Celina Curry
“Crowd”

# Climate change and online social networks

Hywel Williams
7th May 2014

# Content

- Collective behaviour & online social data
- Social network analysis of online climate debate
- Comment on “big data”

Taksim Square, Istanbul -15th June 2013


2 47

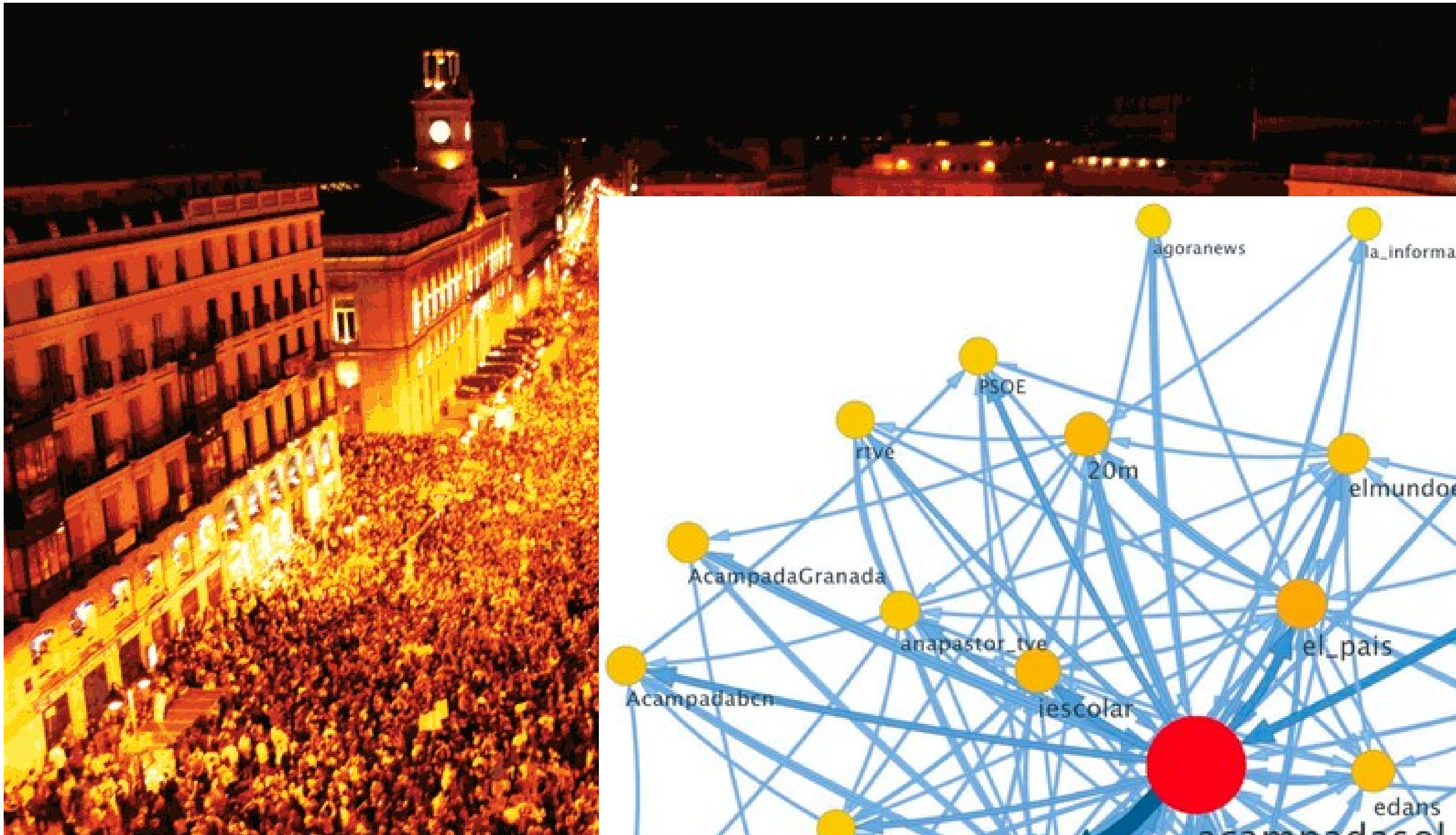

2 47

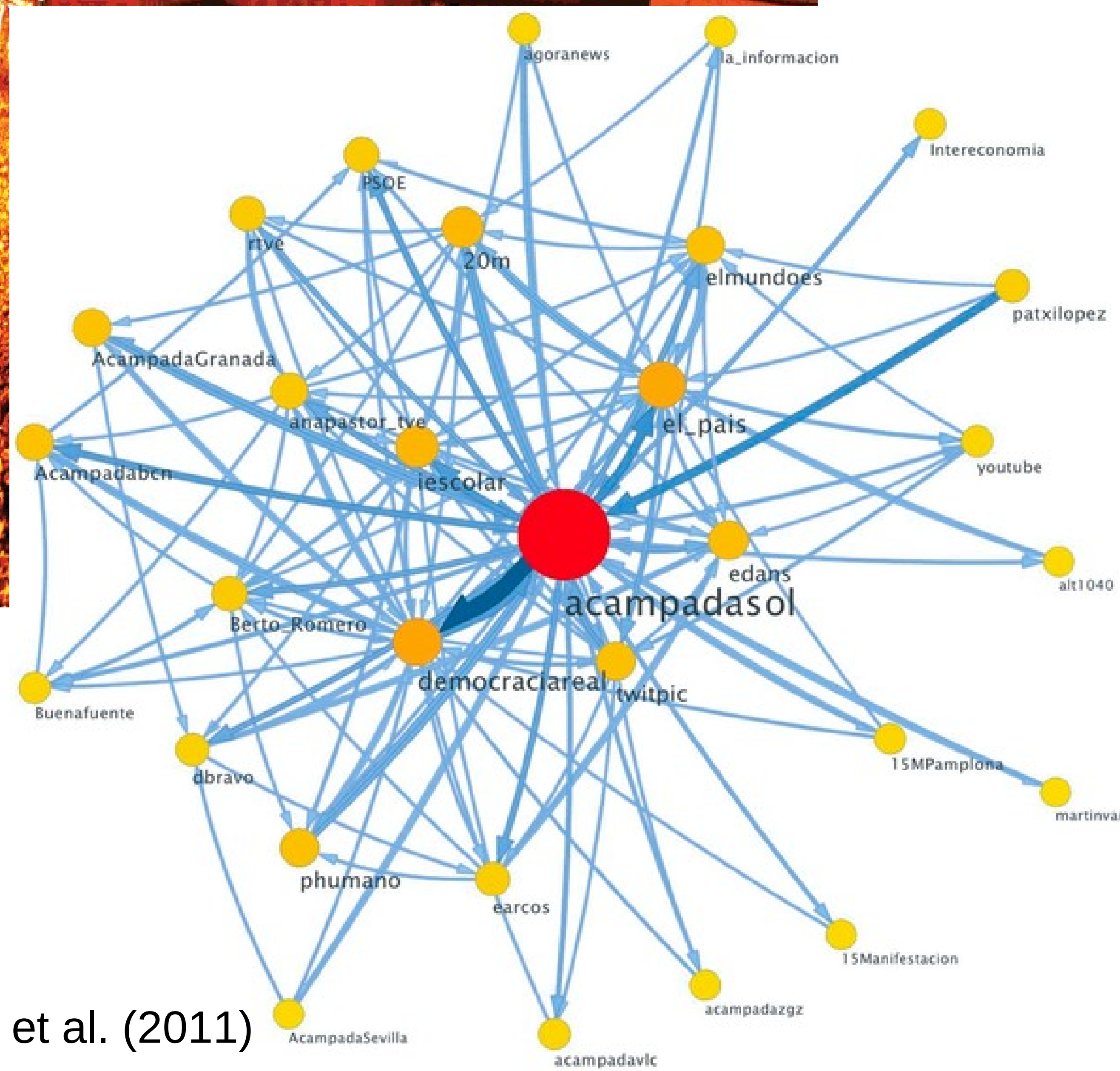
- 138% increase in tweet volume
- #TwitterisblockedinTurkey trends globally

