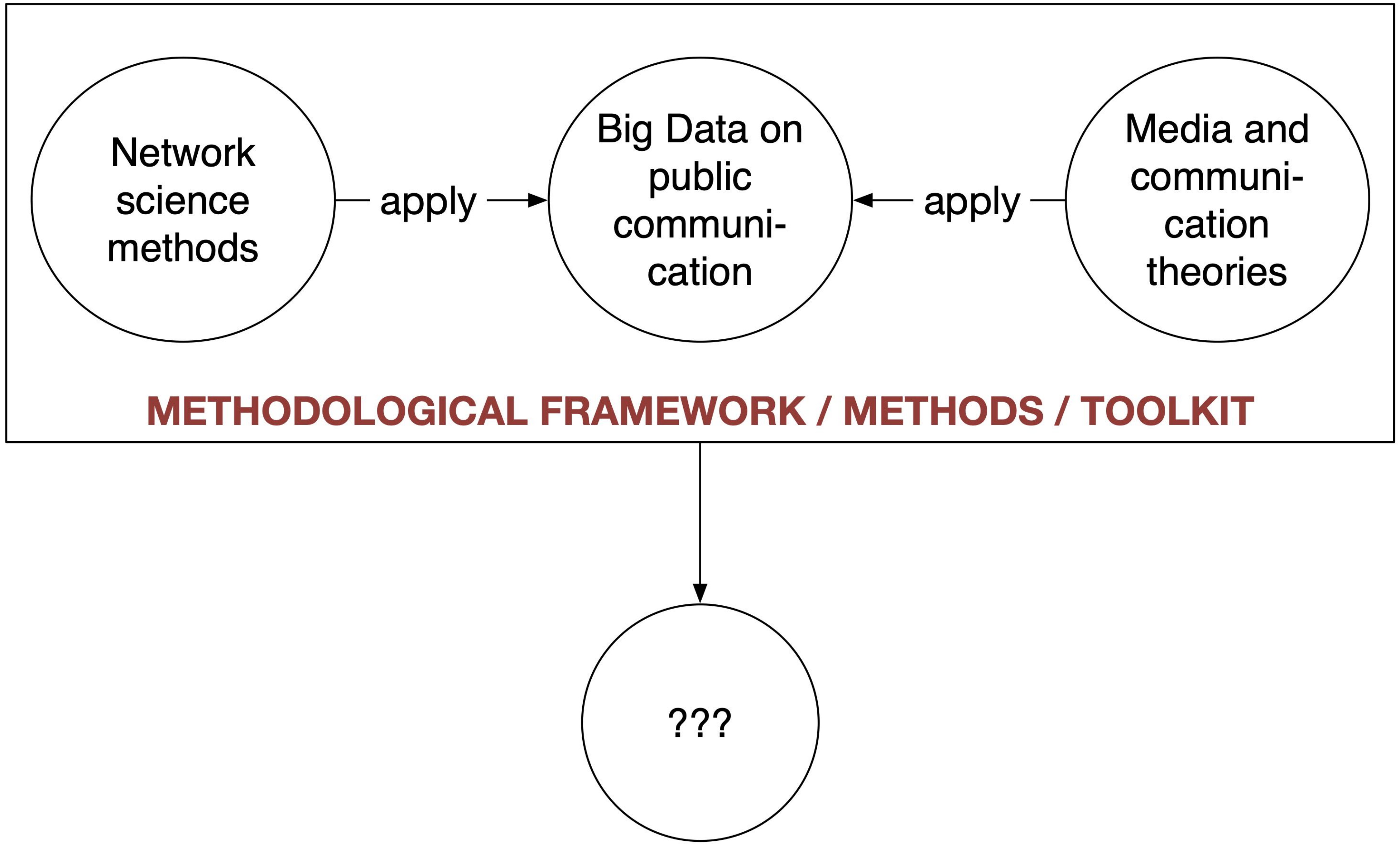
Arts

BIG DATA

Problems in epistemology, methodology, teleology, education, and research training

4. CHALLENGES IN NETWORK SCIENCE FOR COMPUTATIONAL COMMUNICATION SCIENCES

LACK OF A COMMON UNDERSTANDING OF METHODOLOGY AND COMMON GOALS



CHALLENGE 1/3: HOW TO TRANSLATE THEORY?

How can we interpret established concepts and theories regarding public communication from the social sciences and media and communication studies as theories about the structural dynamics of networks?

CHALLENGE 2/3: HOW TO APPLY NETWORK SCIENCE?

How can we apply network science methods to analyse the structural dynamics of networks in online public communication on a scale of whole societies from a perspective that tries to build upon, reject, or extend traditional media and communication studies theory?

CHALLENGE 3/3: BUT WHY?

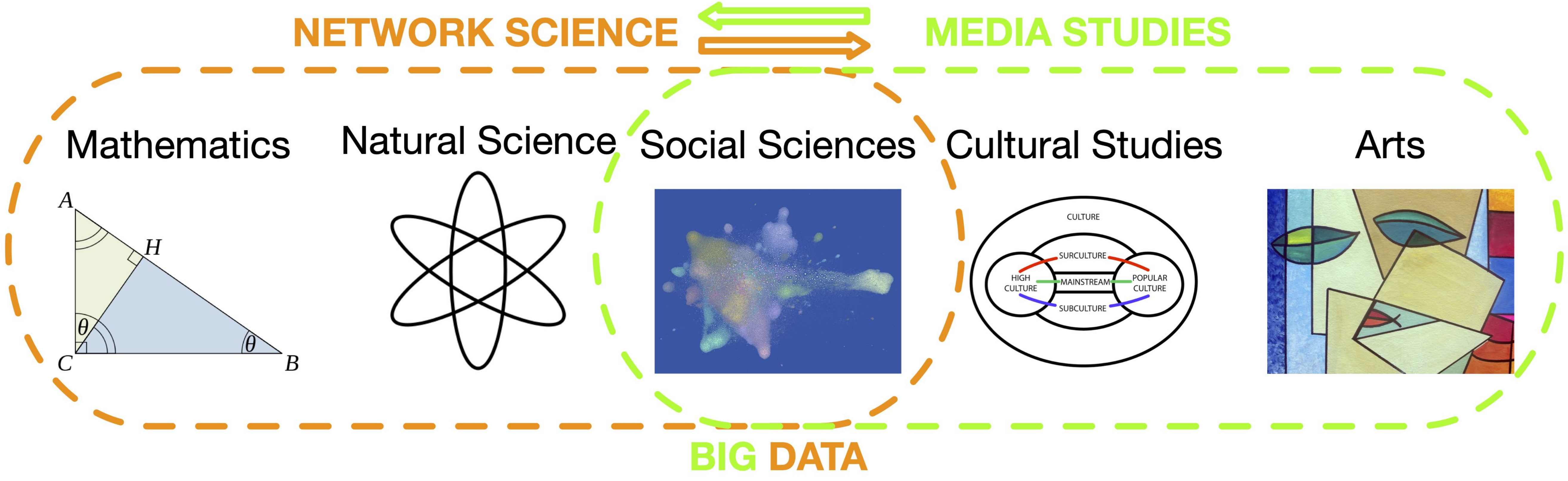
Can this analysis be validated by its utility in confirming established theories or by leading to new theories about online public communication and the public sphere?

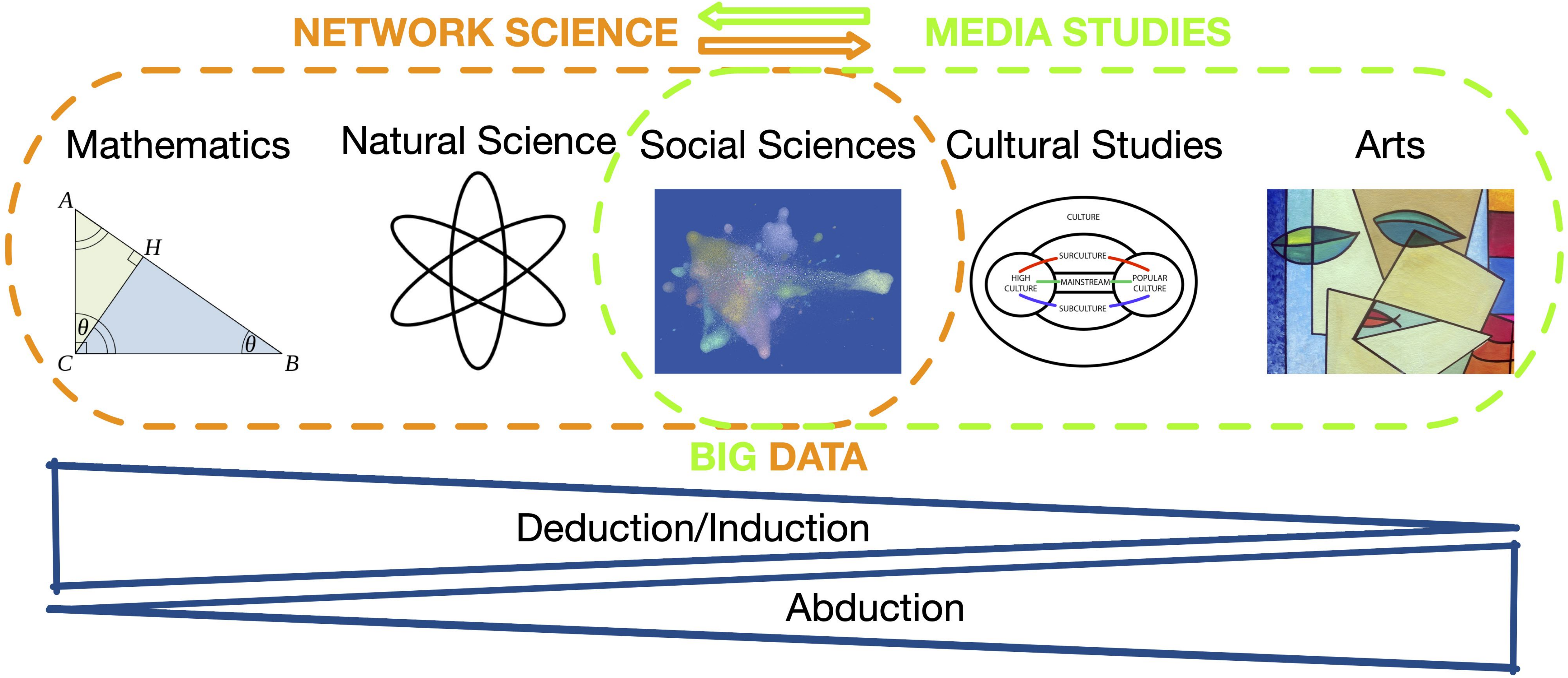
5. A POSSIBLE METHODOLOGICAL FRAMEWORK

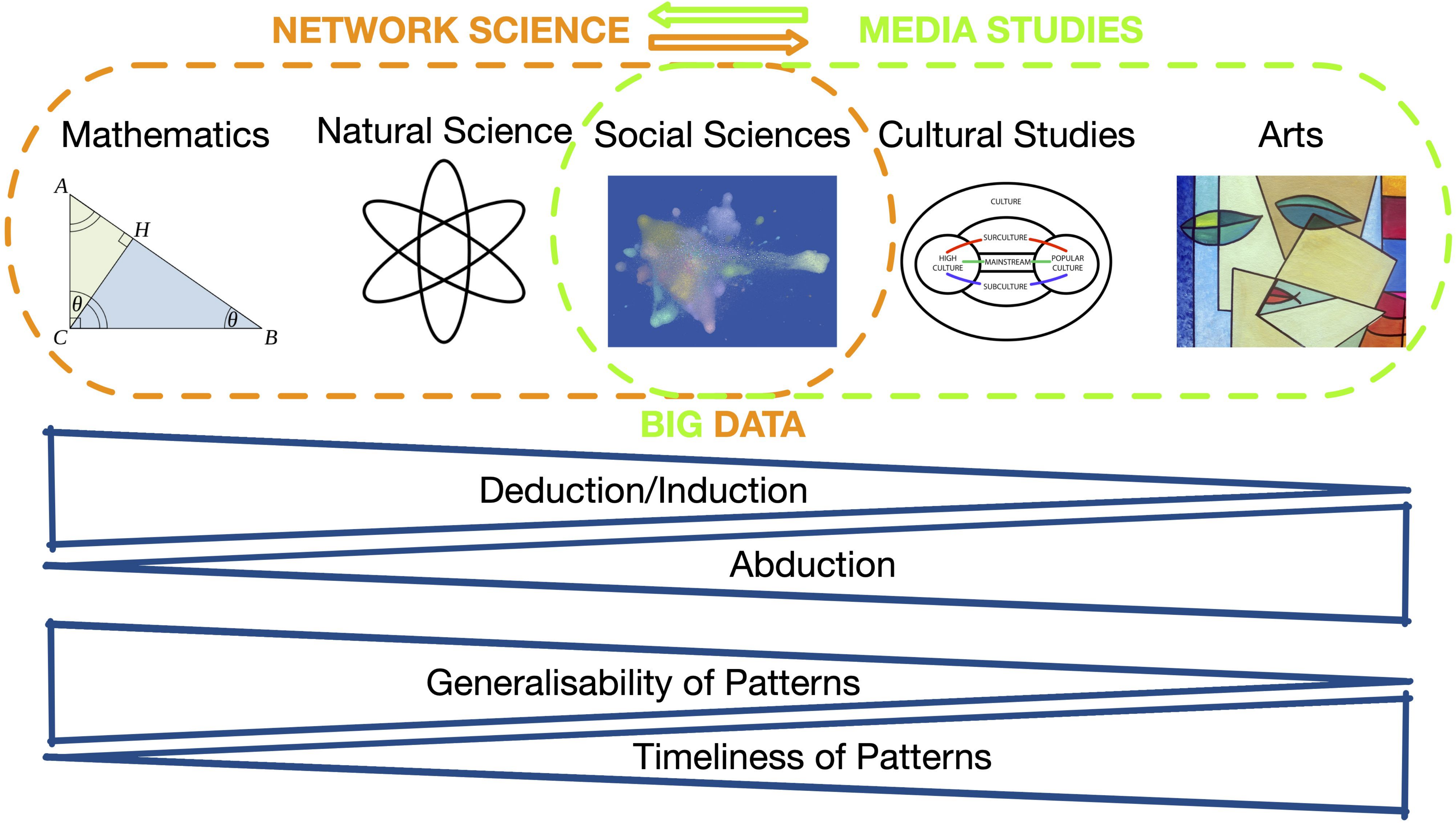
A POSSIBLE METHODOLOGICAL FRAMEWORK

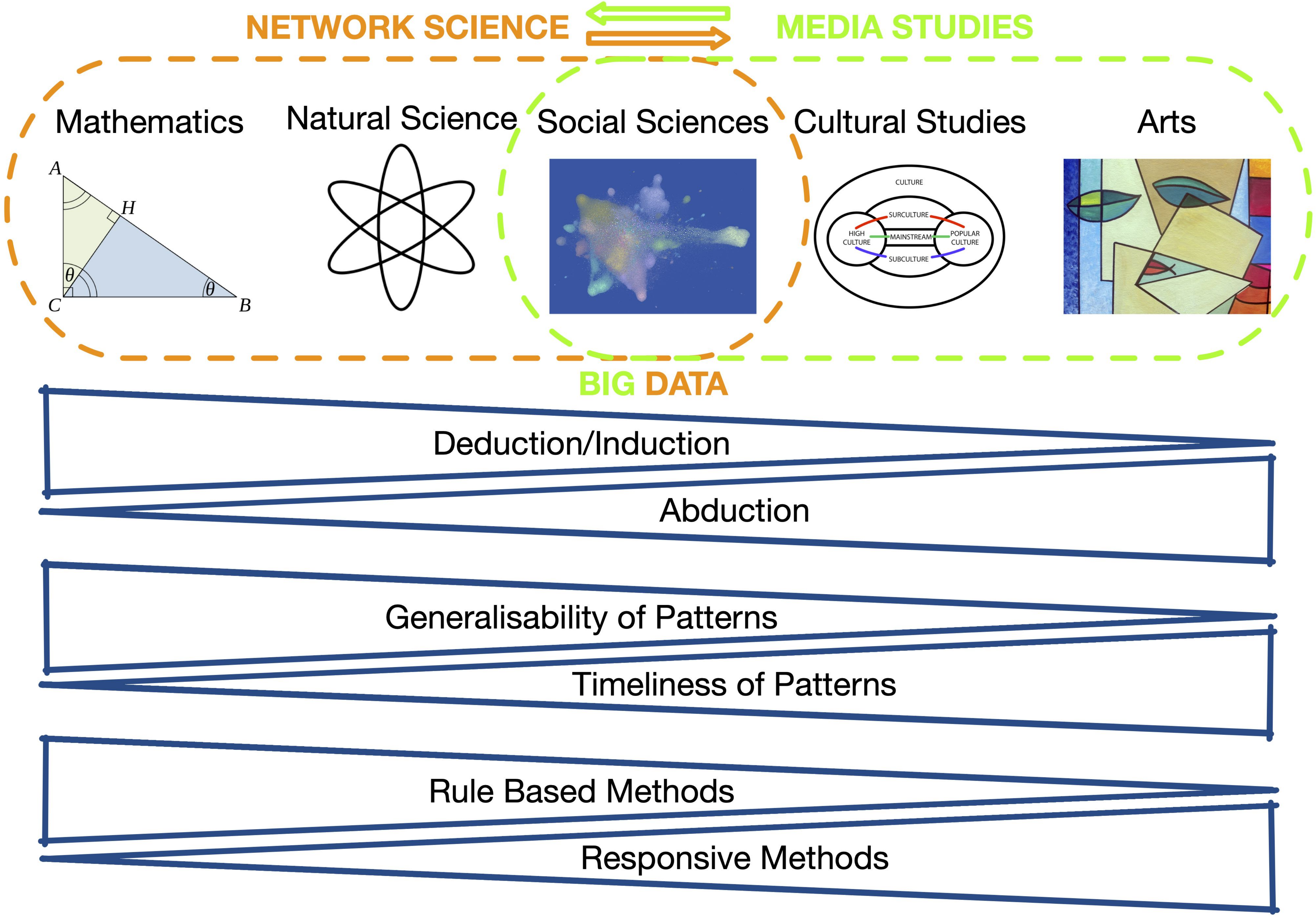
Synthesising ideas by

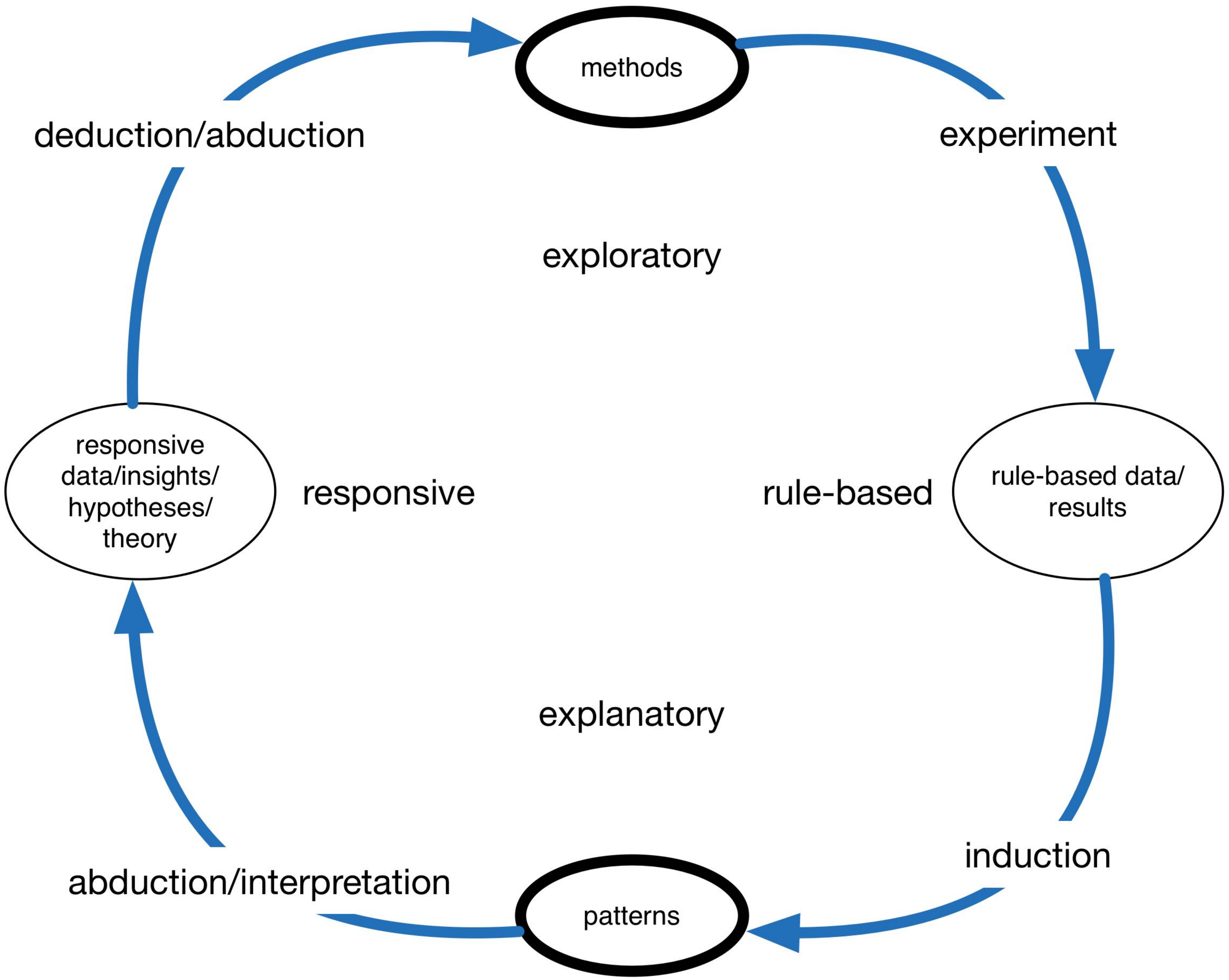
- Berry (2011) on the computational turn in the humanities;
- Halavais (2015) on Big Data interlinking the general with the particular;
- Sechrest & Sidani (1995) on a "false dichotomy" of qual vs quant
- Onwuegbuzie & Leech (2005) on pragmatism as a necessity;
- Dixon (2012) on an epistemology of patterns;


NETWORK SCIENCE
MEDIA STUDIES
Mathematics
Natural Science
Social Sciences
Cultural Studies
Arts
A
H
C
B
θ
θ
CULTURE
SURCULTURE
HIGH CULTURE
MAINSTREAM
POPULAR CULTURE
SUBCULTURE
BIG DATA


NETWORK SCIENCE
MEDIA STUDIES
Mathematics
Natural Science
Social Sciences
Cultural Studies
Arts
A
H
C
B
θ
θ
CULTURE
SURCULTURE
HIGH CULTURE
MAINSTREAM
POPULAR CULTURE
SUBCULTURE
BIG DATA
Deduction/Induction
Abduction