# Social media 101: Erdogan's lesson

- Important communication channel(s)
- Relevant to “real-world” and offline political processes
- Decentralised
  - Many-to-many
  - No “centre”
  - No top-down control
  - Robust to interference

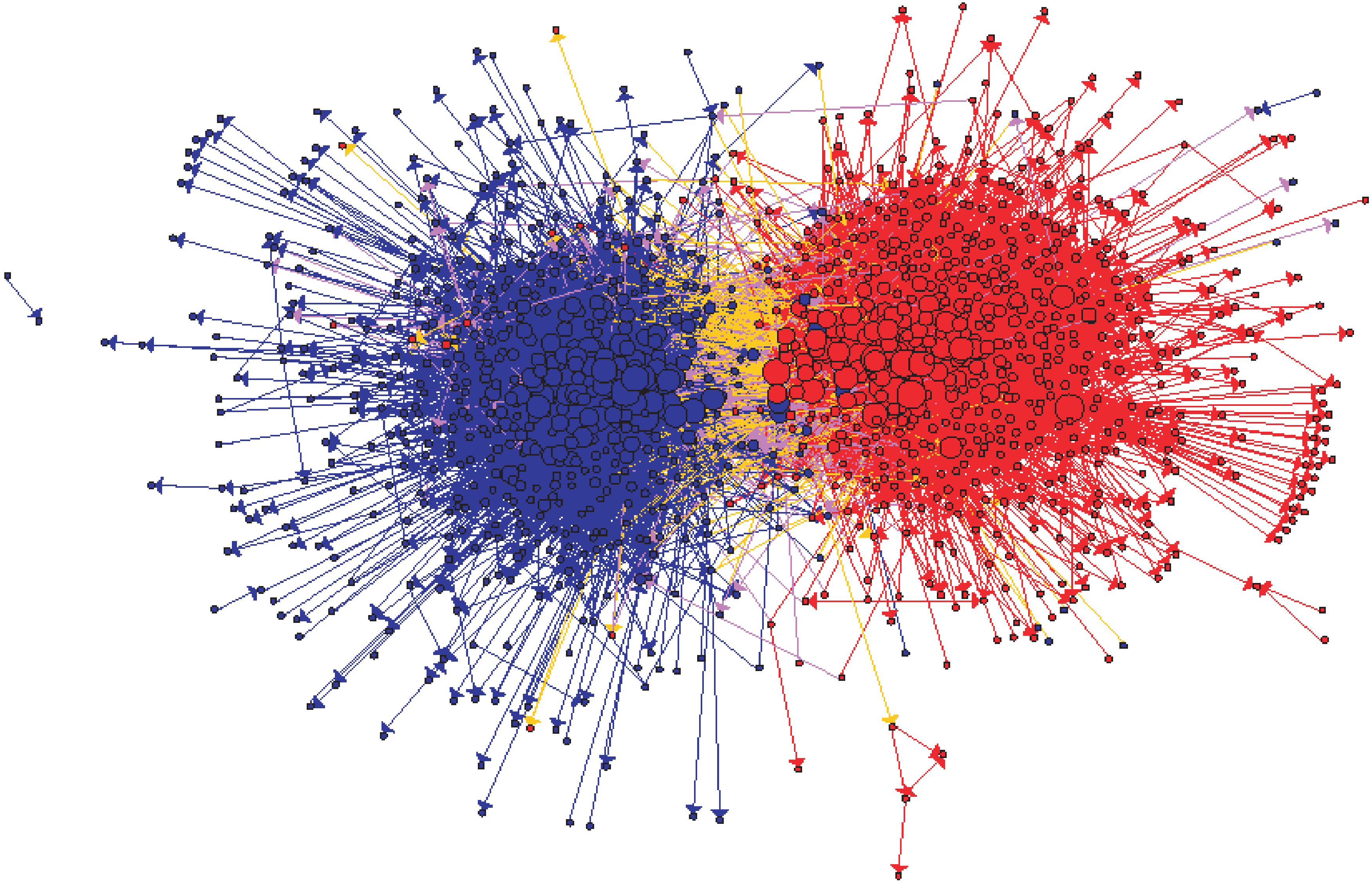
Plaza del Sol, Madrid, 21st May 2011





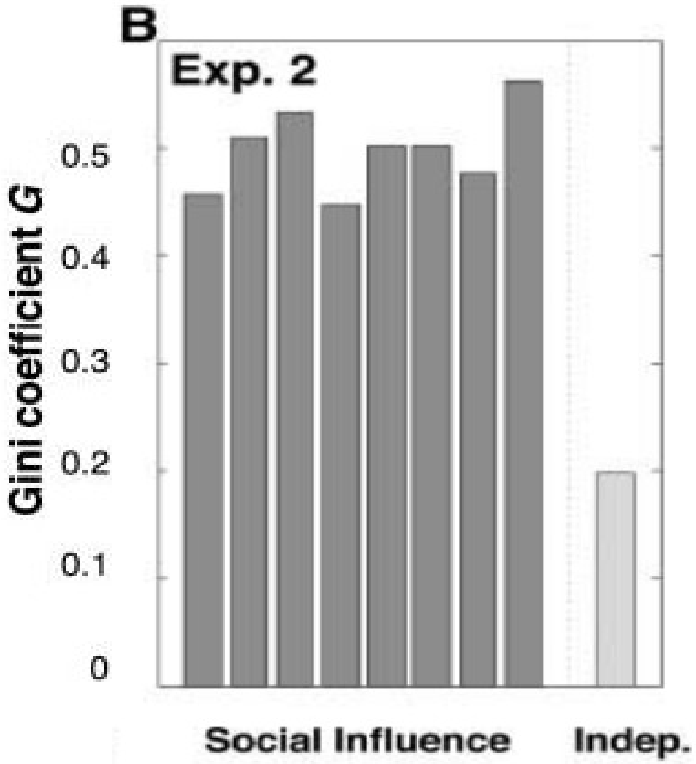


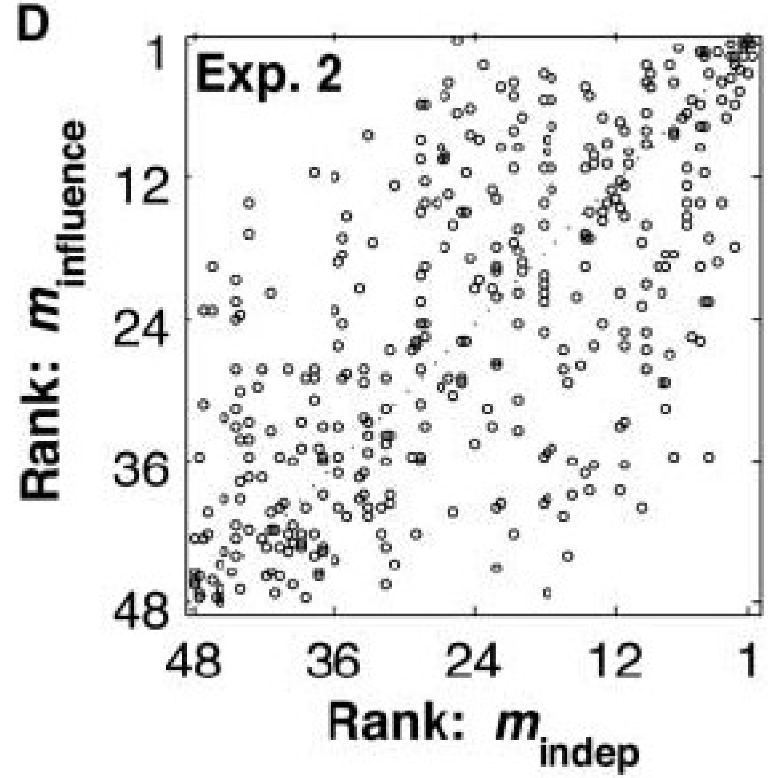
Borge-Holthoefer et al. (2011)



Adamic & Glance (2005)

# Social influence affects opinions



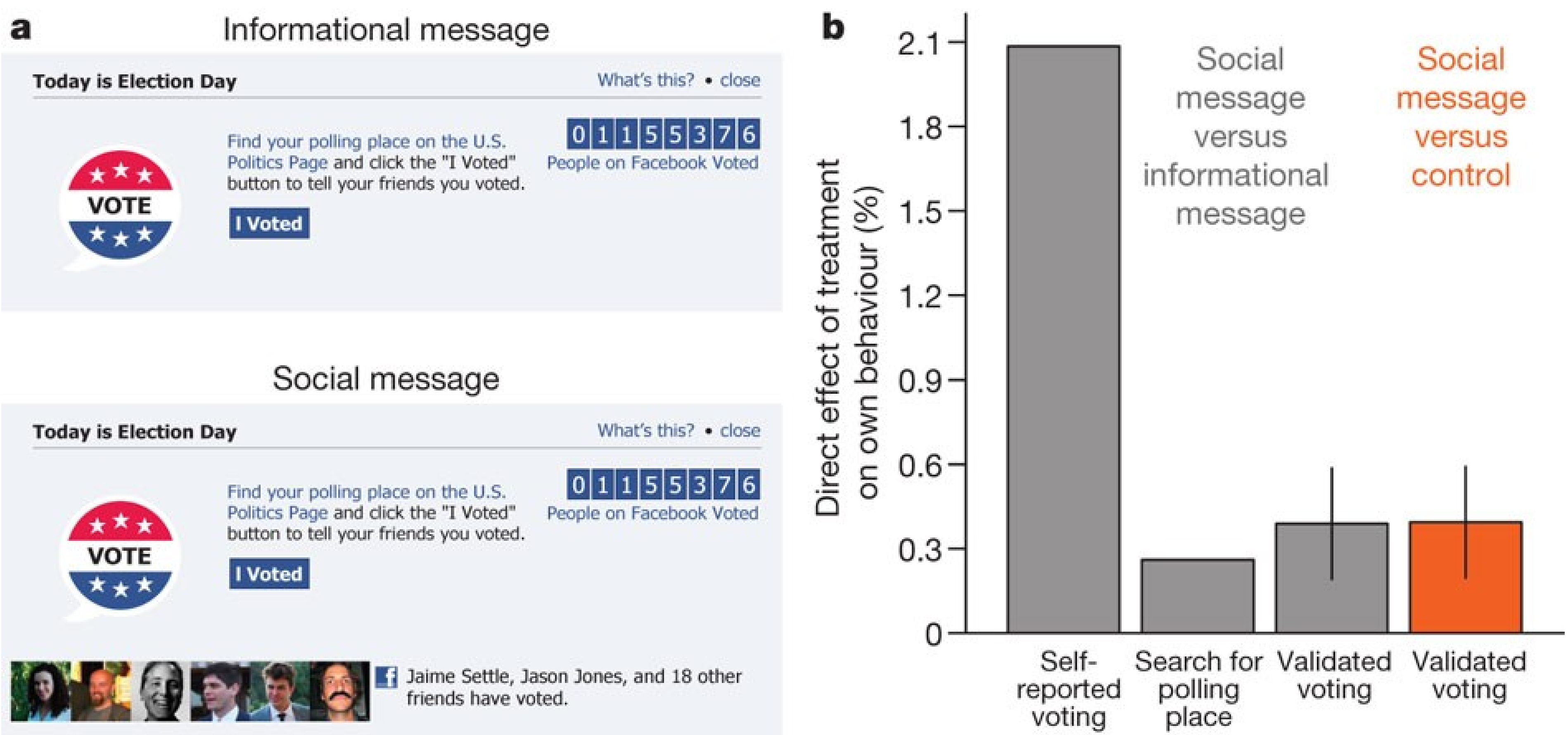




Artificial music market experiment (Salganik et al, 2006):

- Independent choice: Songs shown in random order.
- Social influence: Show number of downloads, order by popularity.