NETWORK SCIENCE
MEDIA STUDIES
Mathematics
Natural Science
Social Sciences
Cultural Studies
Arts
A
H
C
B
θ
θ
CULTURE
SURCULTURE
HIGH CULTURE
MAINSTREAM
POPULAR CULTURE
SUBCULTURE
BIG DATA
Deduction/Induction
Abduction
Generalisability of Patterns
Timeliness of Patterns


NETWORK SCIENCE
MEDIA STUDIES
Mathematics
Natural Science
Social Sciences
Cultural Studies
Arts
A
H
C
B
θ
θ
CULTURE
SURCULTURE
HIGH CULTURE
MAINSTREAM
POPULAR CULTURE
SUBCULTURE
BIG DATA
Deduction/Induction
Abduction
Generalisability of Patterns
Timeliness of Patterns
Rule Based Methods
Responsive Methods


methods
deduction/abduction
experiment
exploratory
responsive data/insights/ hypotheses/ theory
responsive
rule-based
rule-based data/ results
explanatory
induction
abduction/interpretation
patterns

6. EXAMPLE 1: MEASURING COMMUNICATION CASCADES

CASES

Three media items disseminated on Twitter, related to acute events:

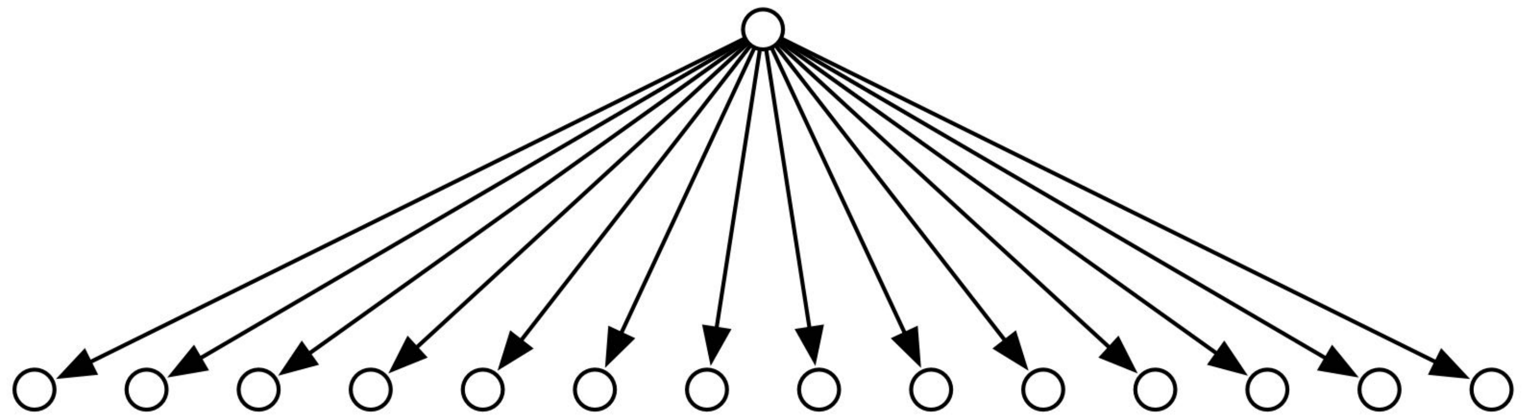
- #sydneysiege – most used
- #illridewithyou – perceived as most 'viral'
- link to a petition to repeat the Brexit referendum – interesting for comparison

WHAT IS VIRALITY (AKA CONTAGION)?

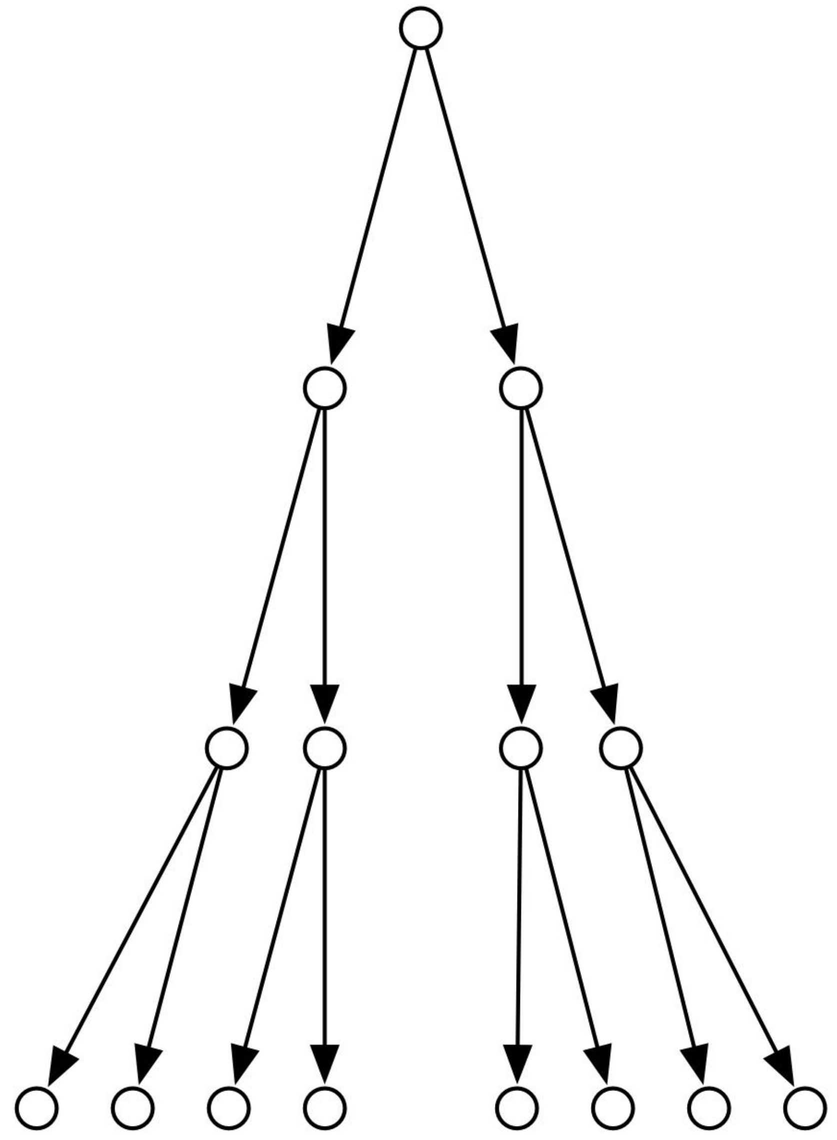
Three perspectives:

- virality of an item
- virality of the structure of a diffusion tree
- qualitative differences: simple (one exposure) vs. complex (threshold models) contagion

STRUCTURAL VIRALITY (AVG. SHORTEST PATH LENGTH)



Broadcast

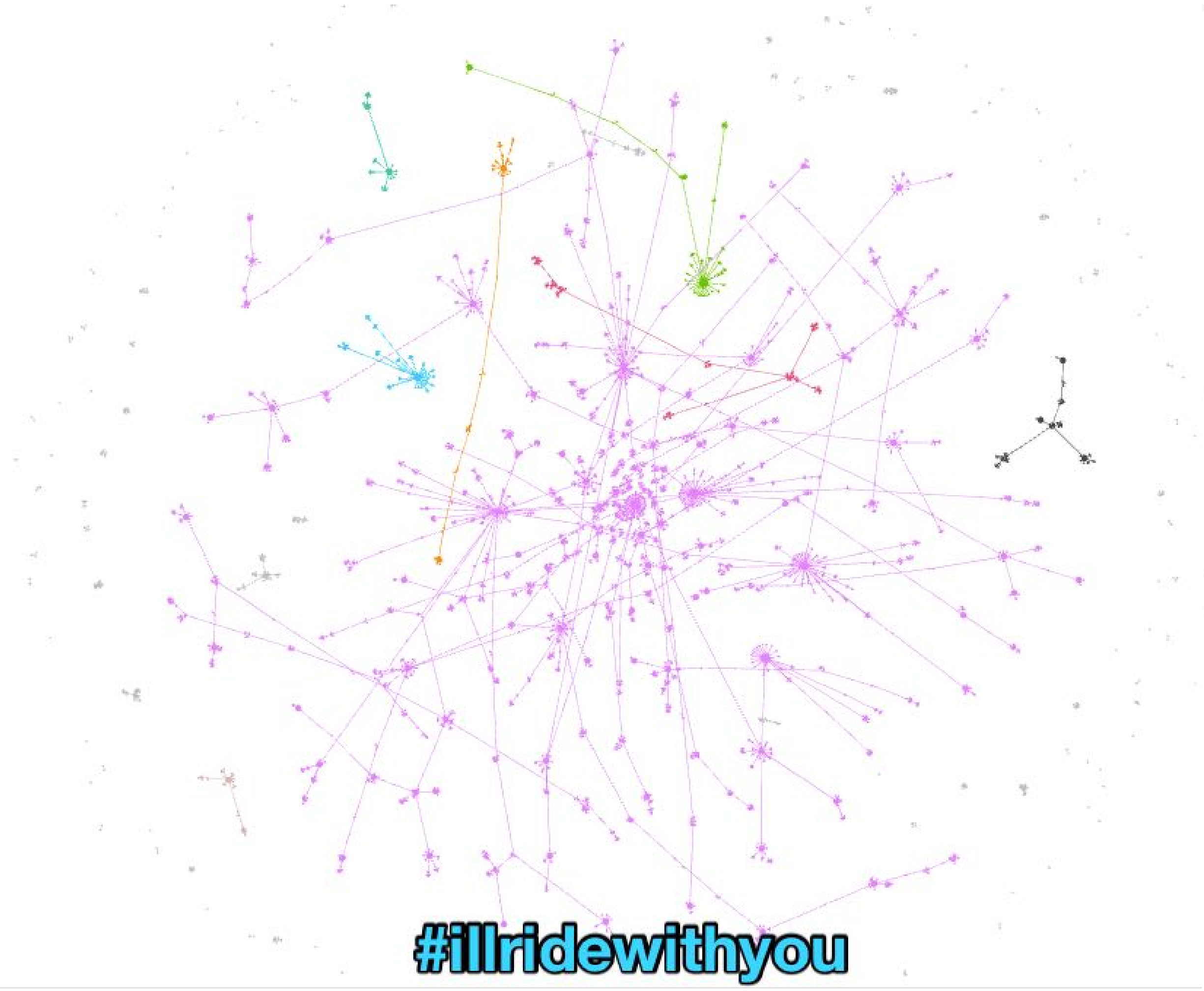


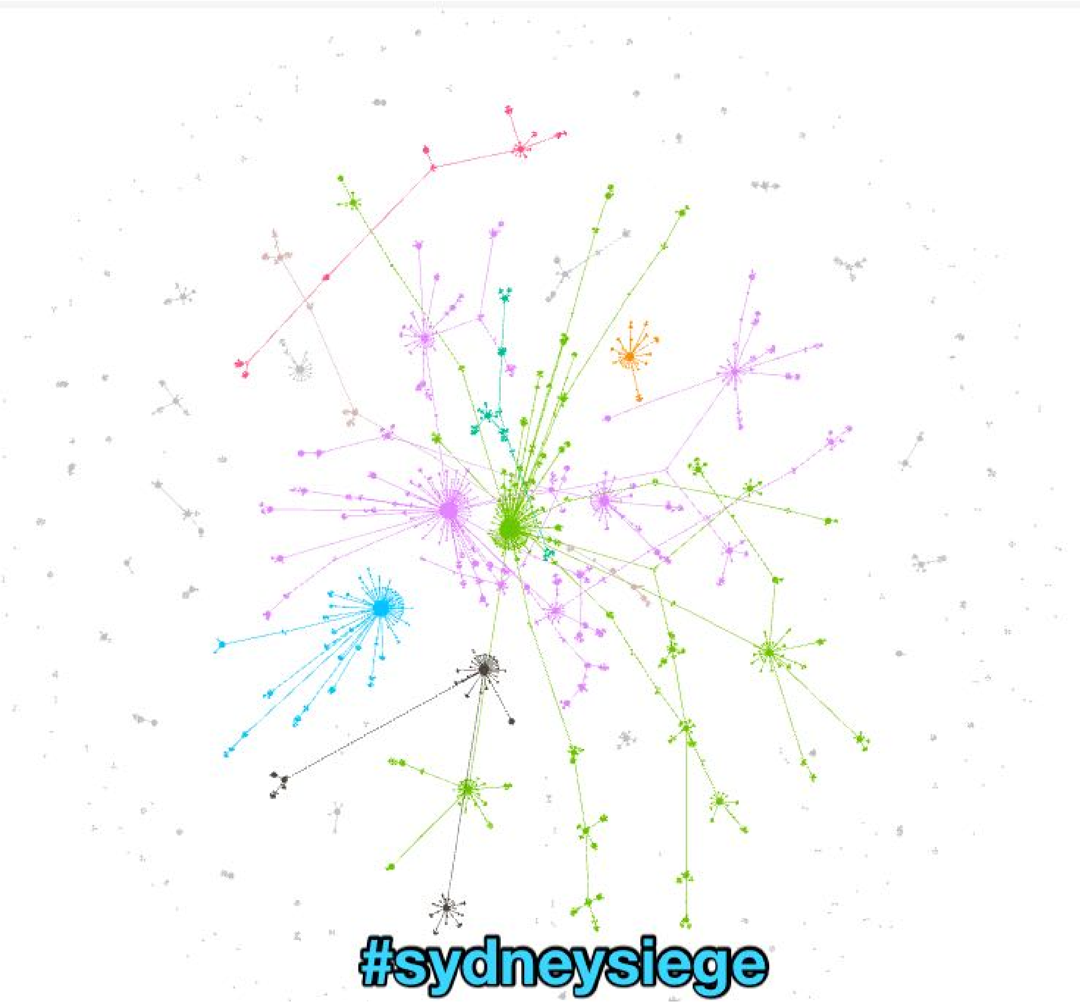
Viral Cascade

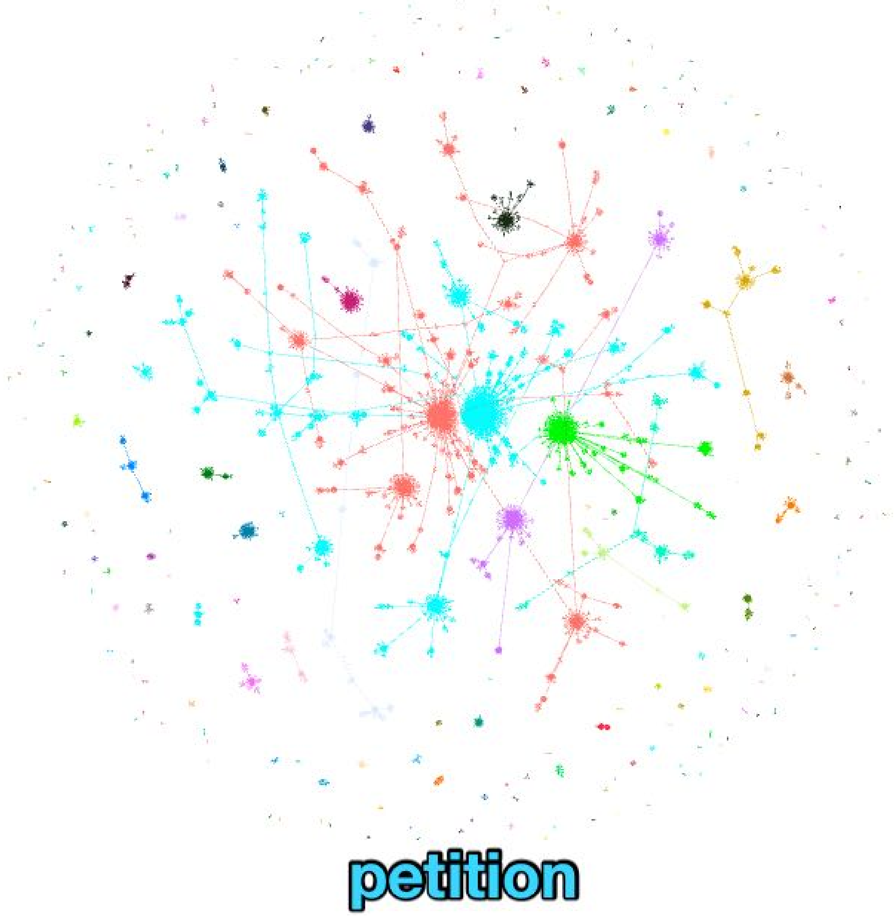
after Goel et al. (2015)

SELECTED RESULTS

CONNECTED COMPONENTS



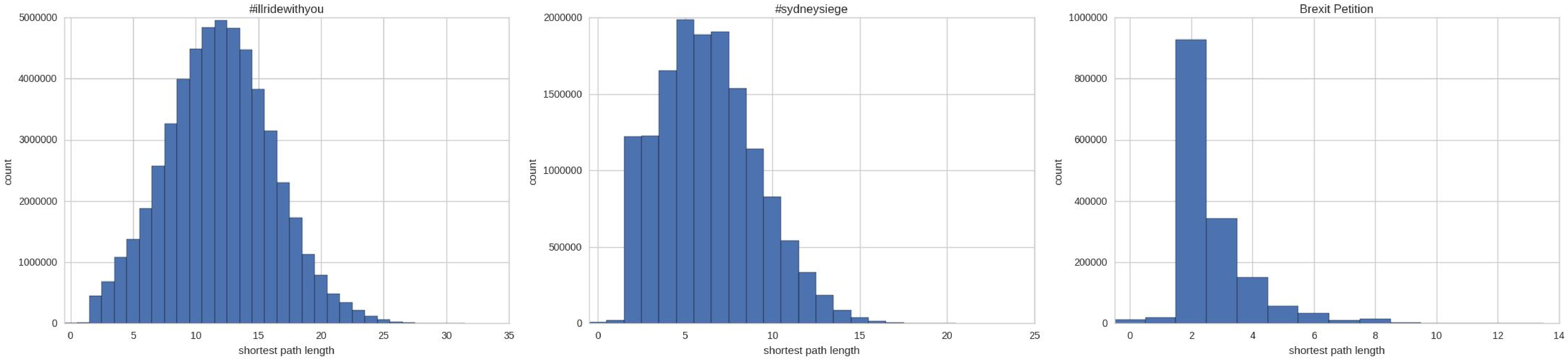




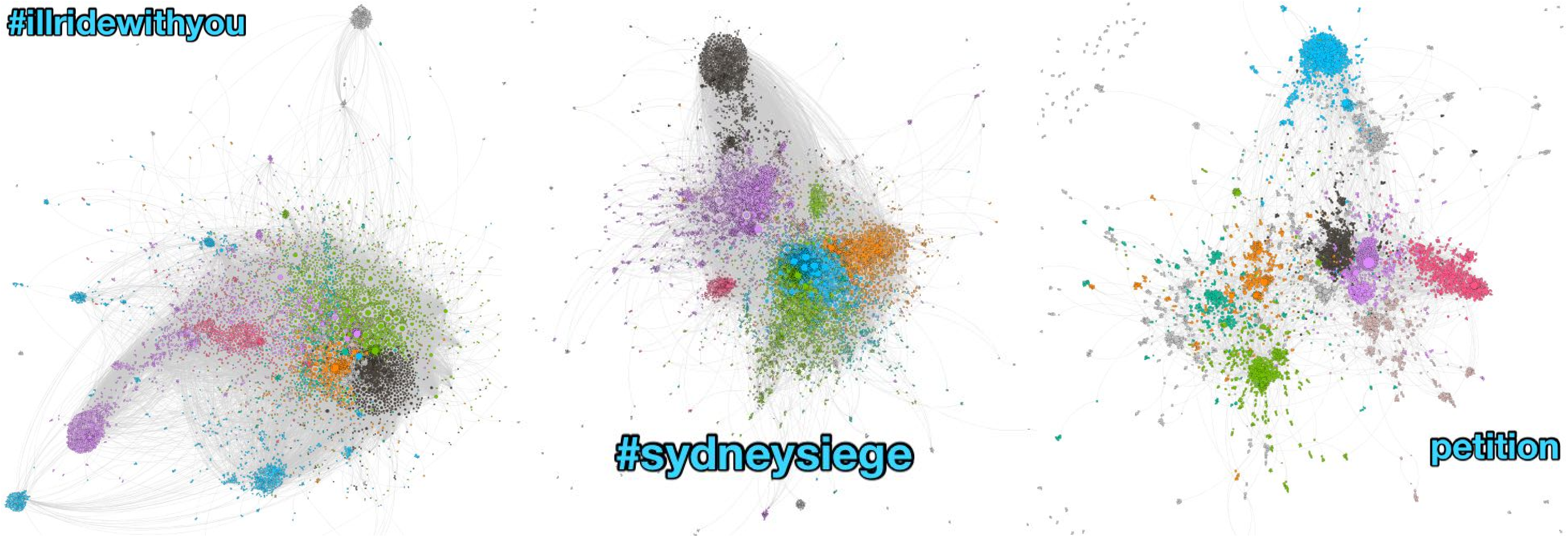
Diffusion trees reconstructed from **first 10,000 participating accounts** in each case.

- reconstructed diffusion network, based on assumption that the last exposure triggered the reshare
- colour by connected component
- decreasing dominance of largest components
- increasing number of components, equivalent to number of entry points -> more viral, more 'organic growth', less viral, more 'outside influence'

STRUCTURAL VIRALITY (AVG. SHORTEST PATH LENGTH)

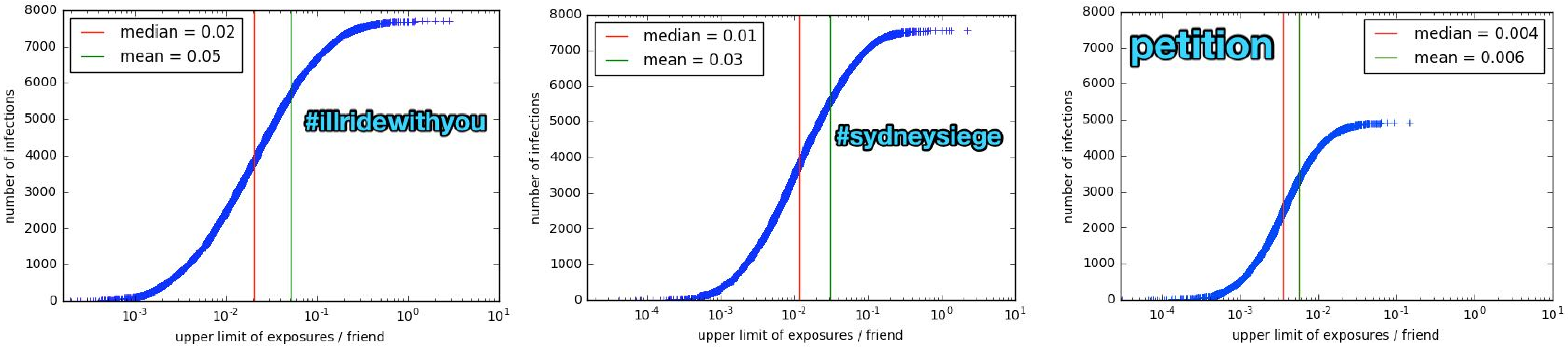


INFLUENCE NETWORK



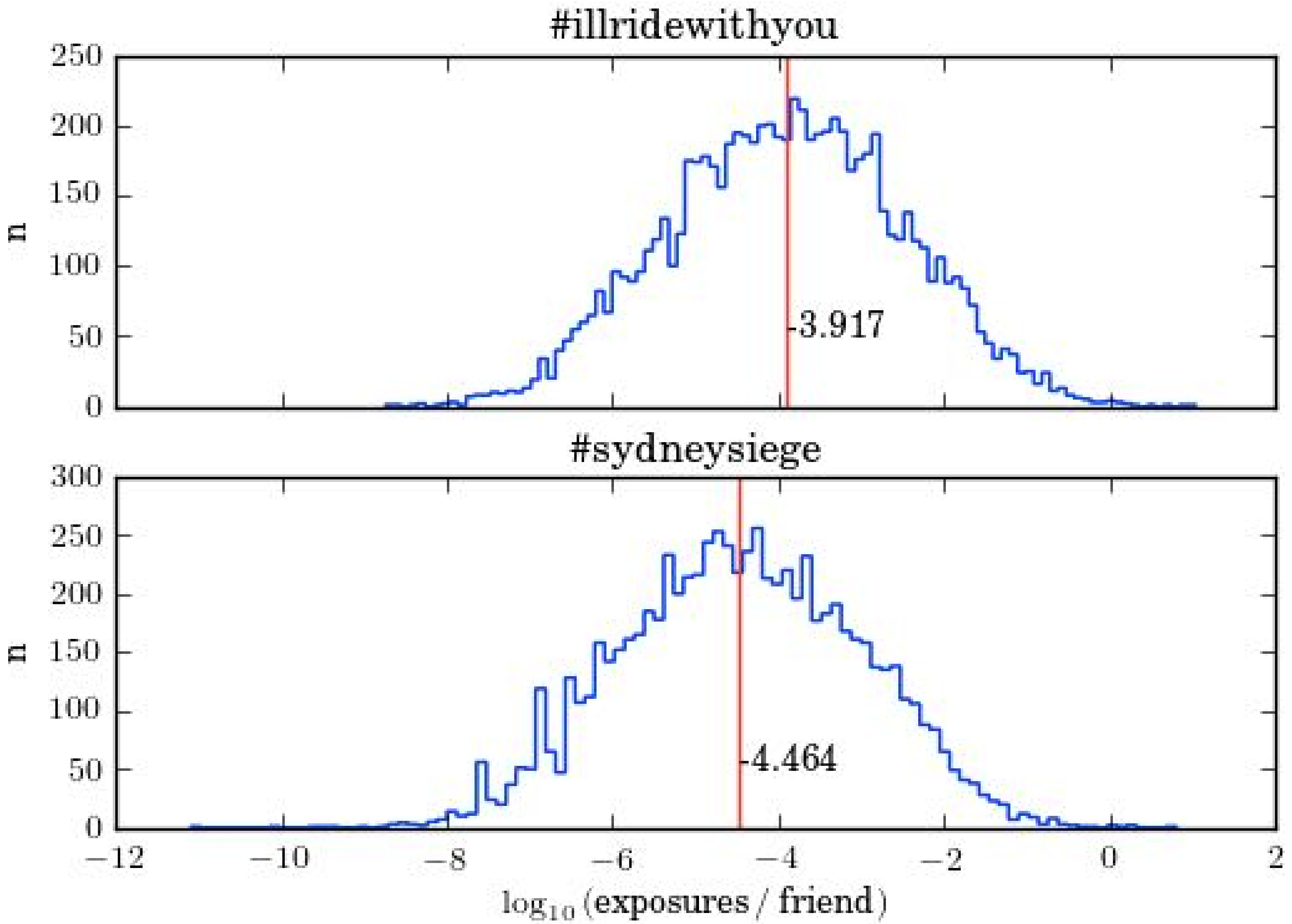
- if I followed another account + this account posted item before I did -> connection
- gives up assumption of single exposure -> enables to analyse possible **complex contagion**
- observation: hashtags both more clustered -> denser underlying community structure in form of follows

COMPLEXITY OF CONTAGION/EXPOSURES PER FRIEND (FOLLOWEE)



- the number of infections increases with the exposures to the respective item per friend
- could be understood as some form of 'peer pressure'
- average highest for illridewithyou (5 exposures per 100 friends)
- middle sydneysiege (3 exposures per 100 friends)
- far below for link (6 per 1000 friends)

COMPLEXITY OF CONTAGION/EXPOSURES PER FRIEND (FOLLOWEE)




```
illridewithyou mean:  -3.91673145599
sydneysiege mean:  -4.46387012522
t =  23.6683039501 , p =  1.18502599027e-121
```

Limitation: the effect is hard to separate from the 'organic grown' nature of #illridewithyou

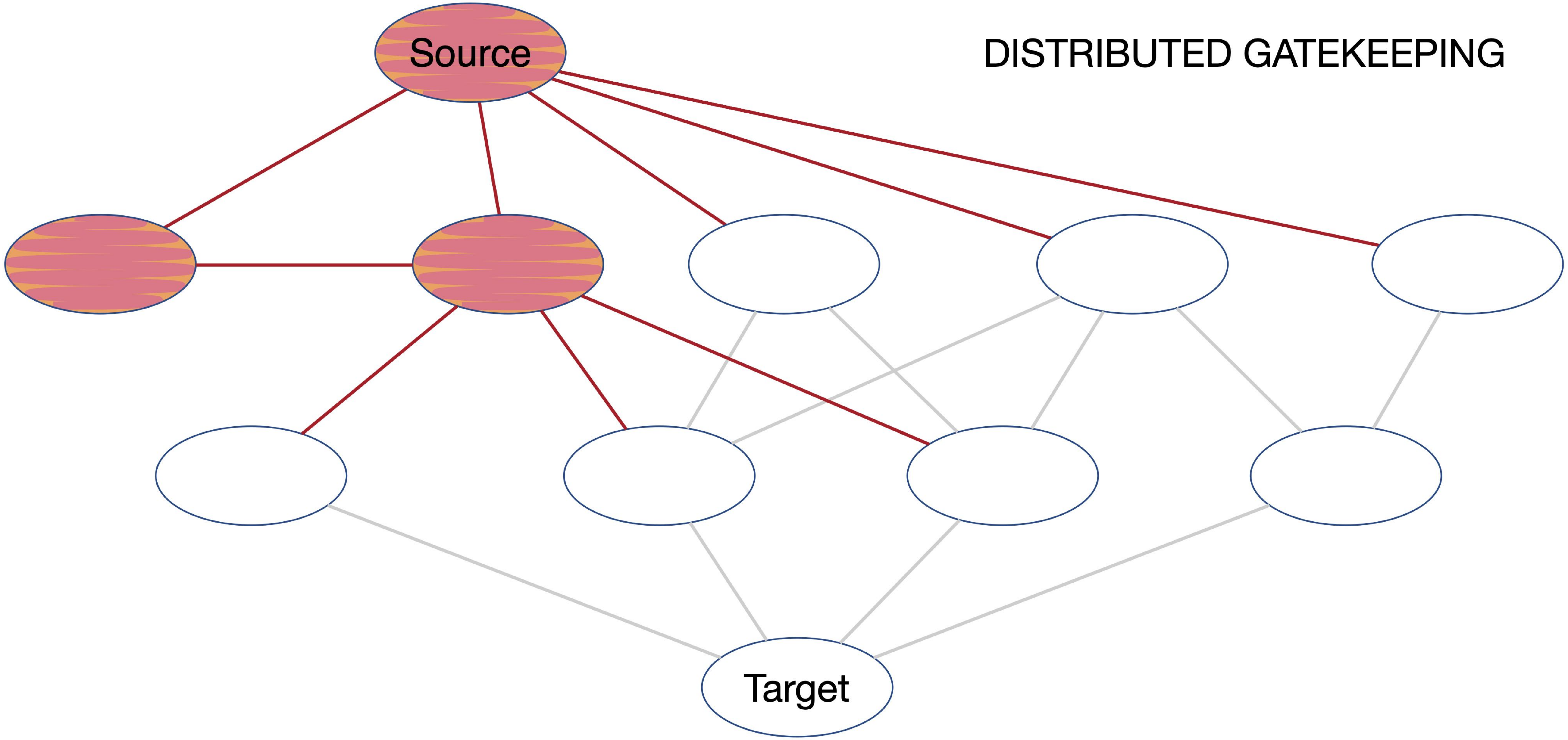
SUMMARY OF RESULTS

	#illridewithyou	#sydneysiege	petition
outside influence	+	++	+++
structural virality	+++	++	+
clustering*	+++	+++	+
impact of single nodes*	+++	++	+
timescales*	+	++	+++
'virality'	**high**	**medium**	**low**
complexity of contagion	+++	++	+

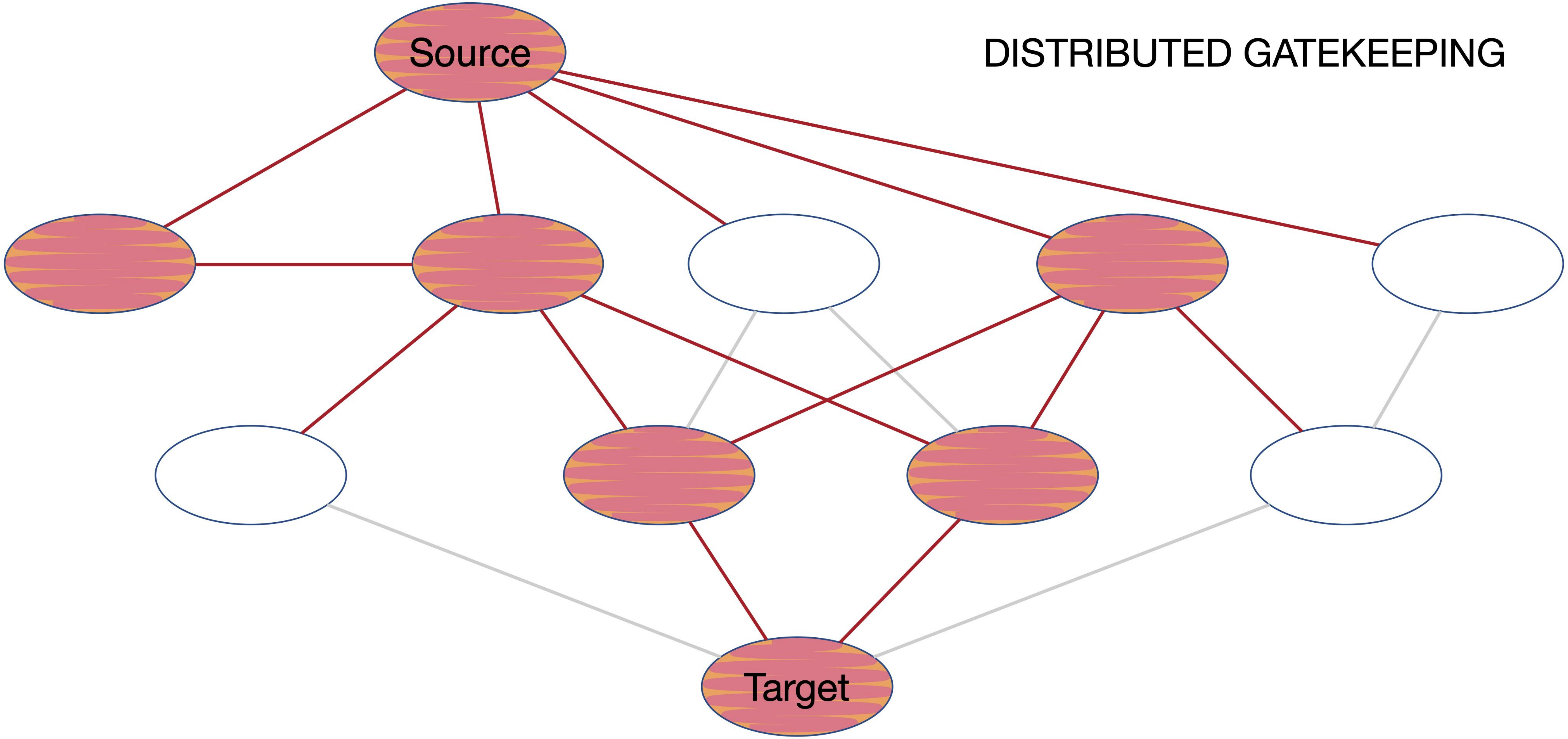
*see Münch (2019)

APPLICATION TO THEORY (EXAMPLE): GATEKEEPING

DISTRIBUTED GATEKEEPING



- concept of distributed gatekeeping in networks, based on limited attention space
- assumption for example: general average threshold of 2 to reshare an item

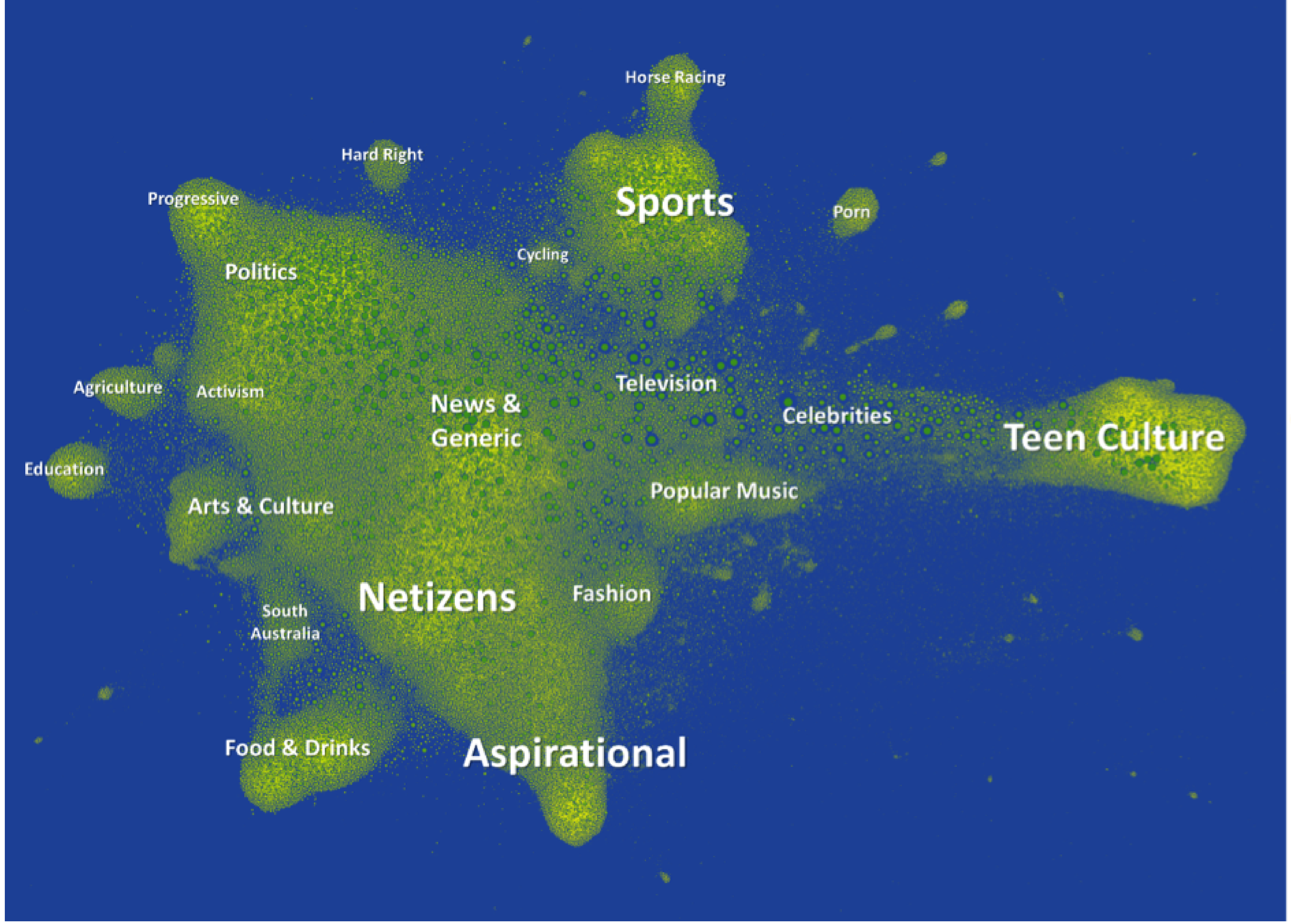


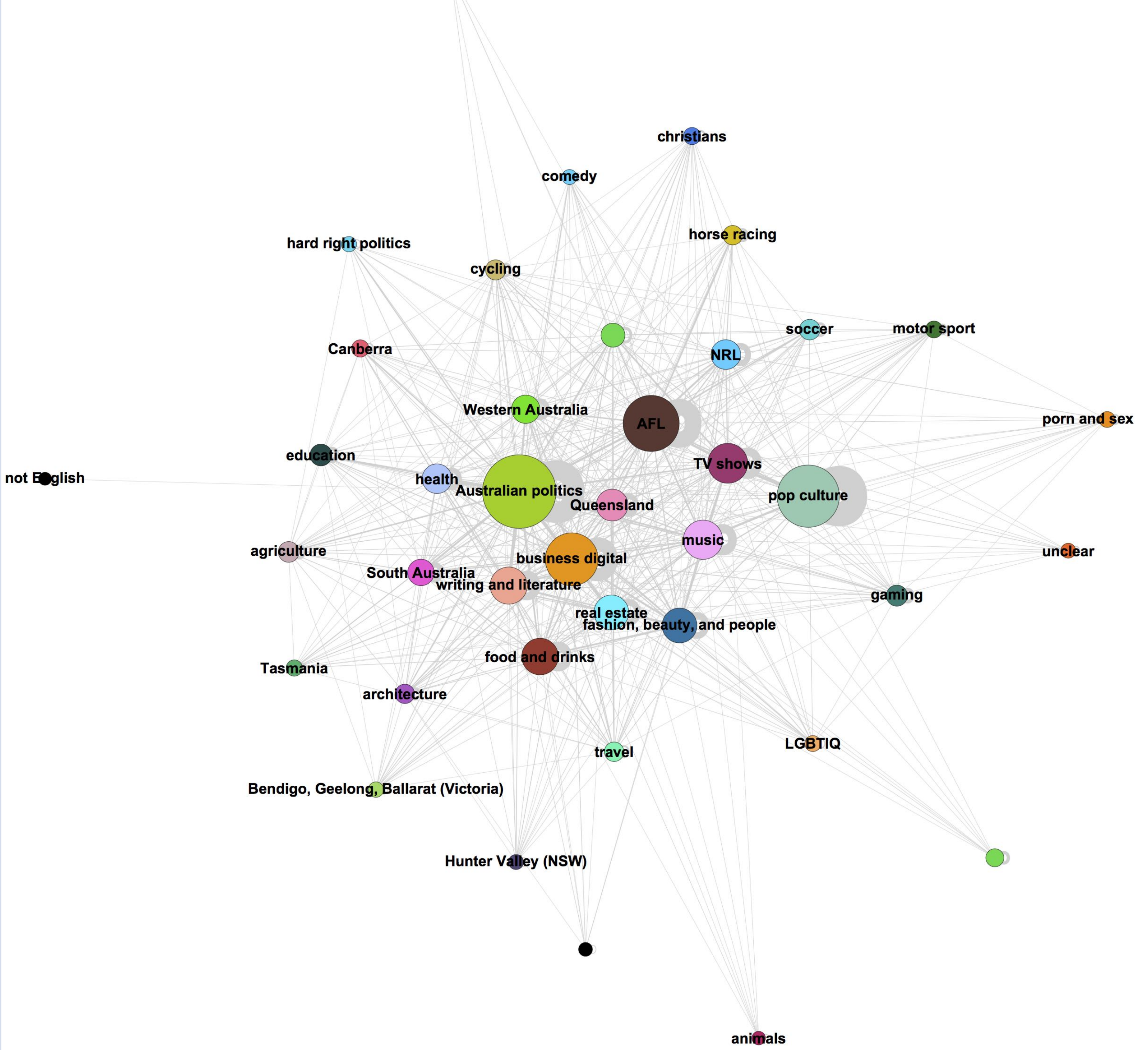
DISTRIBUTED GATEKEEPING

One node acts as a quasi-gatekeeper to get the message through to the target, based on properties of the node (lower threshold), but also on position in the network (node on the far right would not succeed)

7. EXAMPLE 2: MAPPINGS OF A PUBLIC SPHERE

PART 1: THE QUASI-STANDARD – COMMUNITIES BASED ON INNER DENSITY (MODULARITY MAXIMISATION)





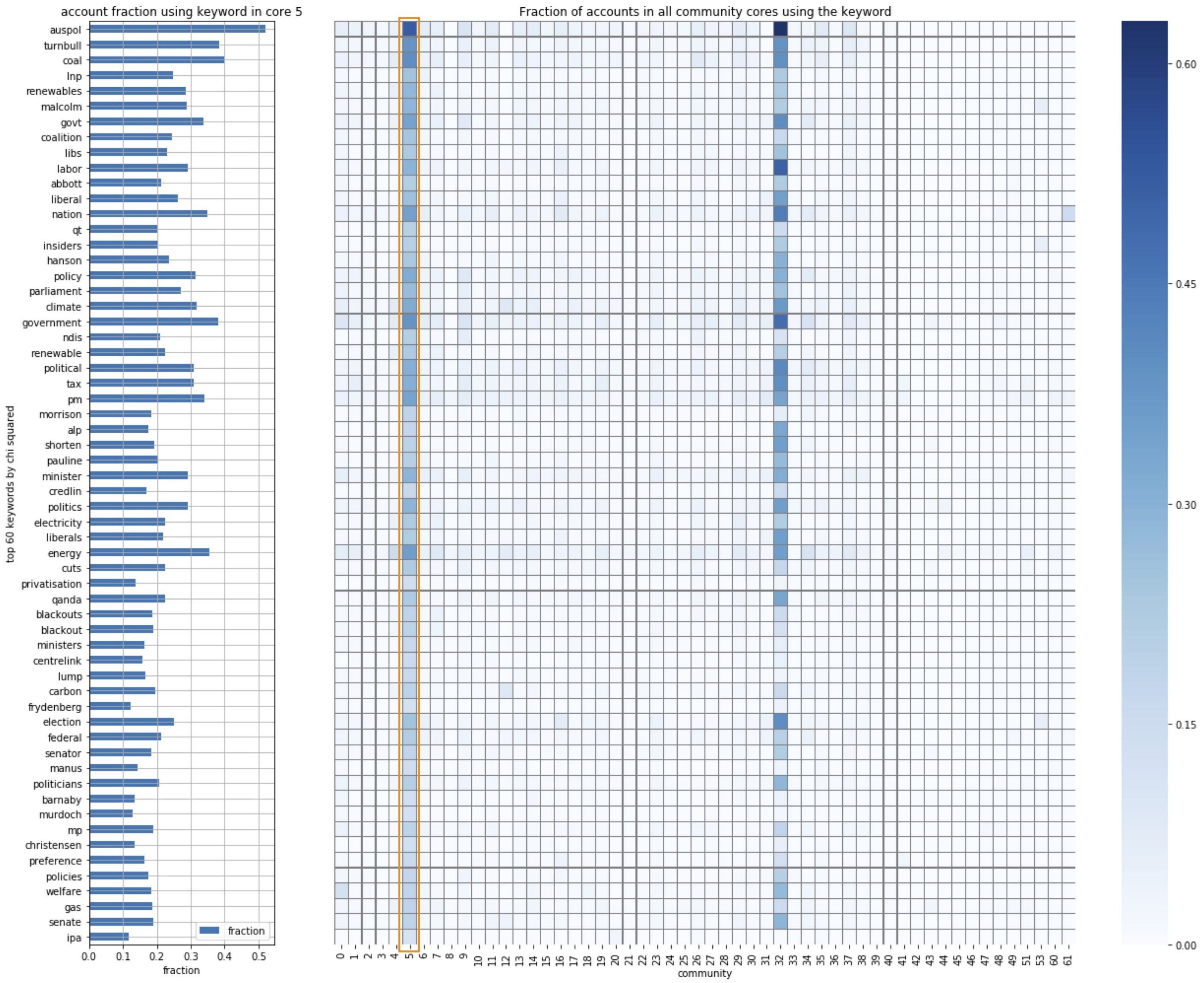
left (Bruns et al., 2017): Australian follow network filtered for **accounts with over 1000 followees + followers**. Labels based on **close reading of profile descriptions**

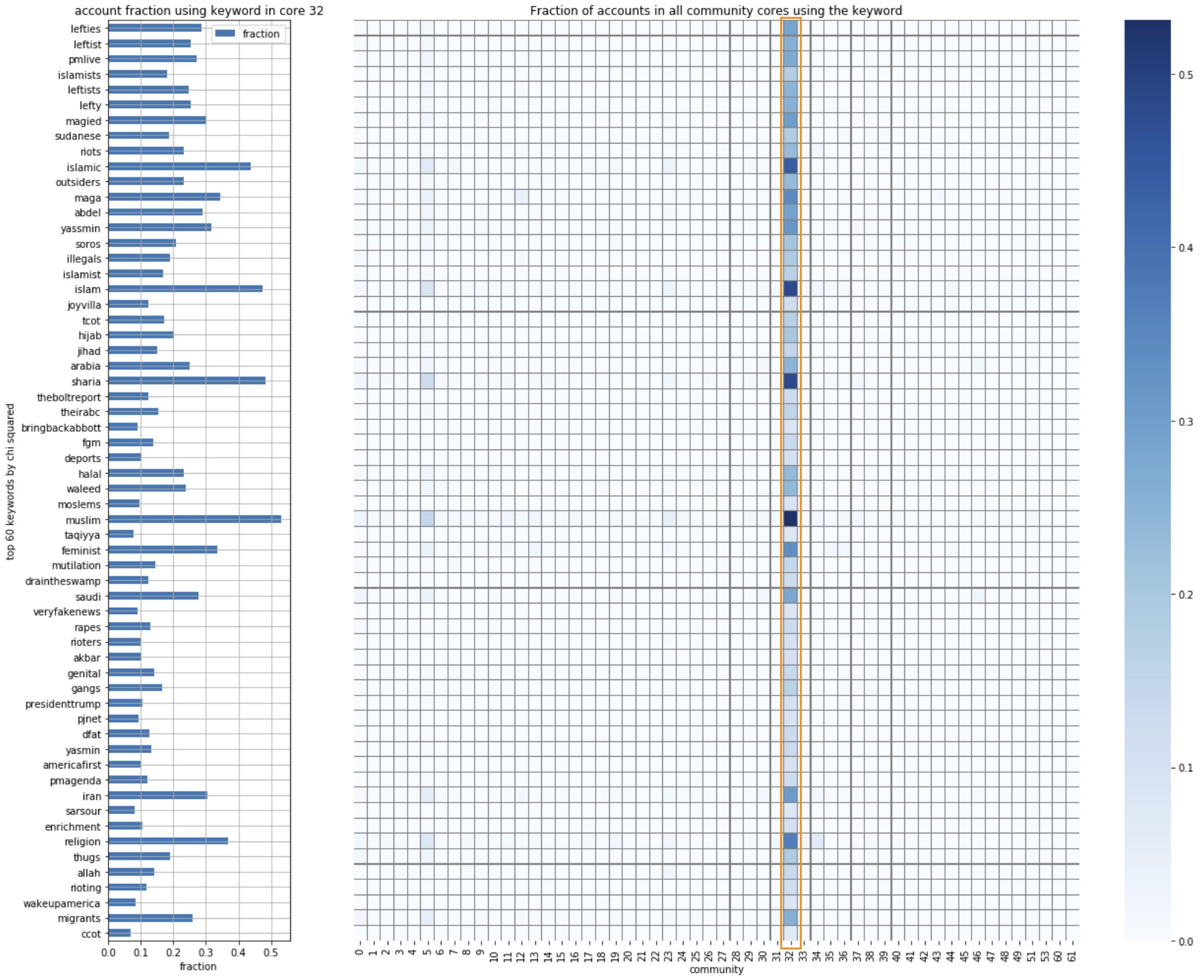
right (Münch, 2019): community graph of communities with degree > 10,000, based on modularity maximisation for the **full Australian follow network** (> 2 million accounts). Labels based on **keyword extraction from tweets by 10% of accounts from each**

community

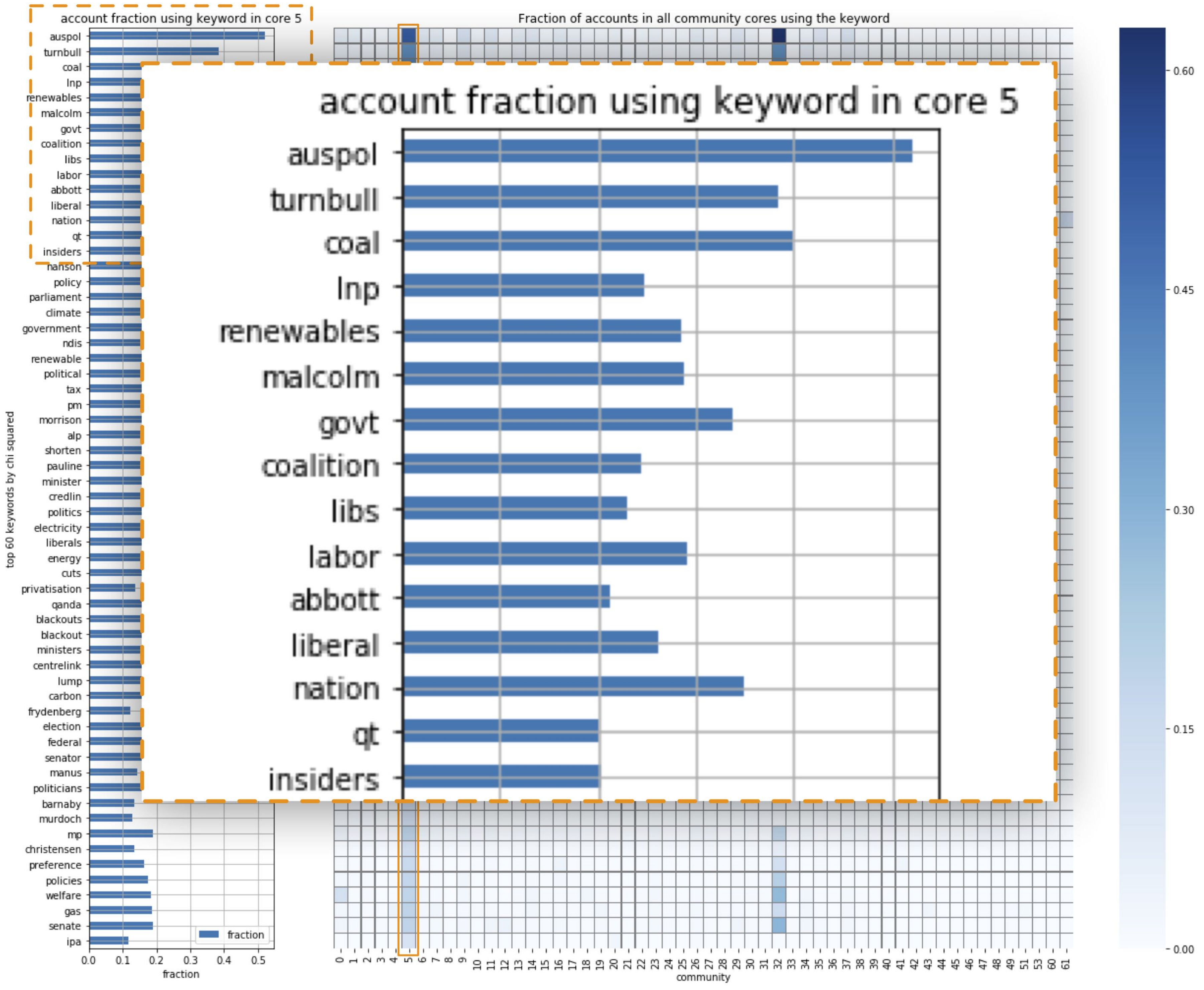
KEYWORD MAPS

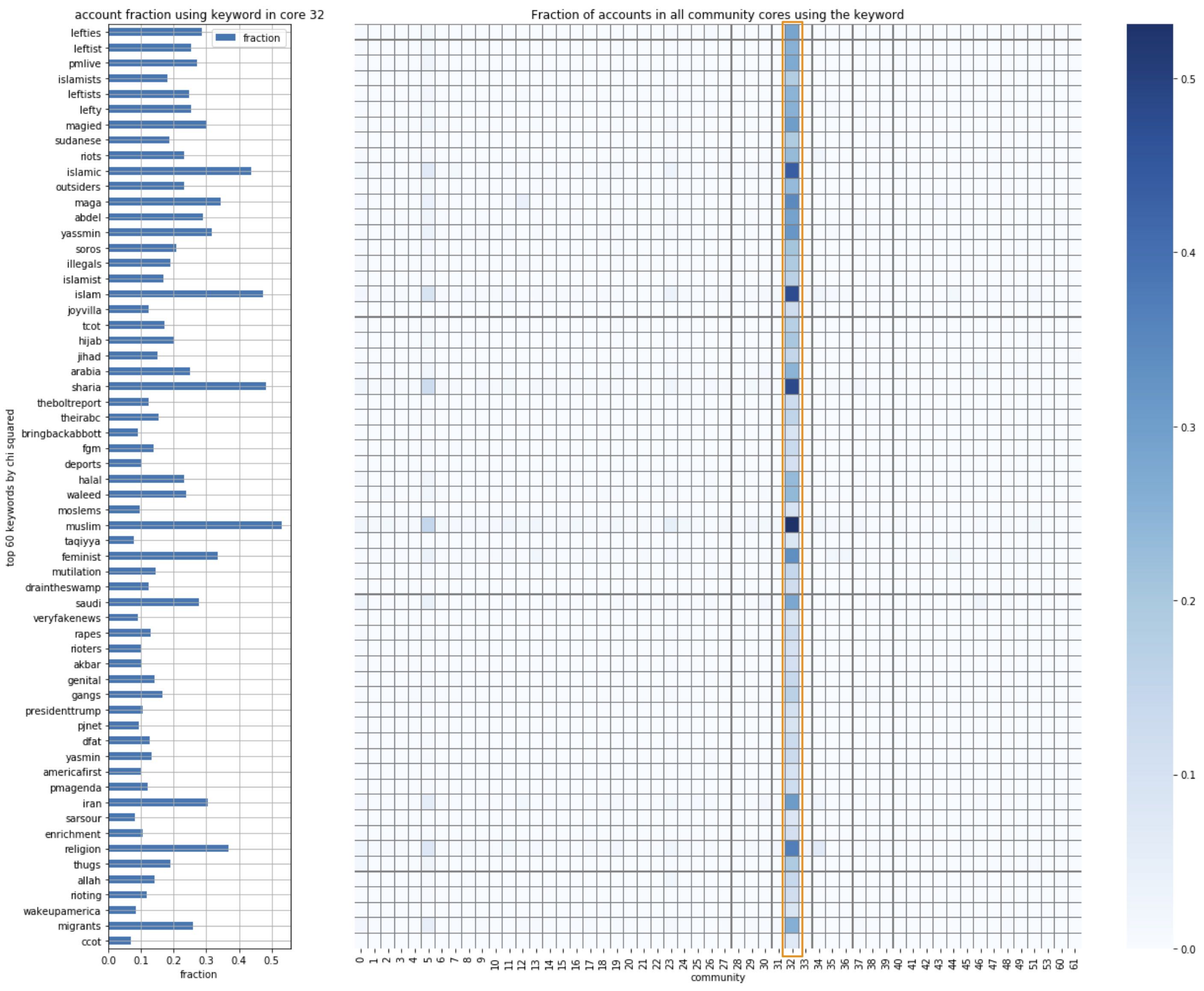
- based on tweets by accounts in community-cores (containing ca. 10% of respective community) from 10-19 February 2017
- done for 53 communities containing over 1000 accounts



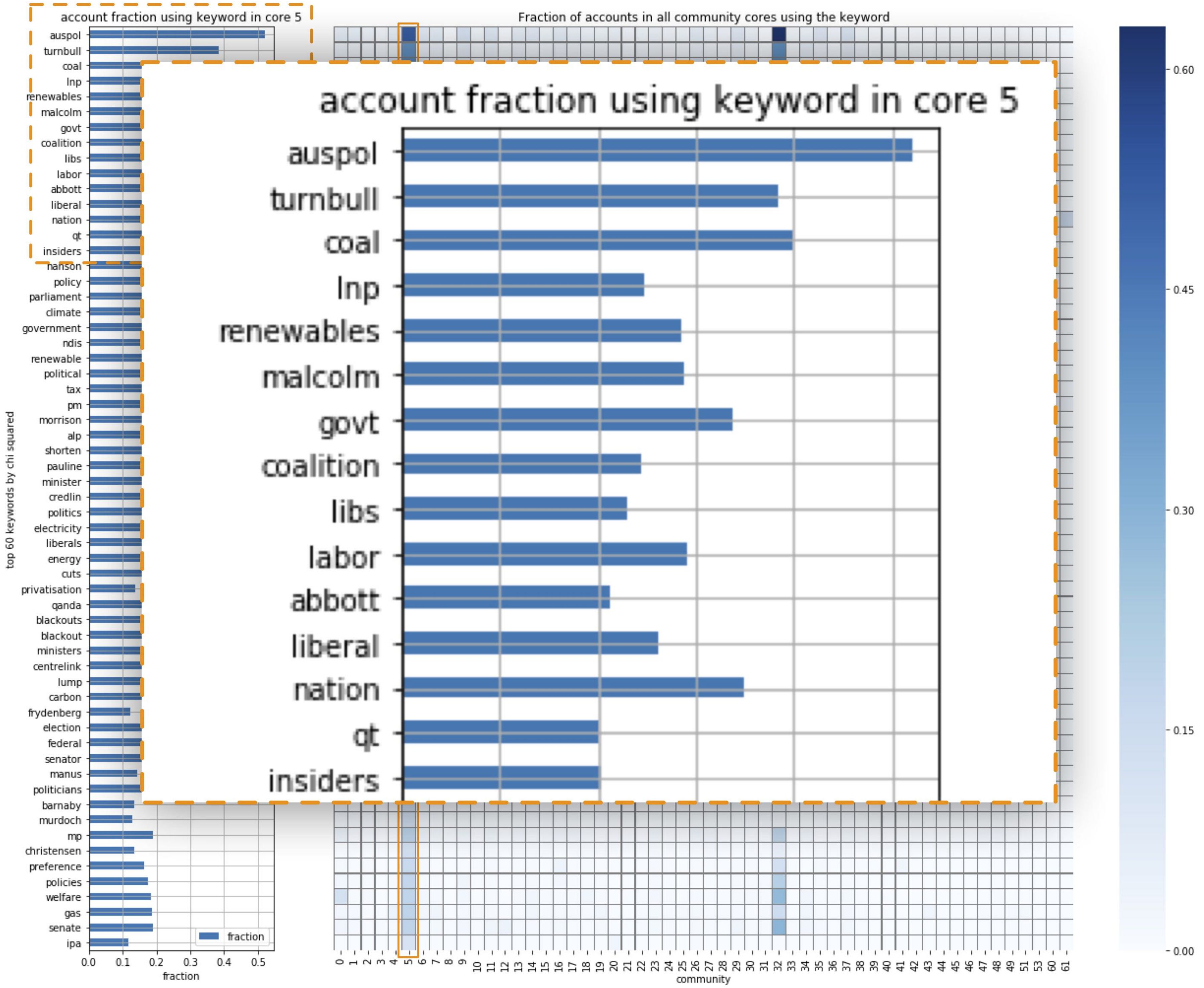


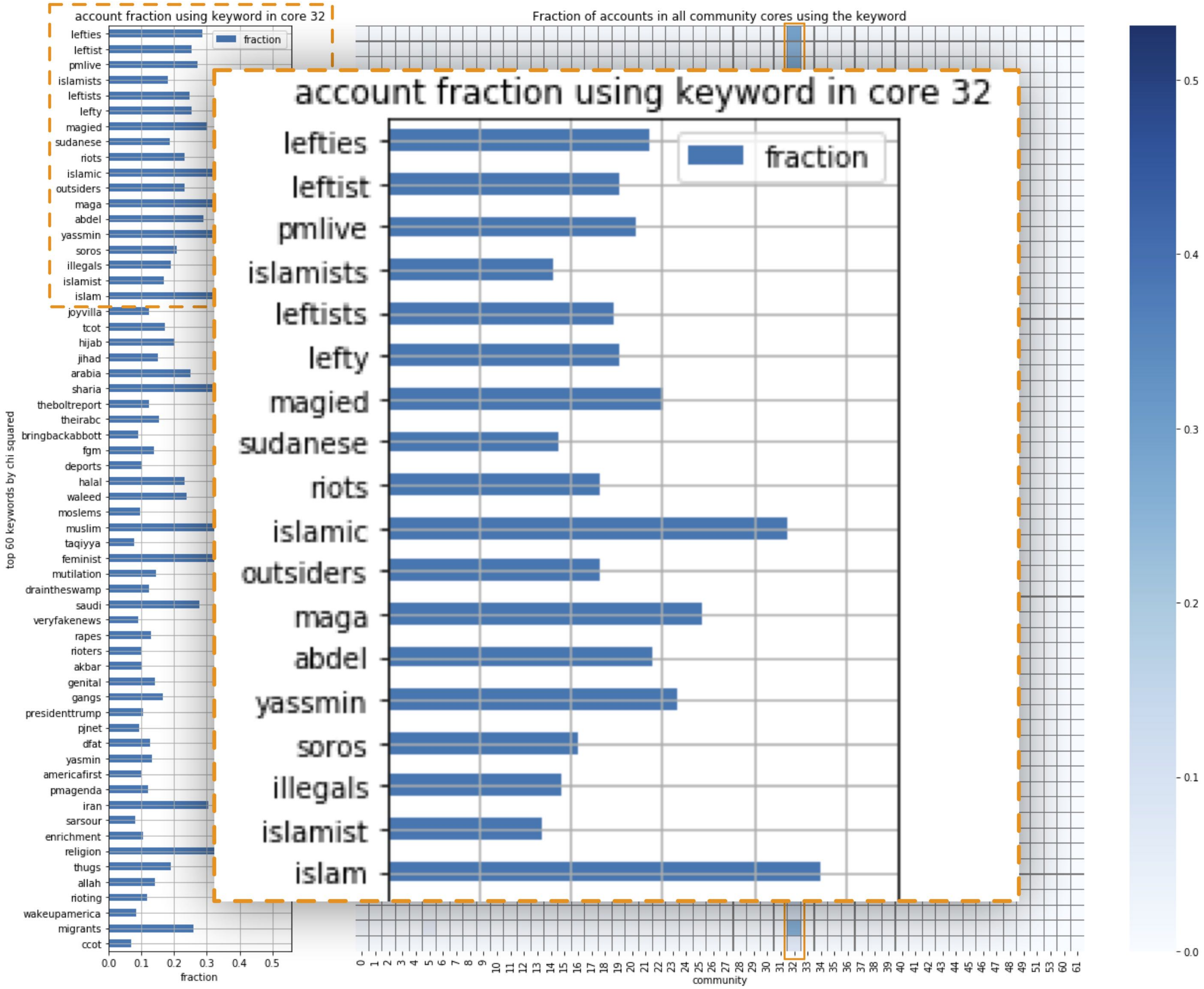
- left side of each figure: most distinct 60 keywords and the fraction of accounts having used them in the community core in focus
- right side of each figure: heatmap showing keyword use in all other communities



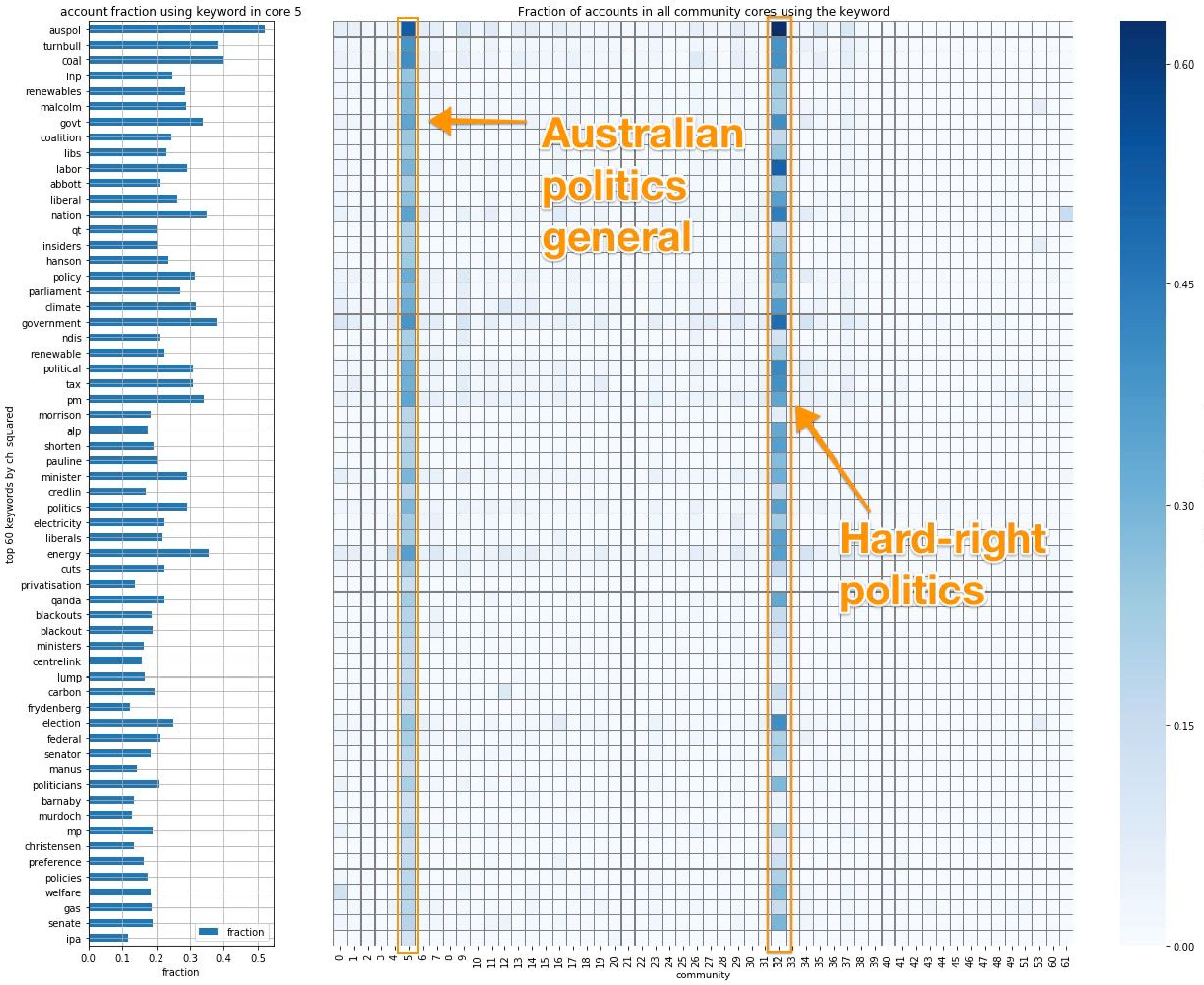


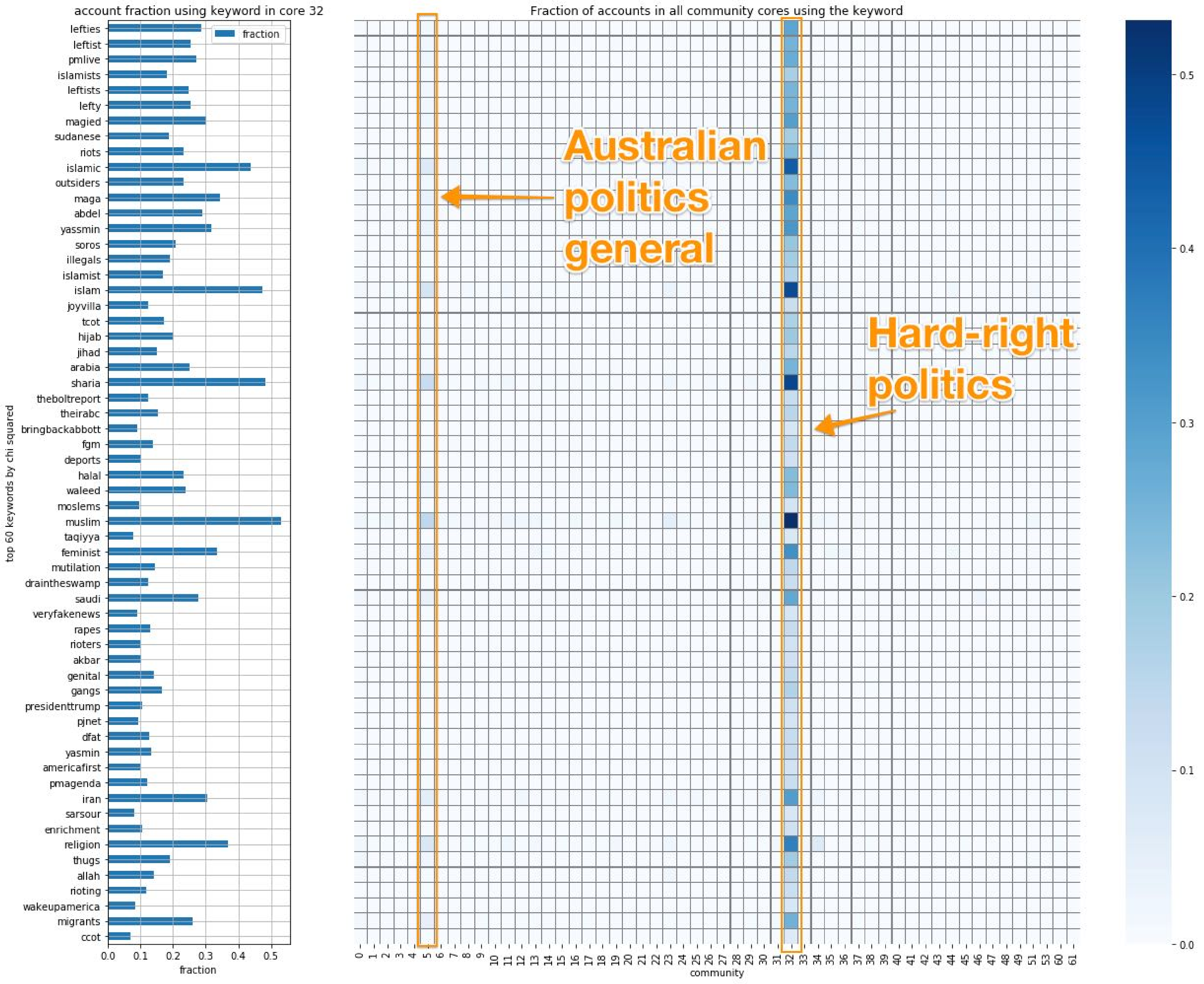
left side: Australian politics, as by the extracted keywords





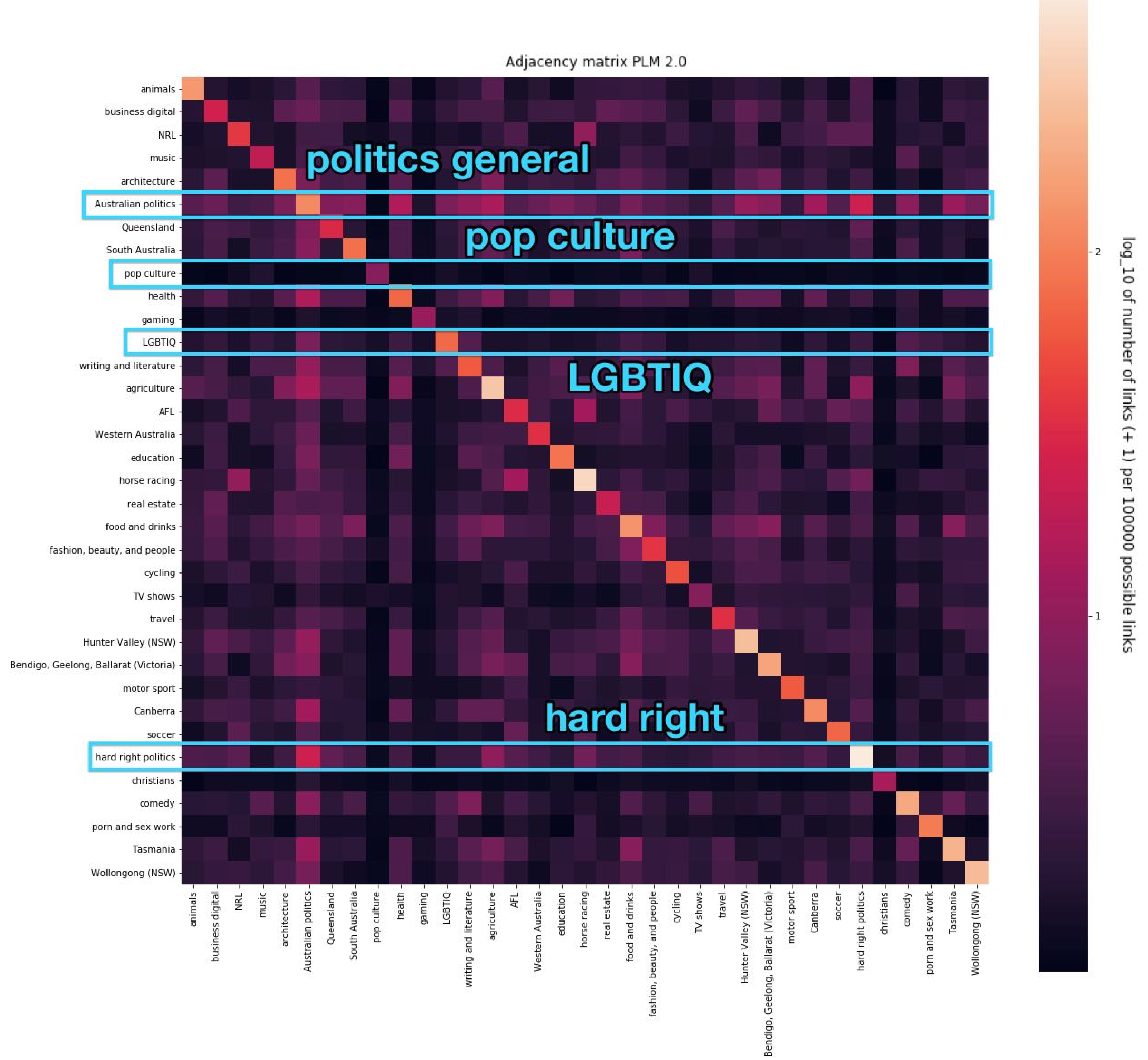
right side: hard-right politics cluster





- visible: one-way relationship regarding topical keywords
- hard-right uses keywords from politics, but not the other way round
- single keywords related to Islam and migrants make it into mainstream discussion
- critique of Islam and immigration as spearhead to enter mainstream discussion?

ADJACENCY MAP



- shows how strongly connected communities are, the lighter the better
- examples: pop culture: with music, and TV shows; politics through whole network; hard right: high inner density, strongly connected with politics

EPISTEMOLOGICAL PROBLEM

MODULARITY MAXIMISATION <-> ECHO CHAMBER

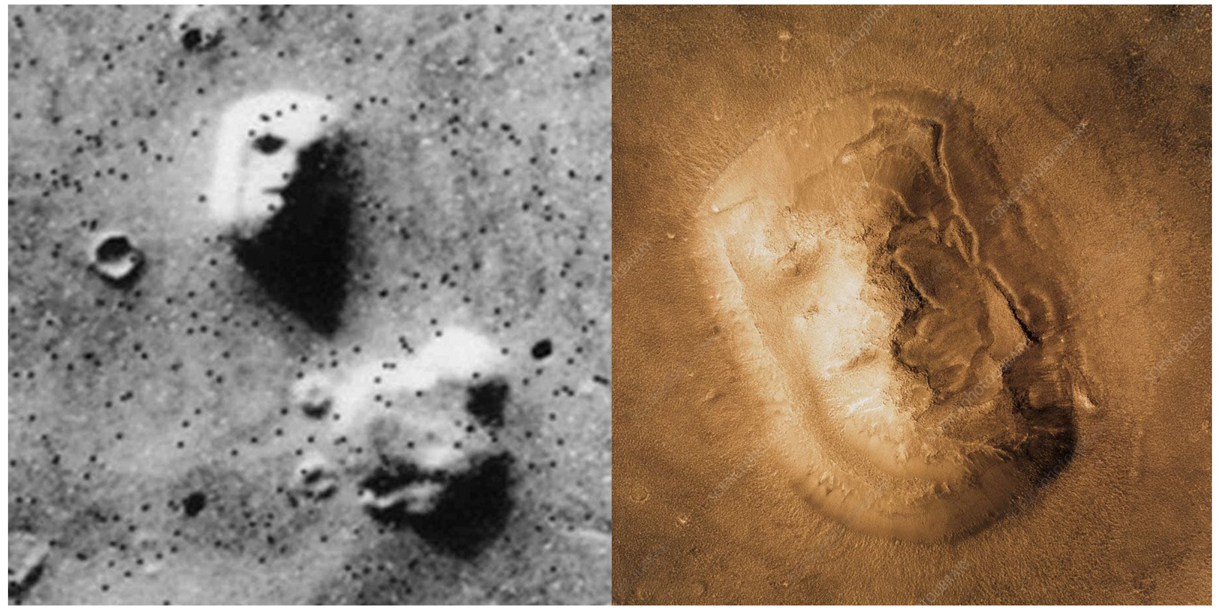
"Networks with high modularity have dense connections between the nodes within modules but sparse connections between nodes in different modules." (Wikipedia, "Modularity (networks)", 11.3.2018)

“An echo chamber comes into being where a group of participants *choose to preferentially connect* with each other, to the exclusion of outsiders.” (Bruns, 2017)

==> significancy problem, as the underlying definition of modularity maximisation is equivalent with echo chamber definition by Bruns

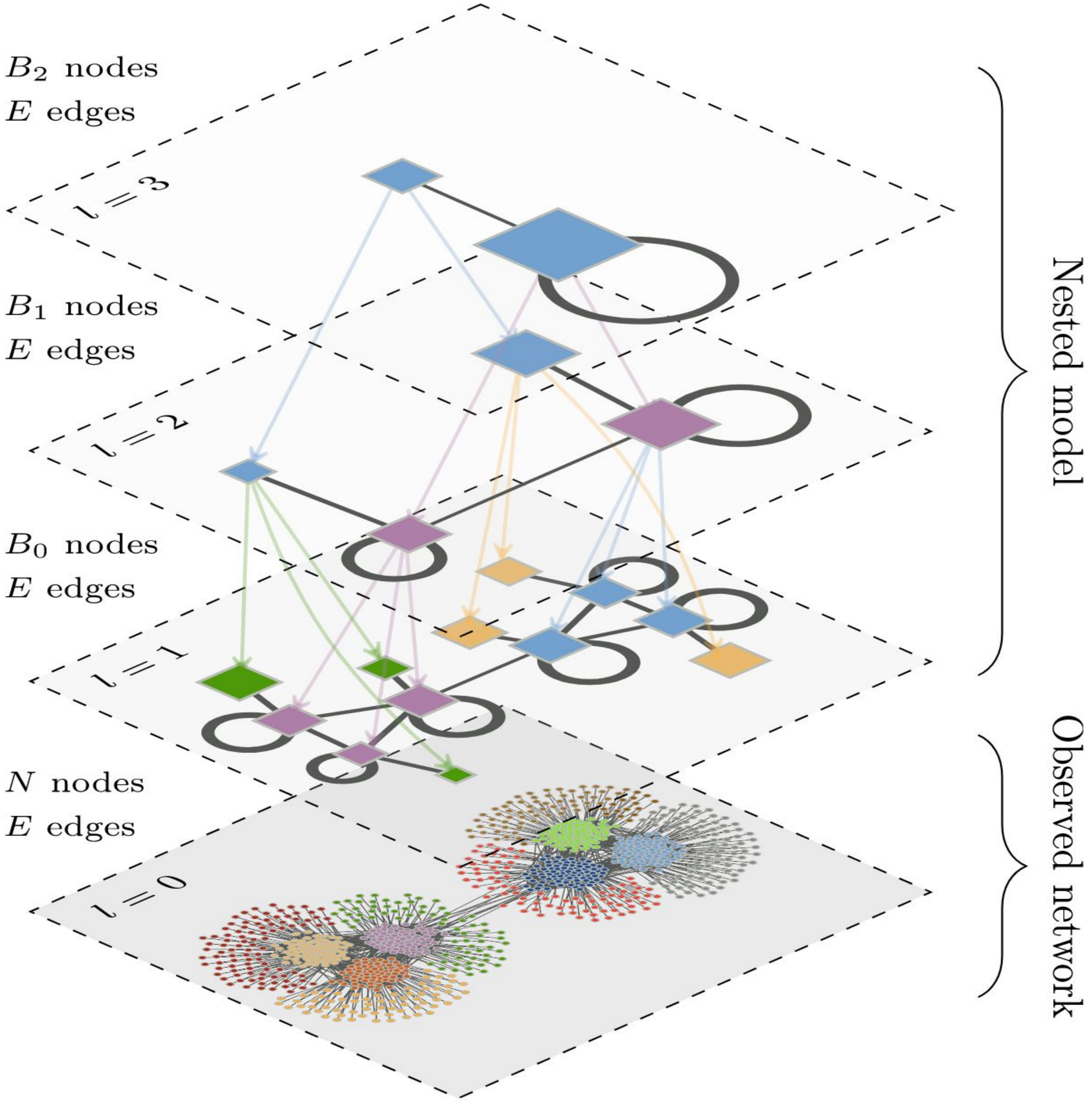
EPISTEMOLOGICAL IMPLICATIONS

- pattern is known, easy to interpret (+)
- efficient (+)
- actually echo-chamber detection method (-)
- resolution limit: either might split up large communities, or ignore small communities (-)
- no statistical tests to distinguish noise from structure, danger of detecting patterns that are not there (apophenia) (-)



PART 2: ALTERNATIVE COMMUNITY DETECTION APPROACH

NESTED STOCHASTIC BLOCK MODEL (SBM) INFERENCE

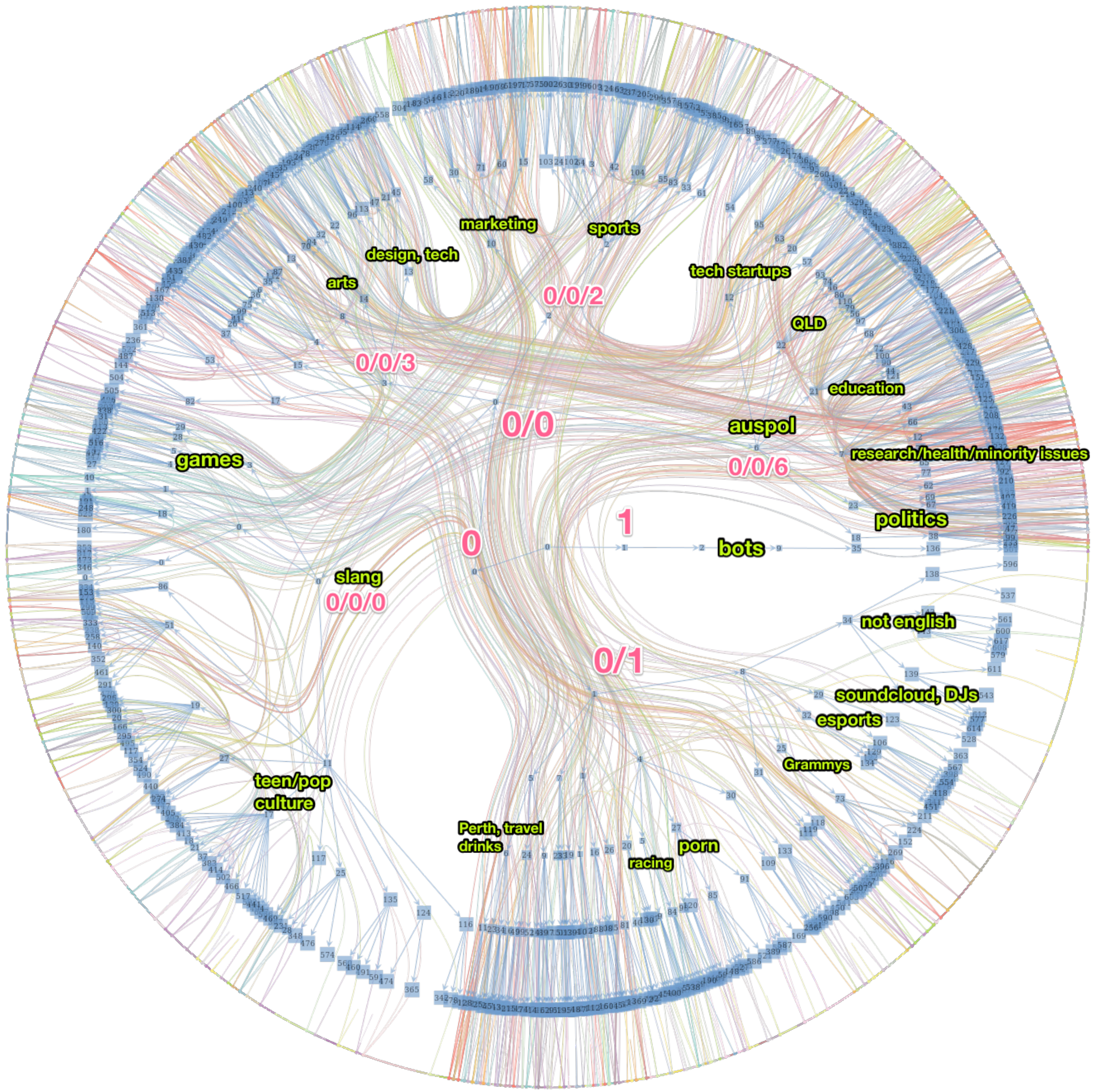


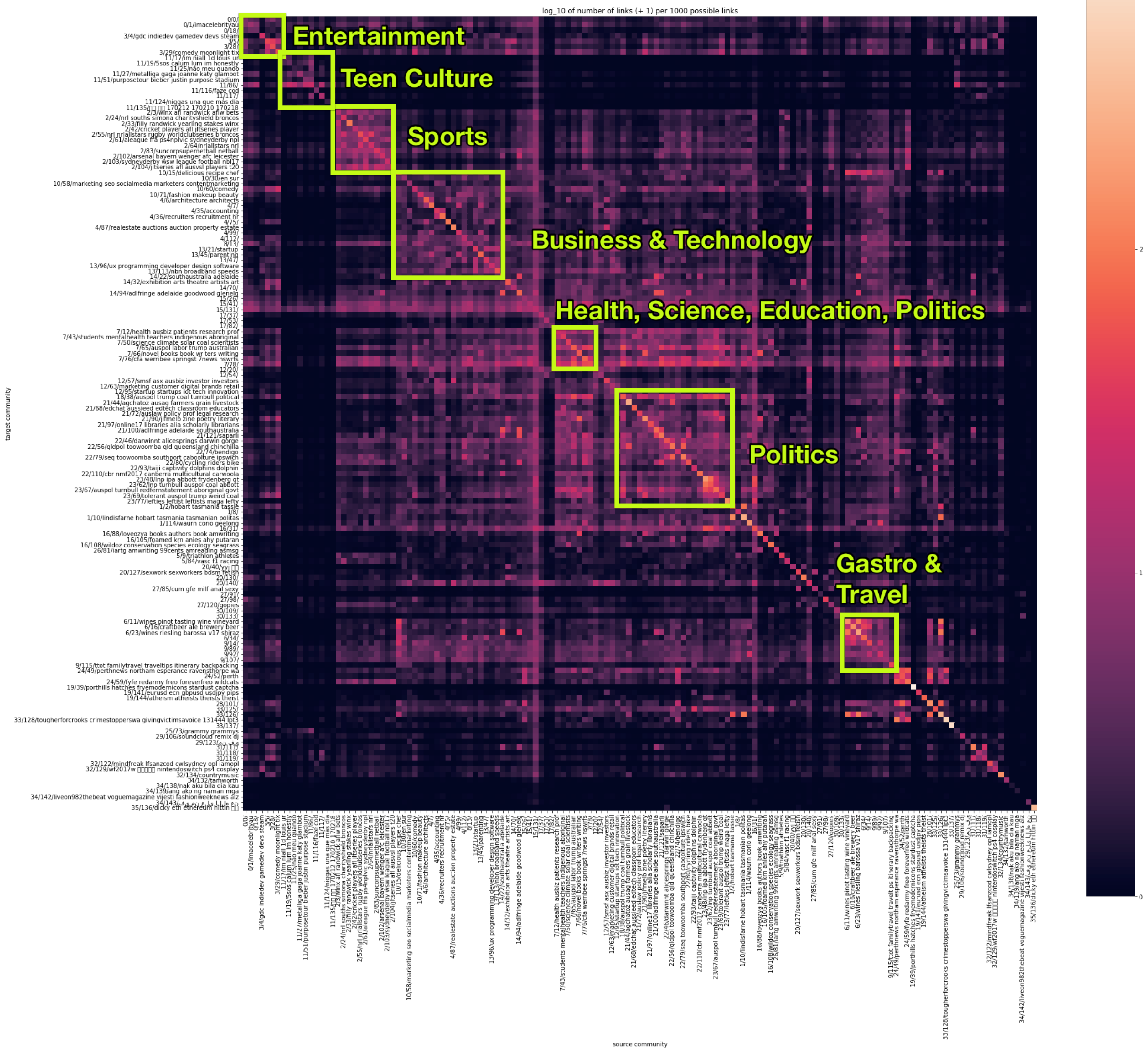
- SBM inference does not look for specific patterns (+-)
- rather grouping nodes with similar 'roles', i.e. connection patterns (think: periphery, centre, bridge, echo chamber) (+)
- hierarchical: circumvents resolution limit (+)
- statistically safeguarded from spotting patterns that are not there (apophenia) (+)
- patterns unknown, analysis after detection, harder to interpret (-)

figure from Peixoto (2017)

RESULTS

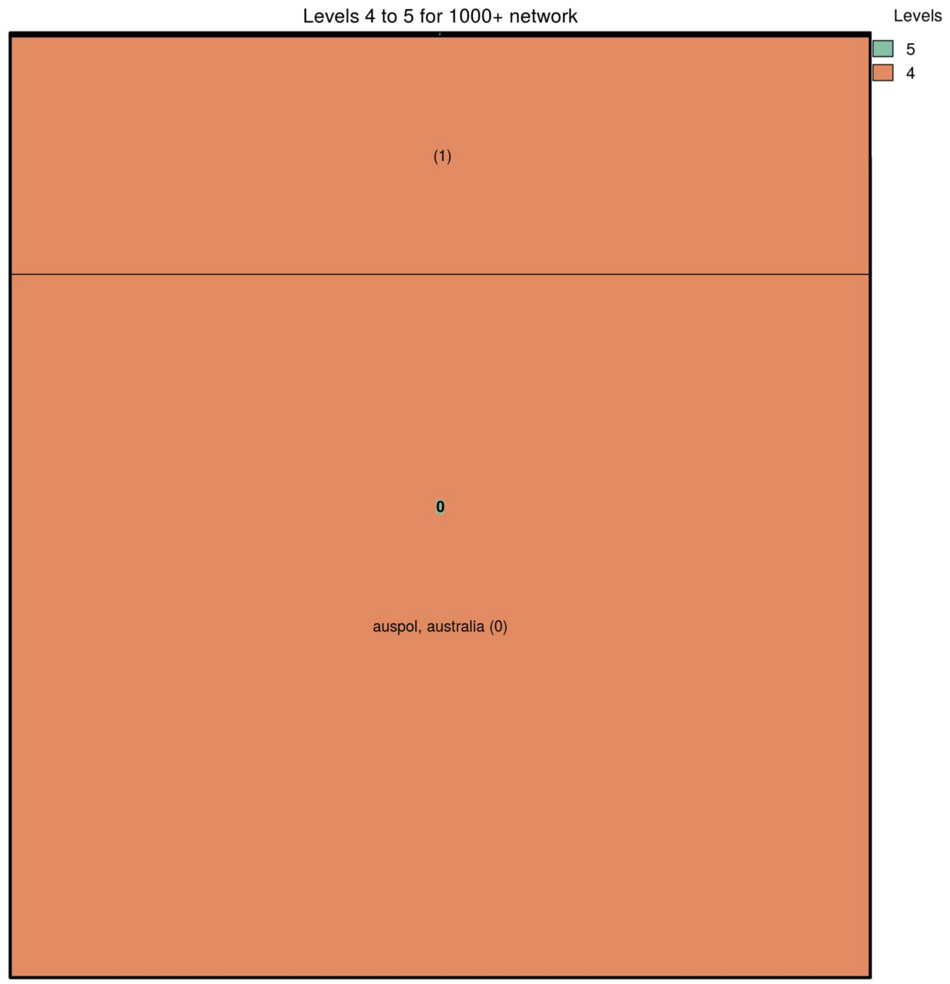
HIERARCHY TREE AND ADJACENCY MAP



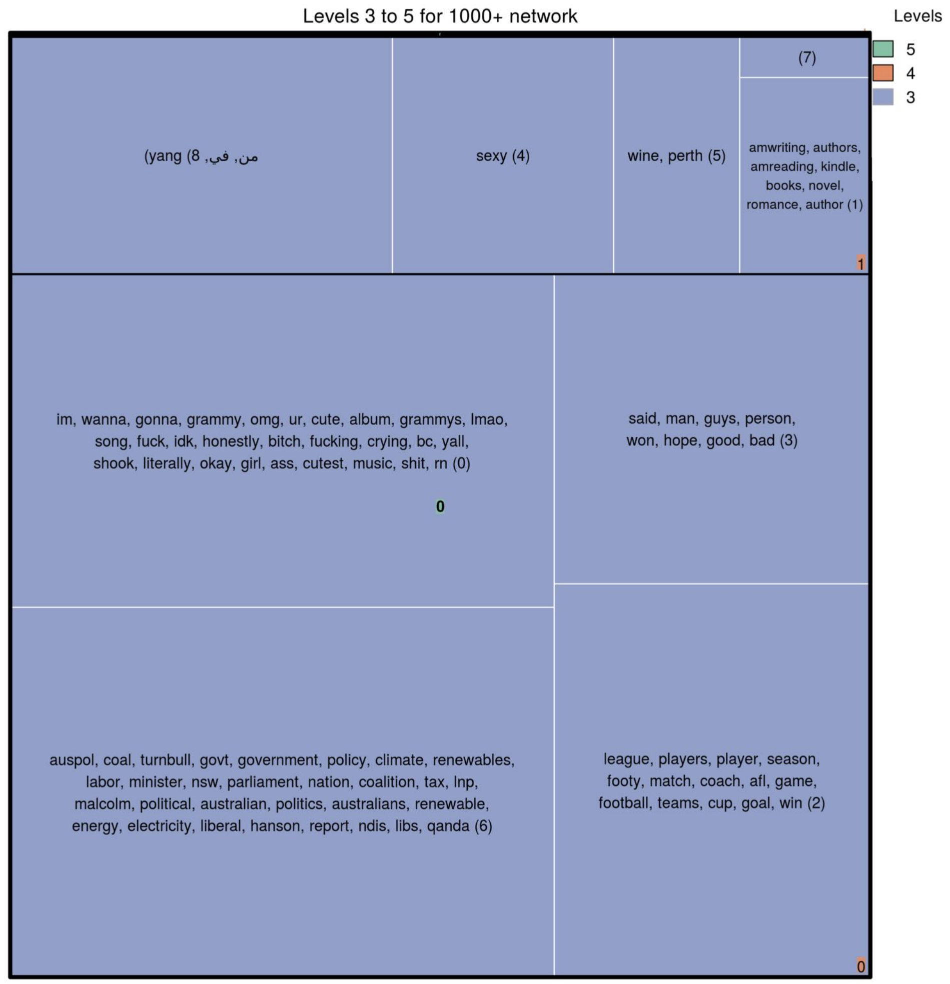


TREEMAPS

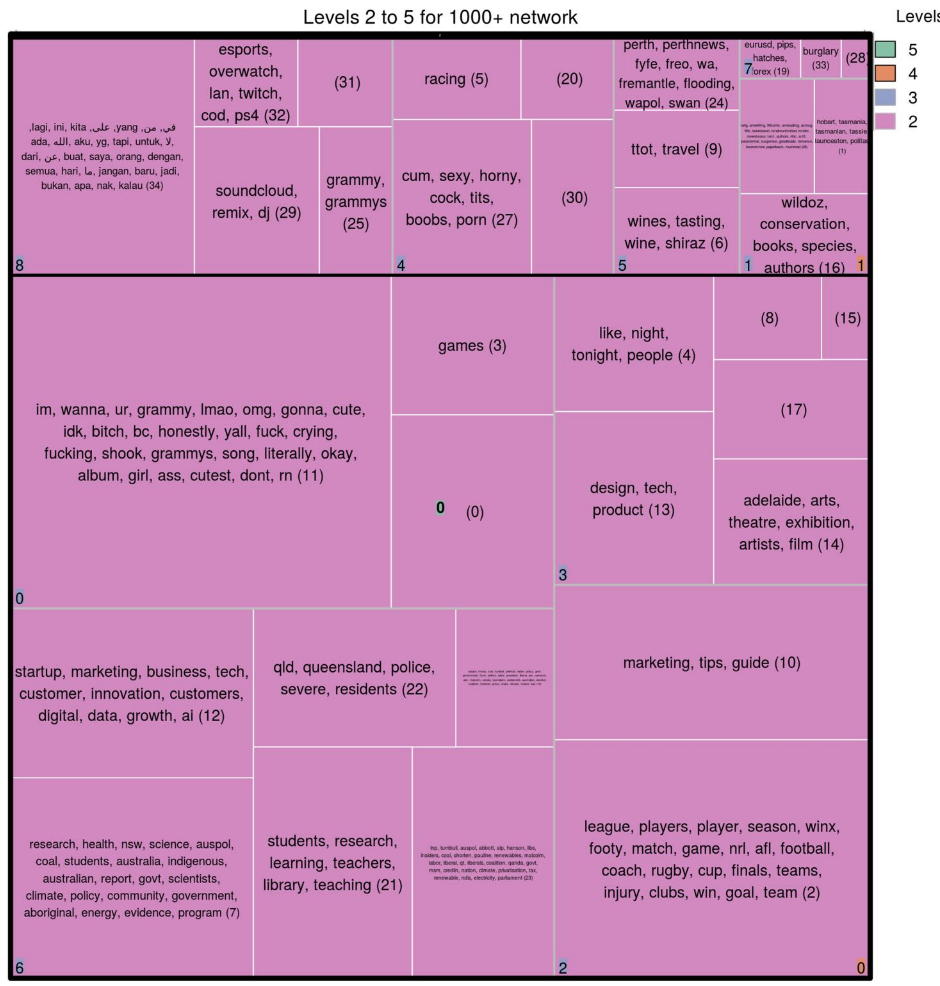
HIERARCHICAL VISUALISATION OF KEYWORDS USED BY ACCOUNTS IN BLOCKS ON EVERY LEVEL



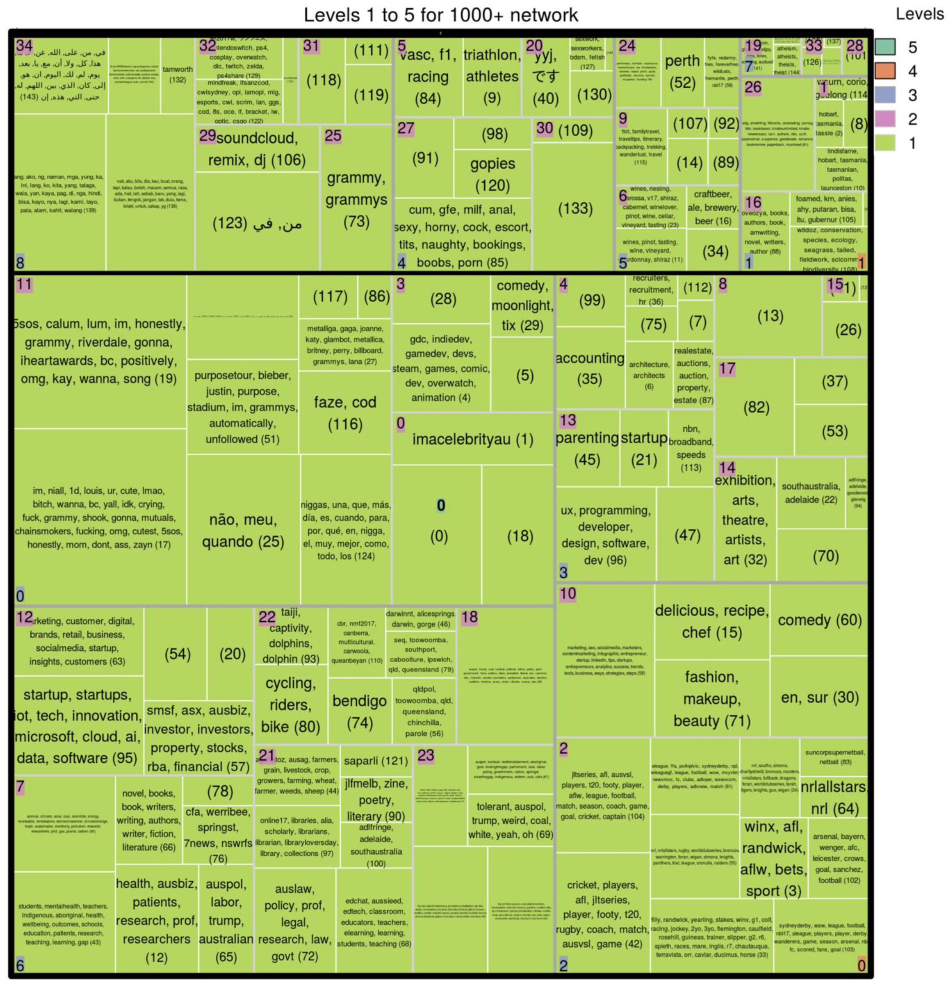
Levels 4 to 5 for 1000+ network
Levels
5
4
(1)
0
auspol, australia (0)



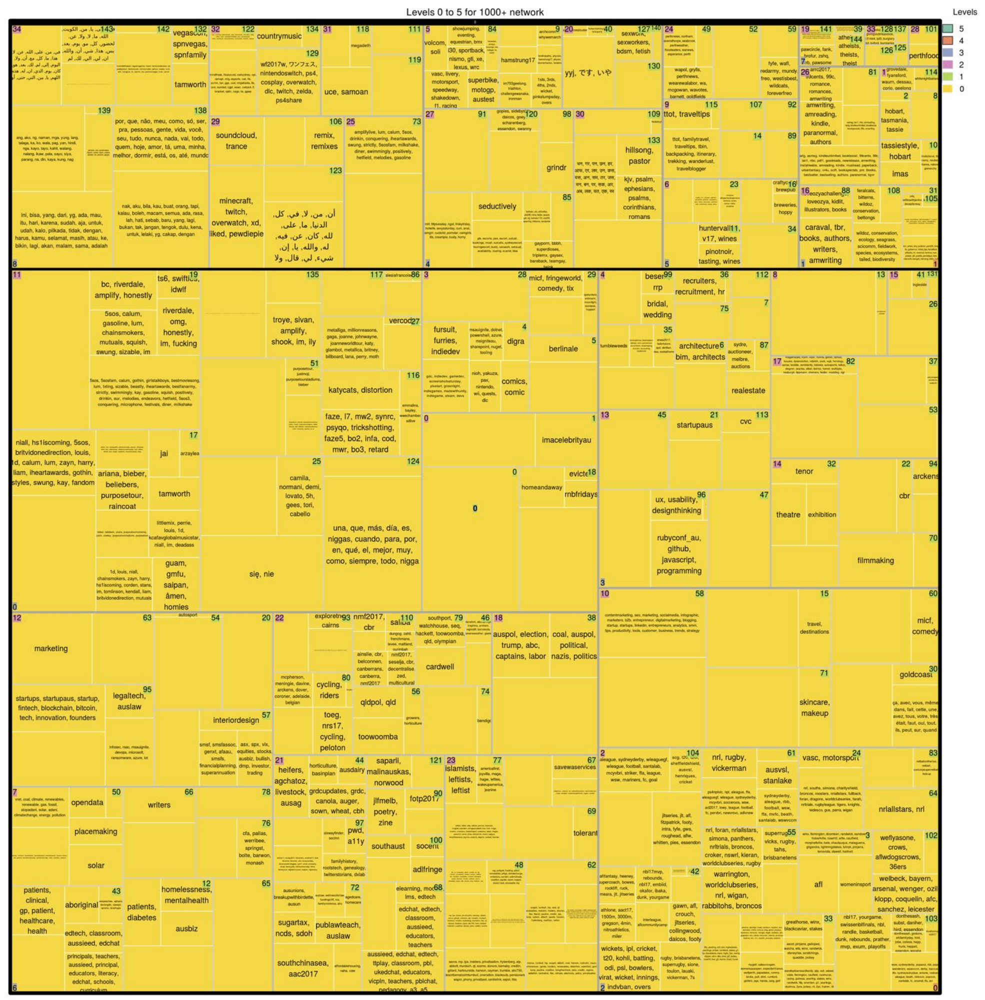
Levels 3 to 5 for 1000+ network
Levels
5
4
3
(yang (8 ,من ,في
sexy (4)
wine, perth (5)
(7)
amwriting, authors, amreading, kindle, books, novel, romance, author (1)
im, wanna, gonna, grammy, omg, ur, cute, album, grammys, lmao, song, fuck, idk, honestly, bitch, fucking, crying, bc, yall, shook, literally, okay, girl, ass, cutest, music, shit, rn (0)
said, man, guys, person, won, hope, good, bad (3)
auspol, coal, turnbull, govt, government, policy, climate, renewables, labor, minister, nsw, parliament, nation, coalition, tax, lnp, malcolm, political, australian, politics, australians, renewable, energy, electricity, liberal, hanson, report, ndis, libs, qanda (6)
league, players, player, season, footy, match, coach, afl, game, football, teams, cup, goal, win (2)



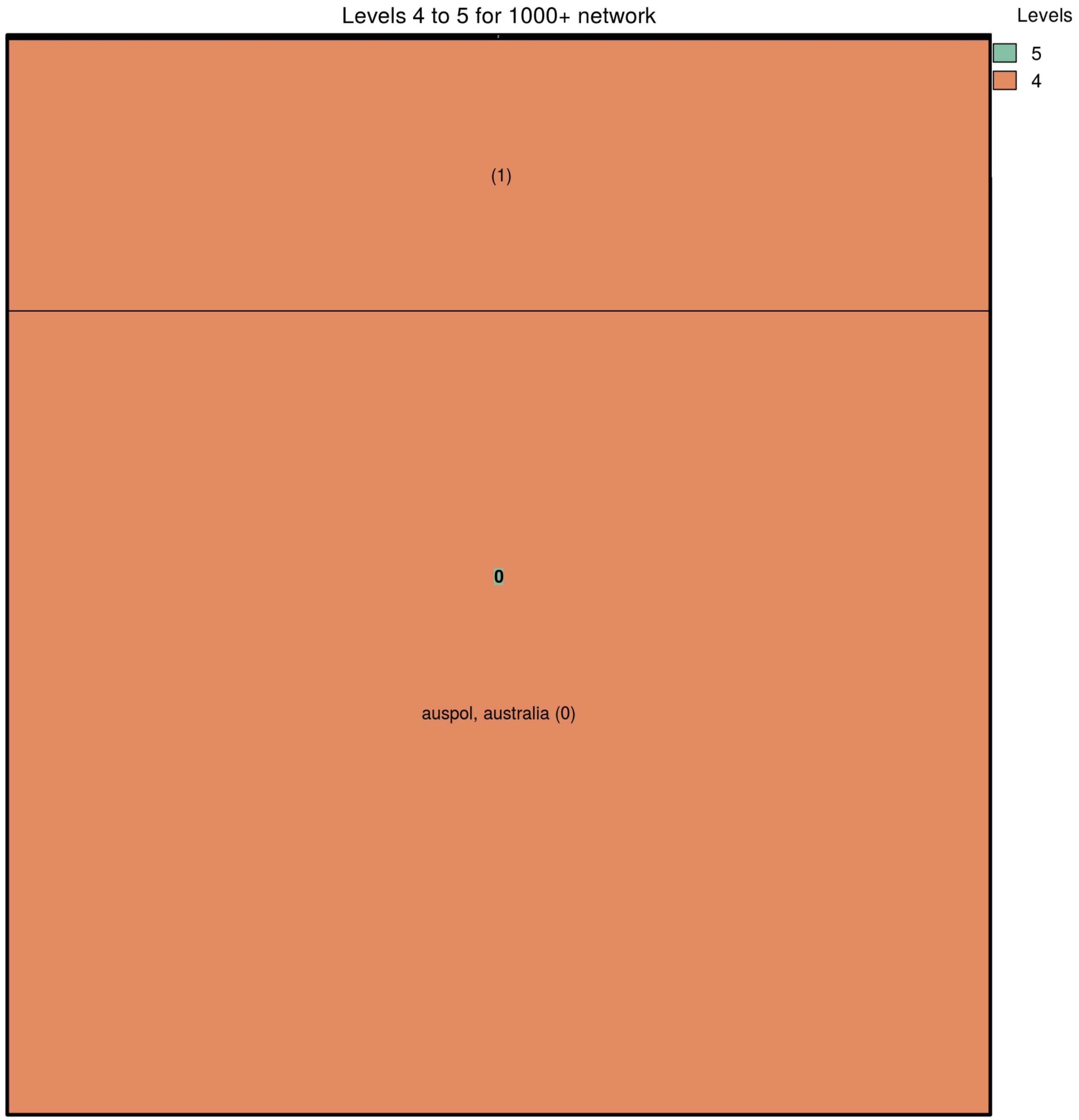
Levels 2 to 5 for 1000+ network
Levels
5
4
3
2
esports, overwatch, lan, twitch, cod, ps4 (32)
(31)
racing (5)
(20)
perth, perthnews, fyfe, freo, wa, fremantle, flooding, wapol, swan (24)
soundcloud, remix, dj (29)
grammy, grammys (25)
cum, sexy, horny, cock, tits, boobs, porn (27)
(30)
ttot, travel (9)
wines, tasting, wine, shiraz (6)
wildoz, conservation, books, species, authors (16)
games (3)
like, night, tonight, people (4)
(8)
(15)
(17)
im, wanna, ur, grammy, lmao, omg, gonna, cute, idk, bitch, bc, honestly, yall, fuck, crying, fucking, shook, grammys, song, literally, okay, album, girl, ass, cutest, dont, rn (11)
(0)
design, tech, product (13)
adelaide, arts, theatre, exhibition, artists, film (14)
startup, marketing, business, tech, customer, innovation, customers, digital, data, growth, ai (12)
qld, queensland, police, severe, residents (22)
marketing, tips, guide (10)
research, health, nsw, science, auspol, coal, students, australia, indigenous, australian, report, govt, scientists, climate, policy, community, government, aboriginal, energy, evidence, program (7)
students, research, learning, teachers, library, teaching (21)
league, players, player, season, winx, footy, match, game, nrl, afl, football, coach, rugby, cup, finals, teams, injury, clubs, win, goal, team (2)



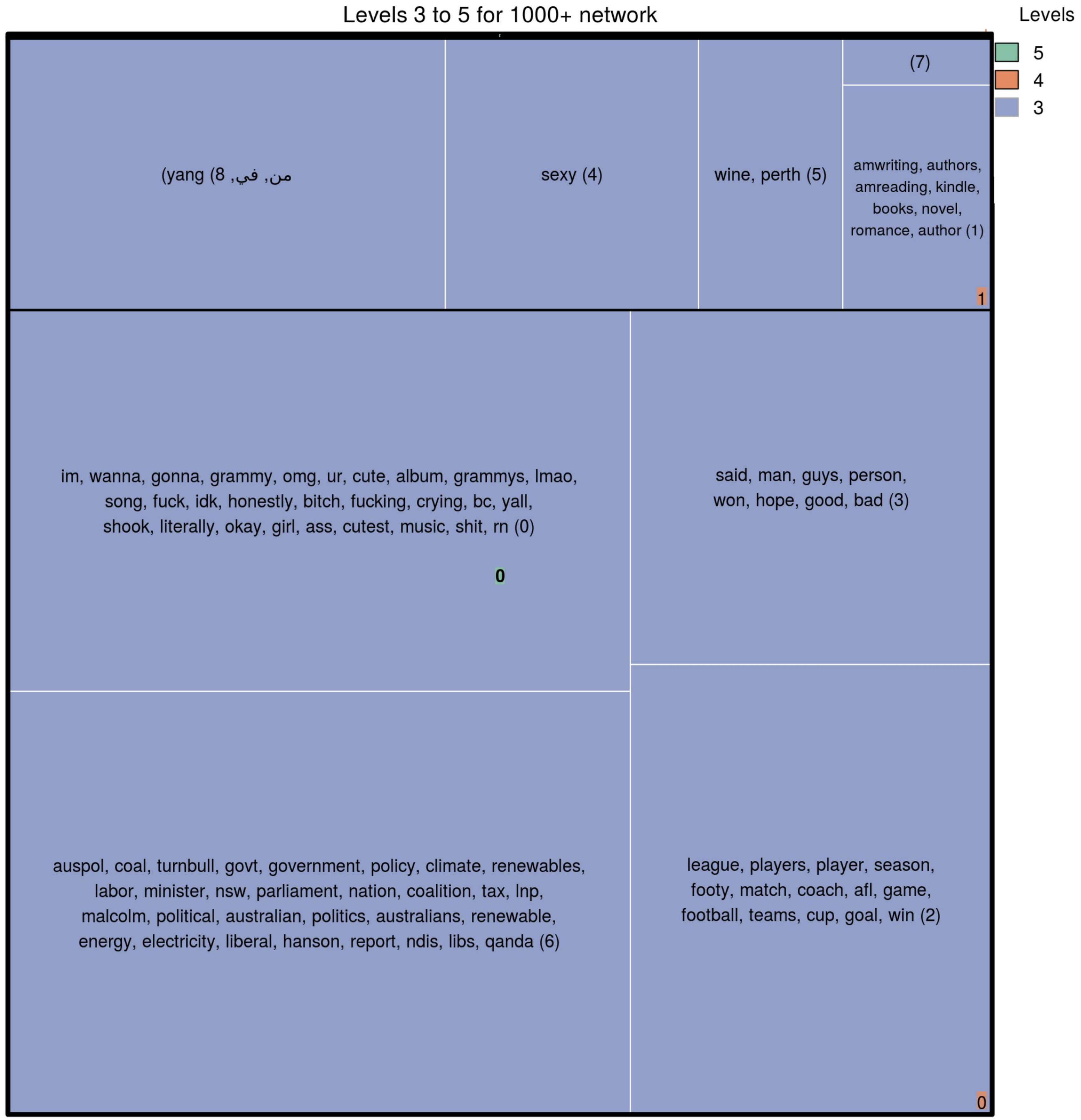
Levels 1 to 5 for 1000+ network
Levels
5
4
3
2
1
soundcloud, remix, dj (106)
grammy, grammys (73)
vasc, f1, racing (84)
triathlon, athletes (9)
gopies (120)
cum, gfe, milf, anal, sexy, horny, cock, escort, tits, naughty, bookings, boobs, porn (85)
5sos, calum, lum, im, honestly, grammy, riverdale, gonna, iheartawards, bc, positively, omg, kay, wanna, song (19)
faze, cod (116)
imacelebrityau (1)
accounting (35)
parenting (45)
startup (21)
não, meu, quando (25)
ux, programming, developer, design, software, dev (96)
exhibition, arts, theatre, artists, art (32)
delicious, recipe, chef (15)
comedy (60)
fashion, makeup, beauty (71)
en, sur (30)
startup, startups, iot, tech, innovation, microsoft, cloud, ai, data, software (95)
cycling, riders, bike (80)
bendigo (74)
nrlallstars nrl (64)
winx, afl, randwick, aflw, bets, sport (3)
health, ausbiz, patients, research, prof, researchers (12)
auspol, labor, trump, australian (65)
auslaw, policy, prof, legal, research, law, govt (72)
cricket, players, afl, jltseries, rugby, coach, match, ausvel, game (42)



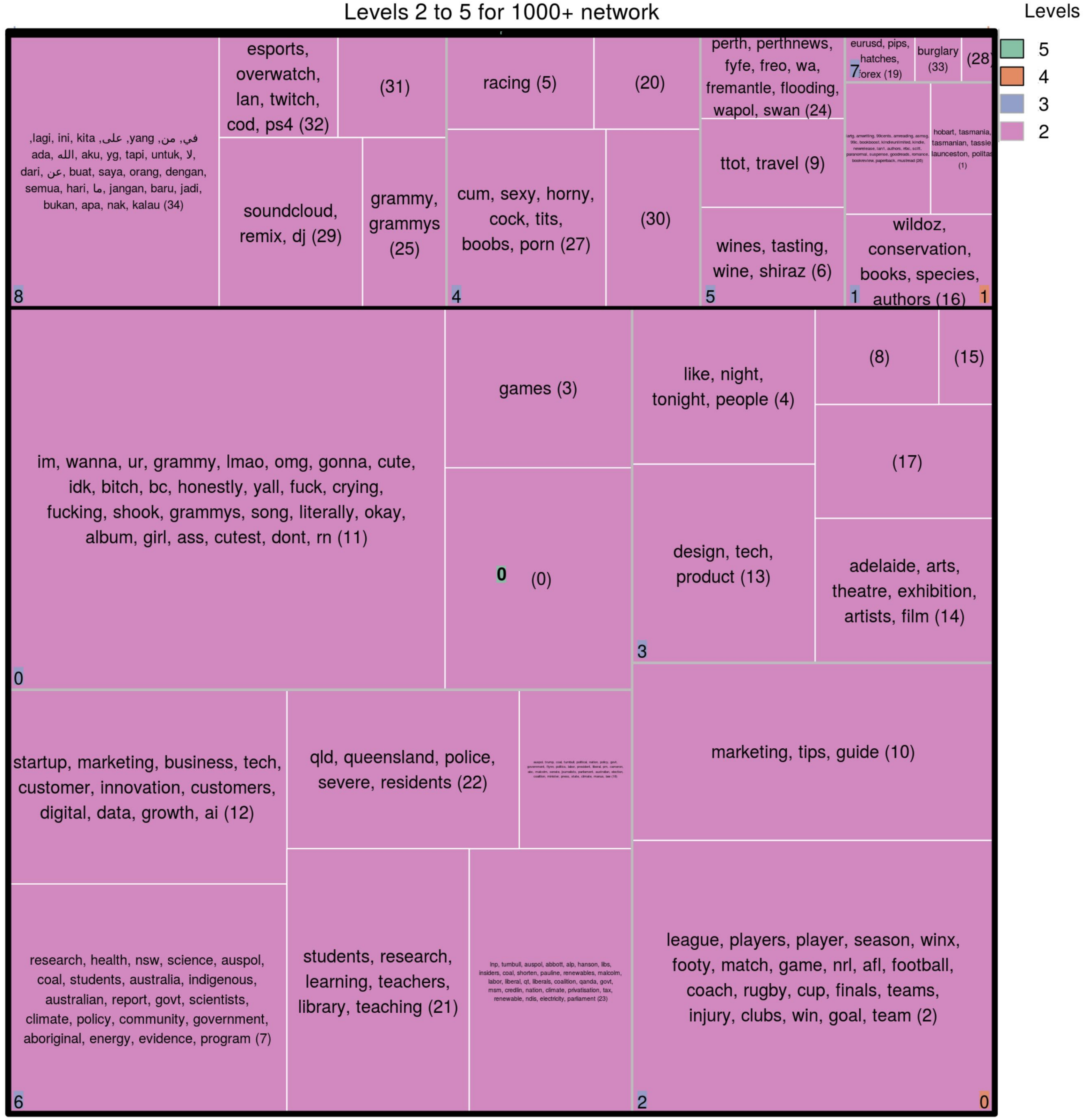
Levels 0 to 5 for 1000+ network

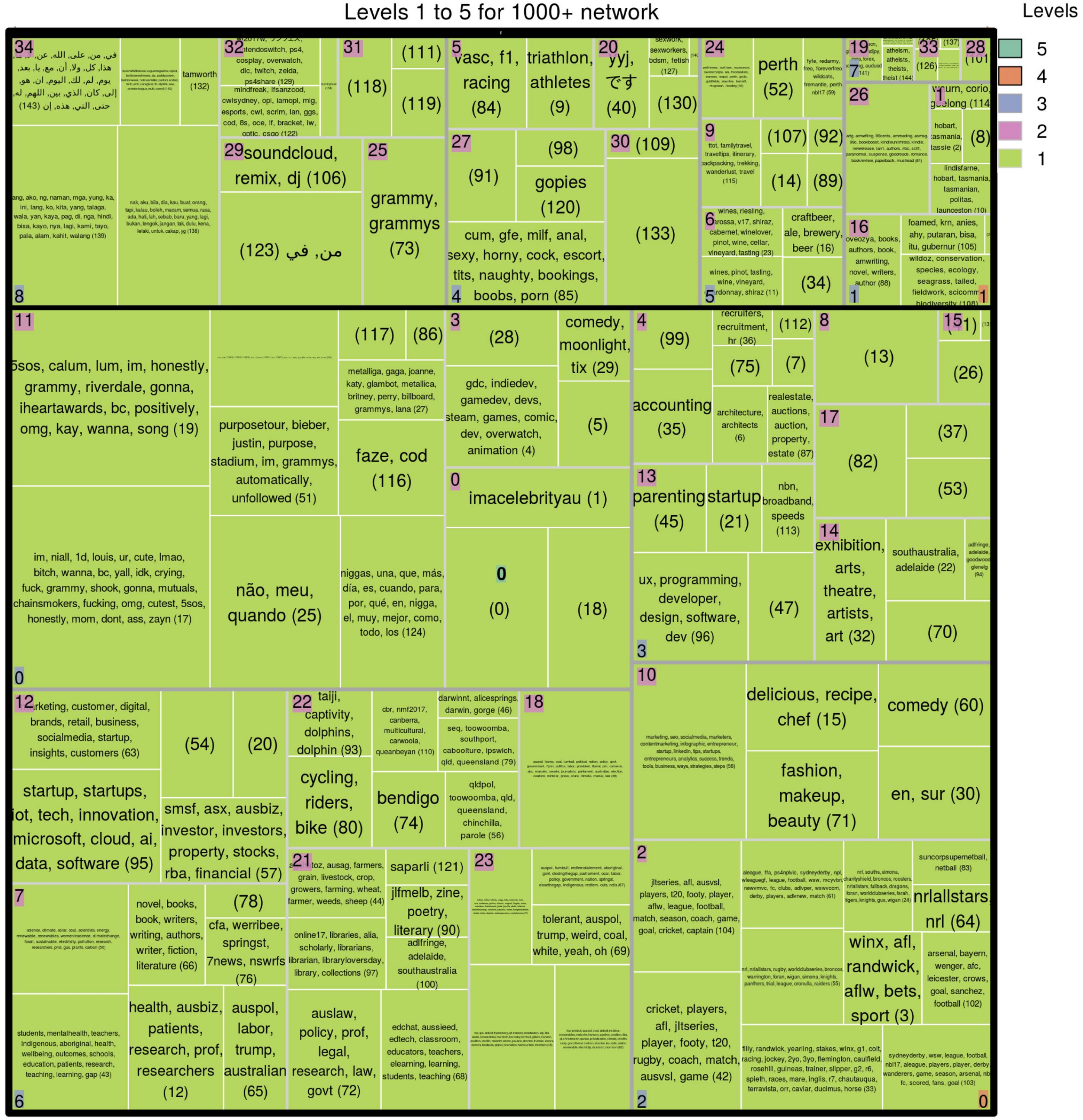


- size of blocks represents active accounts, keywords used by at least 5% of active accounts in block
- visible: divide in periphery & centre; dominance of auspol



politics and sports become clearly visible

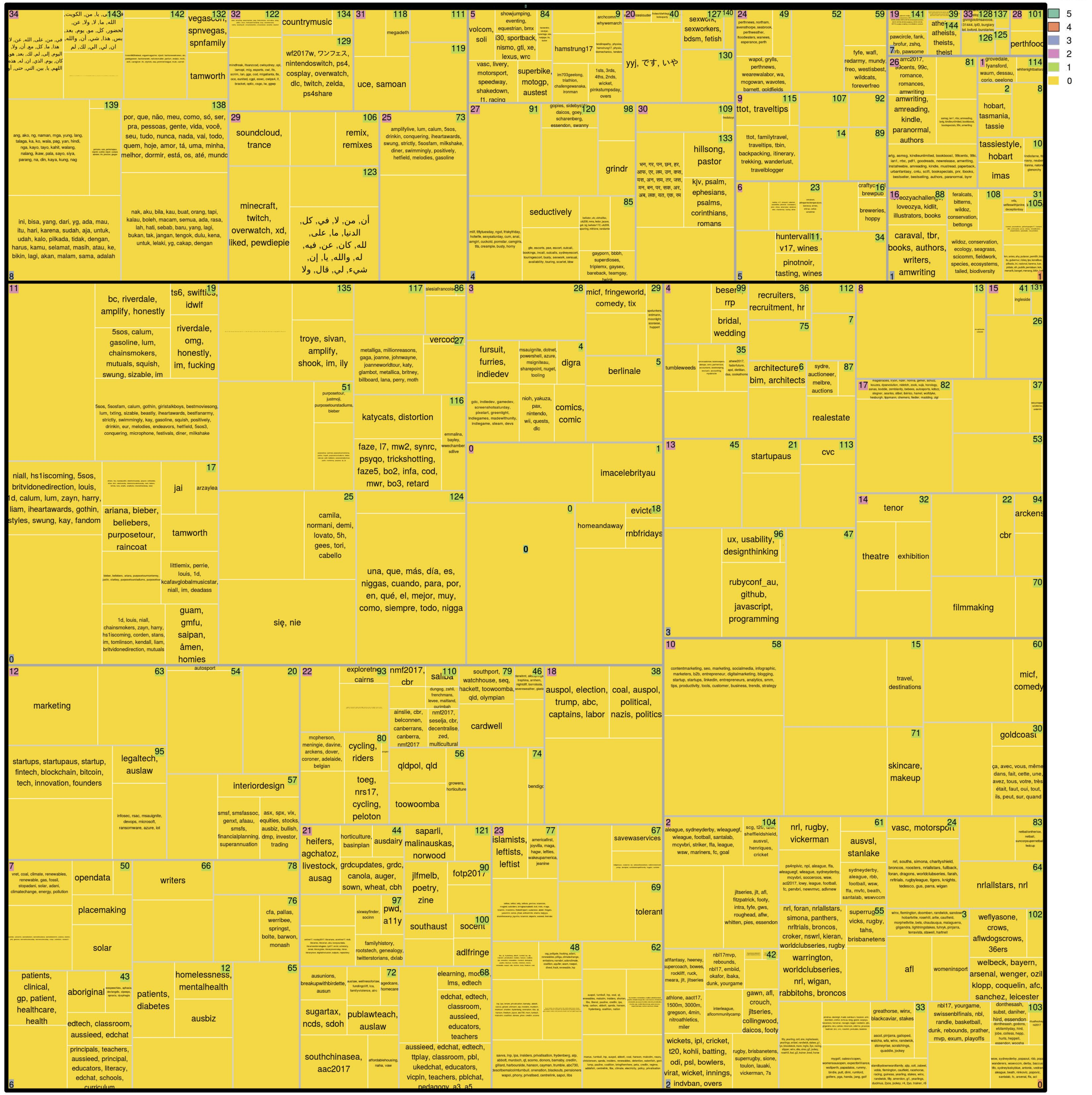




level 2: porn, teens, startups, politics

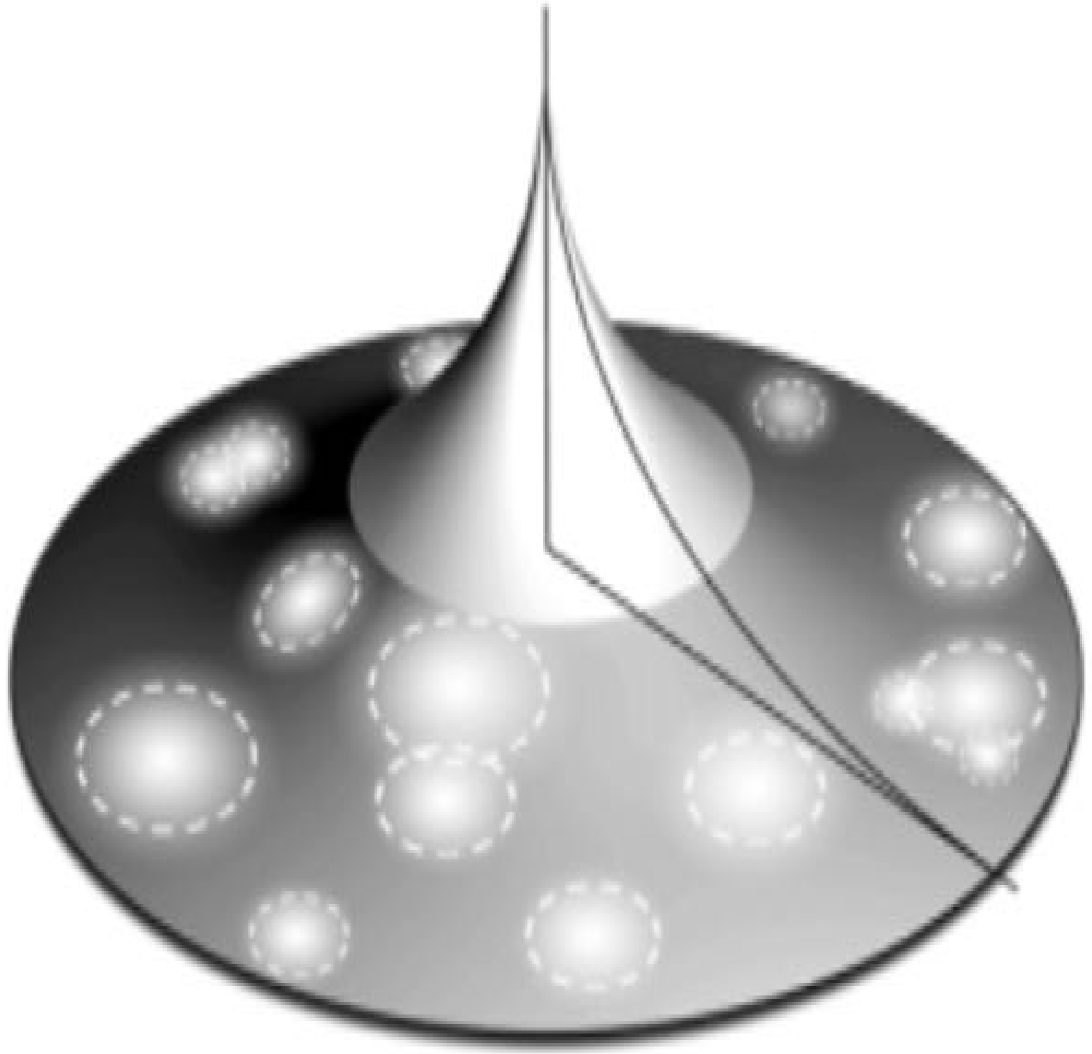
level 1: sports now divided into different forms of sport (e.g. European Soccer vs AFL)

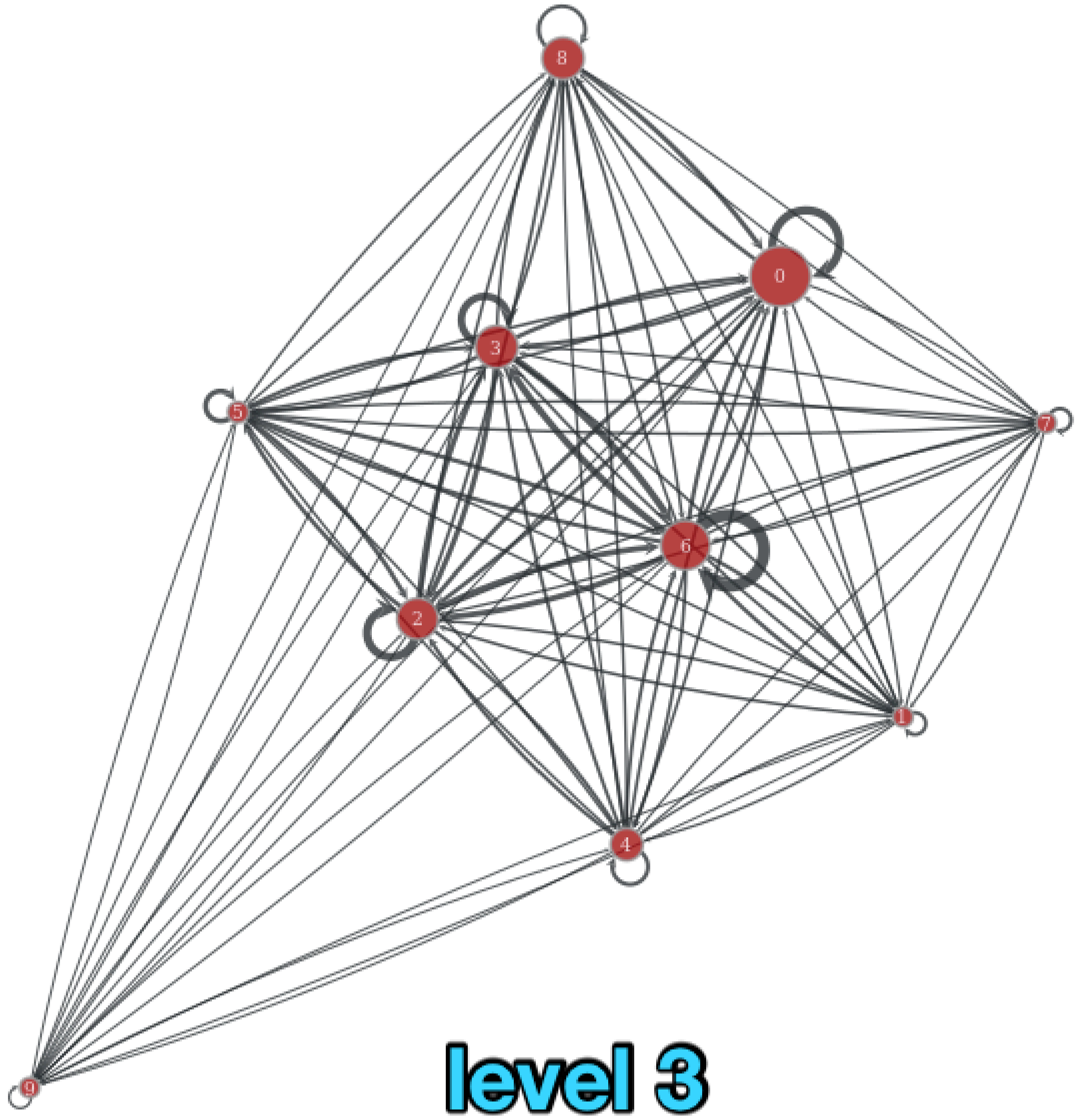
Levels 0 to 5 for 1000+ network

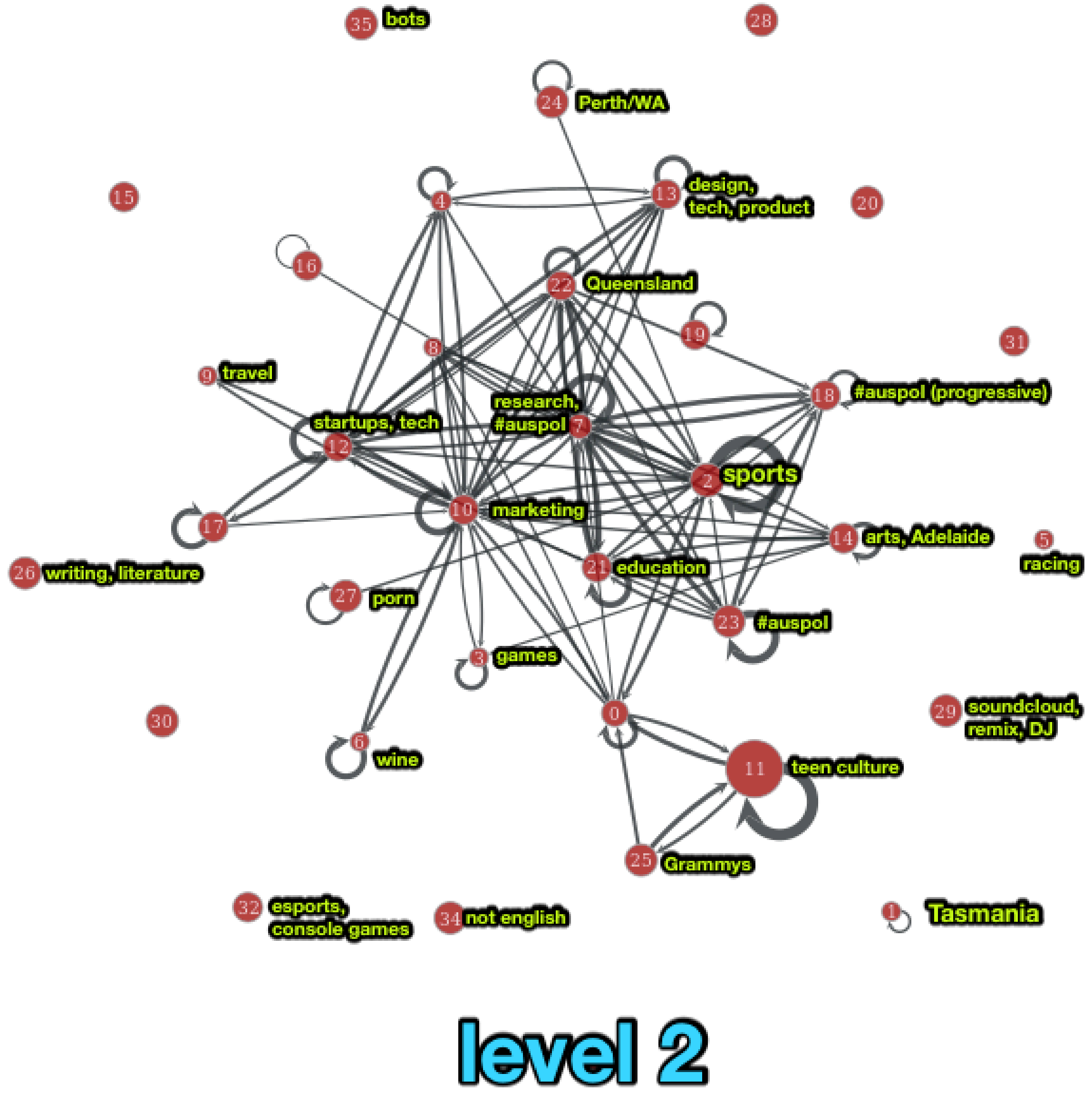


- most detailed level, zoomed in on sports
- now visible special interests such as golf, track athletics, or european soccer

APPLICATION TO THEORY (EXAMPLE): NETWORKED PUBLIC SPHERE

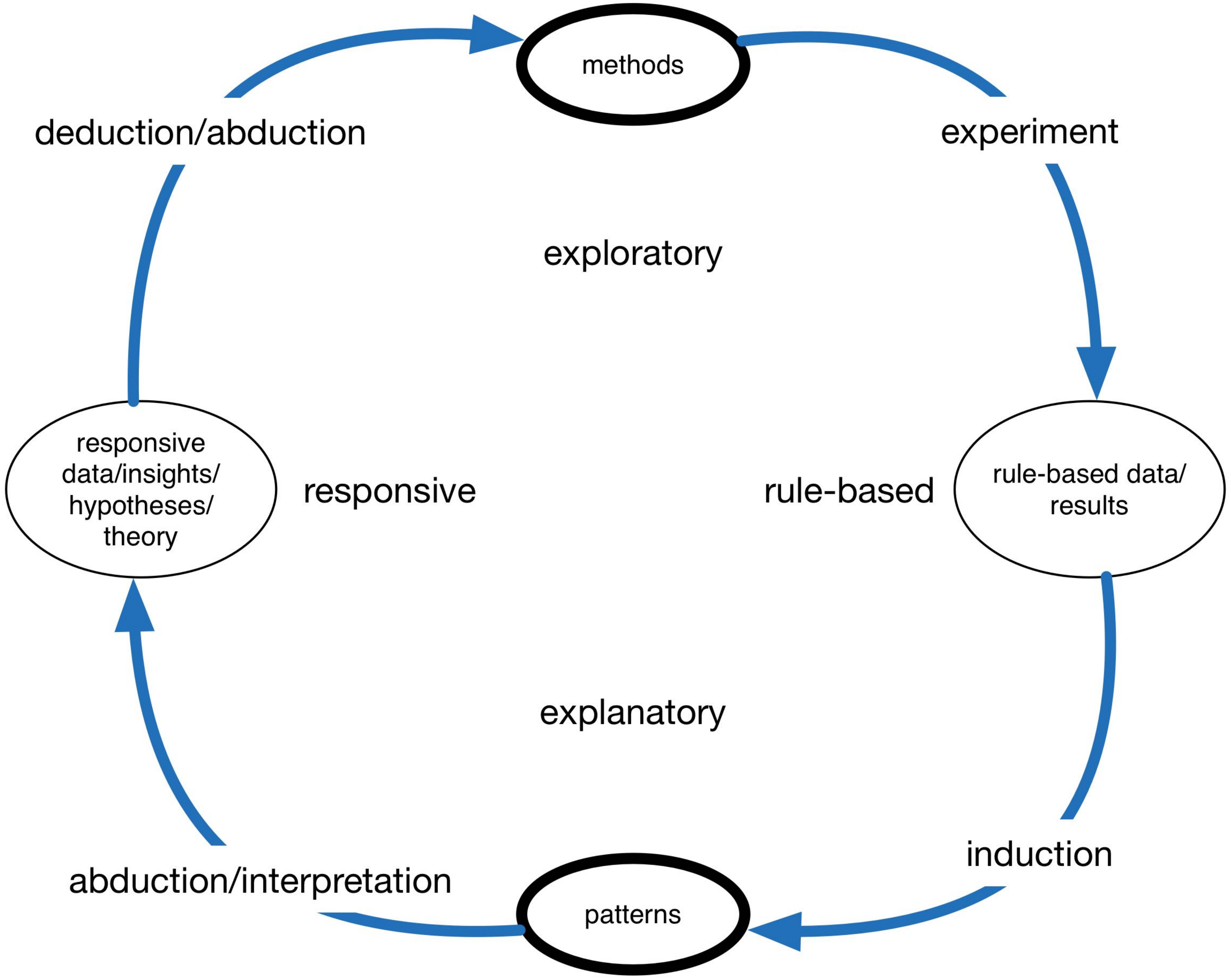






- confirmation of homophily principle, as topics clearly emerge just from the follow network
- block-graph on level 3, reminiscent of mainstream 'mountain' and special interest 'slopes and valleys' as postulated by Bruns (2008)
- issue-publics and smaller communities become visible on right, level 2

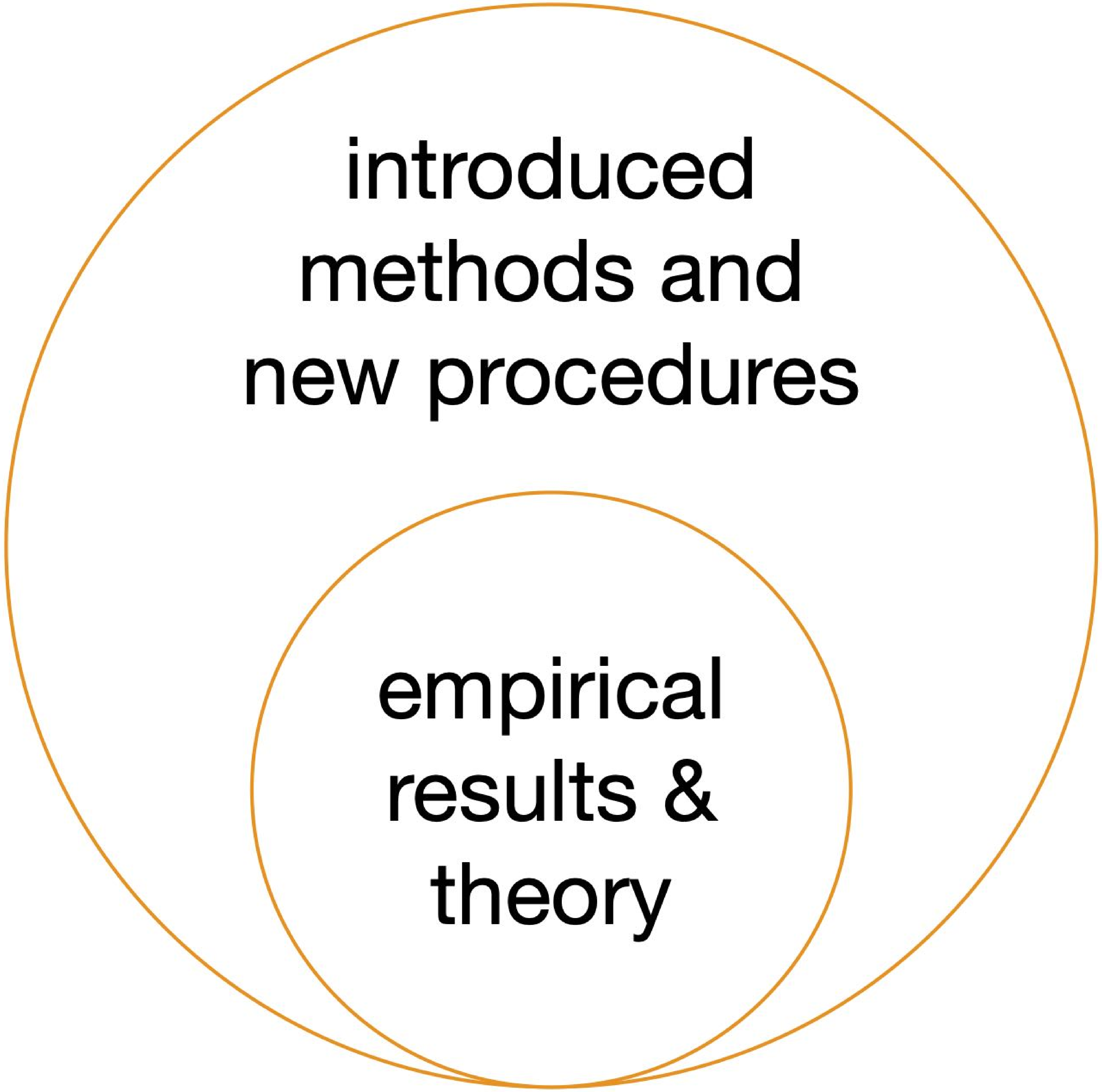
8. WHAT IS IT GOOD FOR?



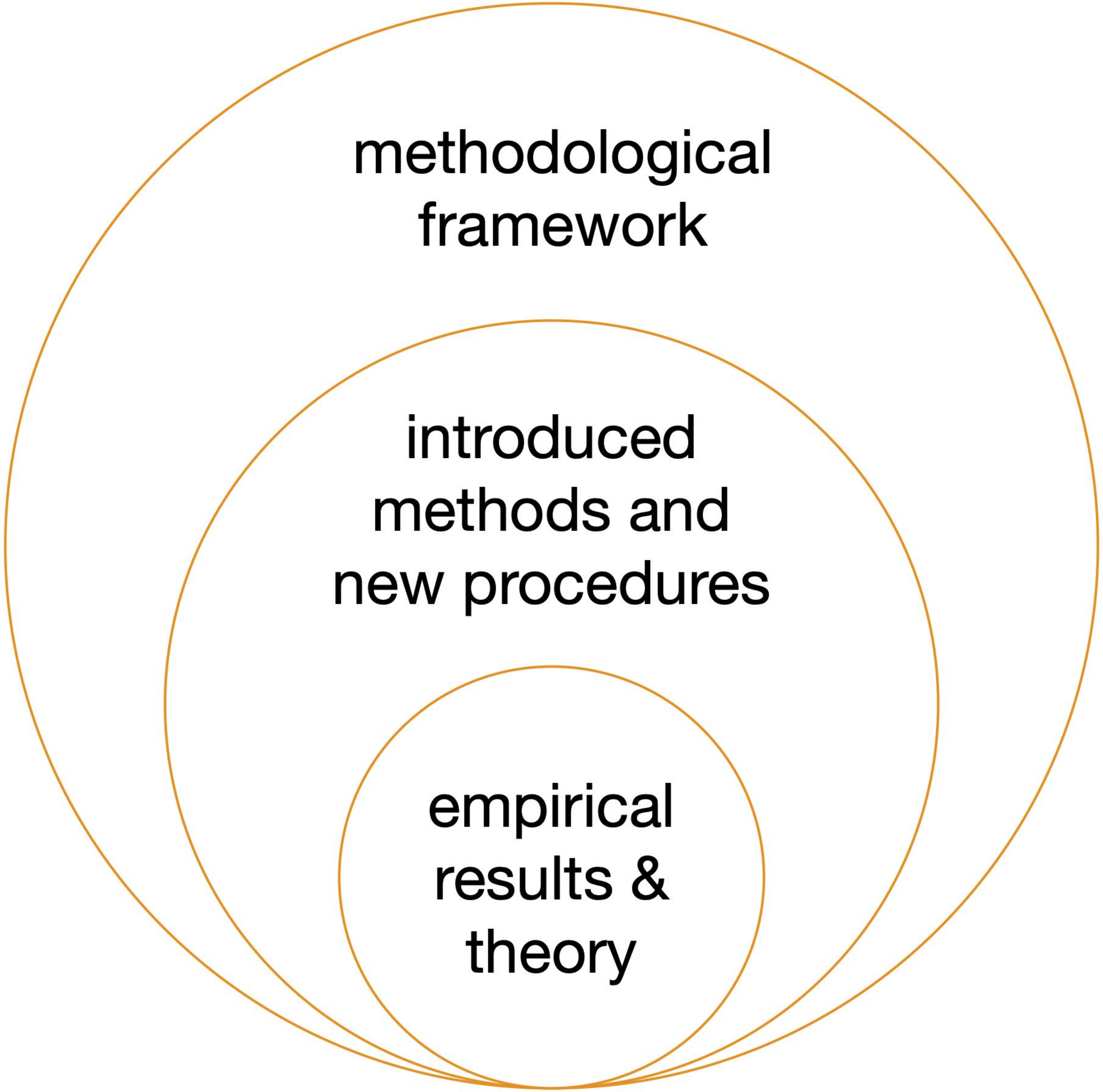
WHAT IS IT GOOD FOR?

empirical
results &
theory

WHAT IS IT GOOD FOR?


introduced
methods and
new procedures
empirical
results &
theory

WHAT IS IT GOOD FOR?


methodological framework
introduced methods and new procedures
empirical results & theory

CHALLENGE 1: HOW TO TRANSLATE THEORY?

- mostly straightforward: model entities as nodes, interactions as edges(, context as layers), or semantically: subjects/nouns as actors, actions/verbs as links
- but: good background knowledge about network analysis methods as well as theories needed to construct useful models
- especially suitable for theories about interaction patterns between individuals/entities, e.g., information diffusion, audience behaviour, publics, and communities
- less suitable for theories that deal with phenomena beyond observable interactions, e.g., media effects (hint: Machine Learning ☝️)

CHALLENGE 2: HOW TO APPLY NETWORK SCIENCE?

- methodological (i.e., epistemological + teleological) framework needed for sustainable integration of methods and theory
- 'rule-based' researchers need to:
  - dare to 'jump over the abductive gap', interpret results
  - understand context
- 'responsive' researchers need:
  - true understanding, not only application, of rule-based methods as providers of patterns
  - more math and programming skills

Practical steps (variations possible):

1. Translating theory;
2. model and collect data; (<-- practical 1)
3. exploration (e.g., visualisation); (<-- practical 2)
4. verification (measuring/content analysis); (<-- practical 3)
5. validation with (mathematical/agent-based) models and experiments;
6. revise theory, rinse, repeat

CHALLGENGE 3: BUT WHY?

- Network science is able to generate, and detect, patterns that can drive the abductive generation, and the inductive/deductive verification of hypotheses central to theory about public communication, especially on a society-wide scale.
- Network science is particularly important for dealing with problems of organised complexity.
- Public communication faces society-wide challenges (hate-speech, extremism, fake news, polarisation, …) due to the increased structural/organised complexity of a networked public sphere.


AND WE'VE GOT
TO DO SOMETHING TO
STRAIGHTEN THIS MESS OUT.

REFERENCES / FURTHER READING

Berry, D. M. (2011). The computatonal turn: thinking about the digital humanities. Culture Machine, 12, 1–22. Retrieved from http://www.culturemachine.net/index.php/cm/article/view/440/470

Bruns, A. (2008). Life beyond the public sphere: towards a networked model for political deliberation. Information Polity, 13, 71–85.

Bruns, A. (2017). Echo Chamber? What Echo Chamber? Reviewing the Evidence. In Future of Journalism 2017. Cardiff, UK. Retrieved from https://eprints.qut.edu.au/113937/

Bruns, A., Moon, B., Münch, F., & Sadkowsky, T. (2017). The Australian Twittersphere in 2016: Mapping the Follower/Followee Network. Social Media + Society, 3(4). http://doi.org/10.1177/2056305117748162

Dixon, D. (2012). Analysis tool or research methodology: is there an epistemology for patterns? In Understanding Digital Humanities. Palgrave Macmillan.

Goel, S., Anderson, A., Hofman, J., & Watts, D. (2015). The structural virality of online diffusion. Management Science. Retrieved from http://dx.doi.org/10.1287/mnsc.2015.2158

Habermas, J. (2006). Political communication in media society: does democracy still enjoy an epistemic dimension? The impact of normative theory on empirical research. Communication Theory, 16(4), 411–426. http://doi.org/10.1111/j.1468-2885.2006.00280.x

Halavais, A. (2015). Bigger sociological imaginations: framing big social data theory and methods. Information, Communication & Society, 18(5), 583–594. http://doi.org/10.1080/1369118X.2015.1008543

Luhmann, N. (1997). Die Gesellschaft der Gesellschaft. Frankfurt am Main: Suhrkamp.

McQuail, D. (2010). Mass communication theory (6th ed.). SAGE Publications.

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

Onwuegbuzie, A. J., & Leech, N. L. (2005). On Becoming a Pragmatic Researcher: The Importance of Combining Quantitative and Qualitative Research Methodologies. International Journal of Social Research Methodology, 8(5), 375–387. http://doi.org/10.1080/13645570500402447

Peixoto, T. P. (2017). The graph-tool python library. http://doi.org/10.6084/M9.FIGSHARE.1164194.V14

Sechrest, L., & Sidani, S. (1995). Quantitative and qualitative methods: Is There an Alternative? Evaluation and Program Planning, 18(1), 77–87. http://doi.org/10.1016/0149-7189(94)00051-X

Weaver, W. (1948). Science and complexity. American Scientist, 36(4), 536–544. Retrieved from http://www.jstor.org/stable/27826254

THANKS!

🖖

TWITTER/MASTODON: @FLXVCTR(@MAS.TO)