Salganik, Dodds & Watts (2006)

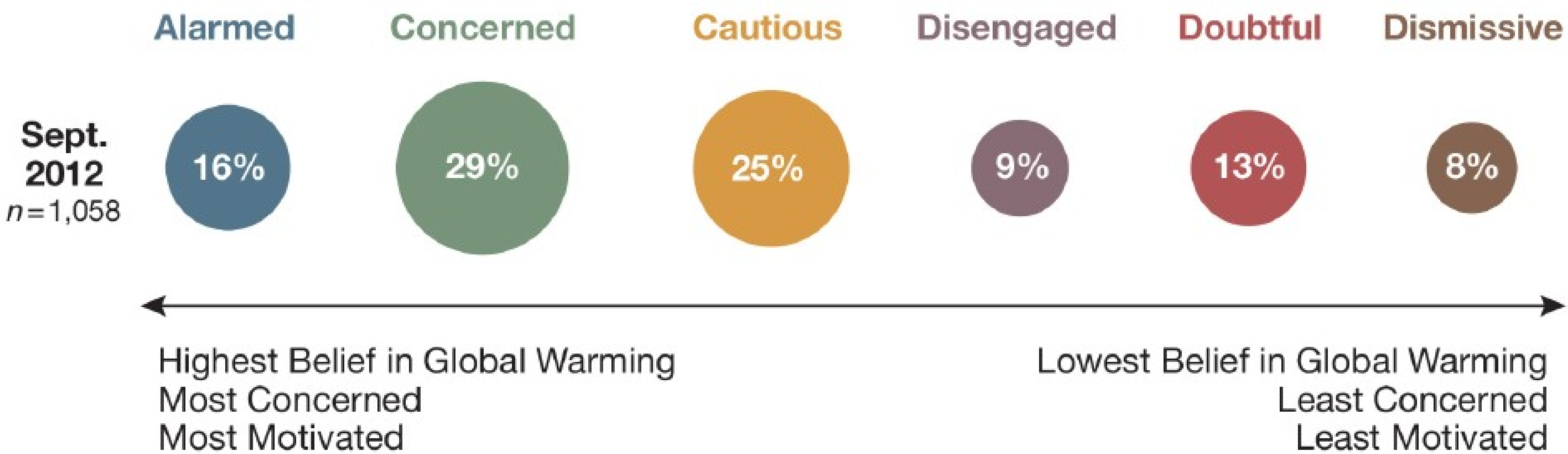
# Social influence affects behaviours




RM Bond *et al.* *Nature* **489**, 295-298 (2012) doi:10.1038/nature11421

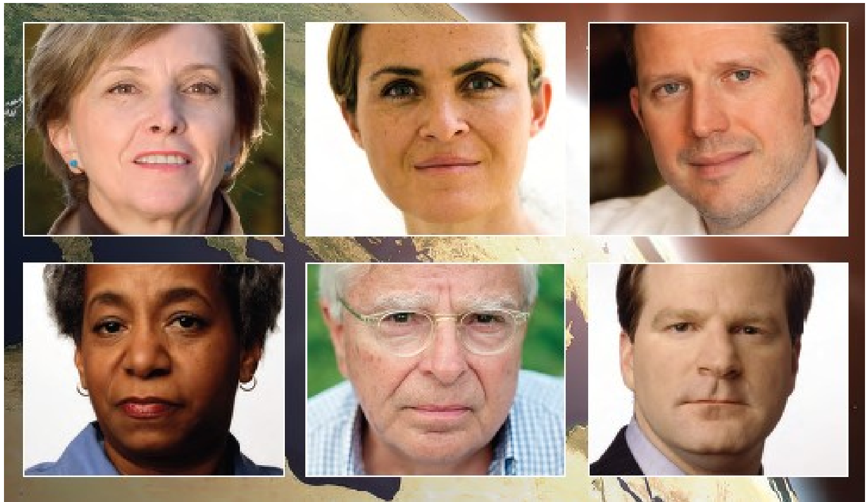

ipcc
INTERGOVERNMENTAL PANEL ON climate change
CLIMATE CHANGE 2013
The Physical Science Basis
WG I
WORKING GROUP I CONTRIBUTION TO THE
FIFTH ASSESSMENT REPORT OF THE
INTERGOVERNMENTAL PANEL ON CLIMATE CHANGE
WMO
UNEP

# Attitudes to climate change: the “Six Americas”




*Proportion represented by area*

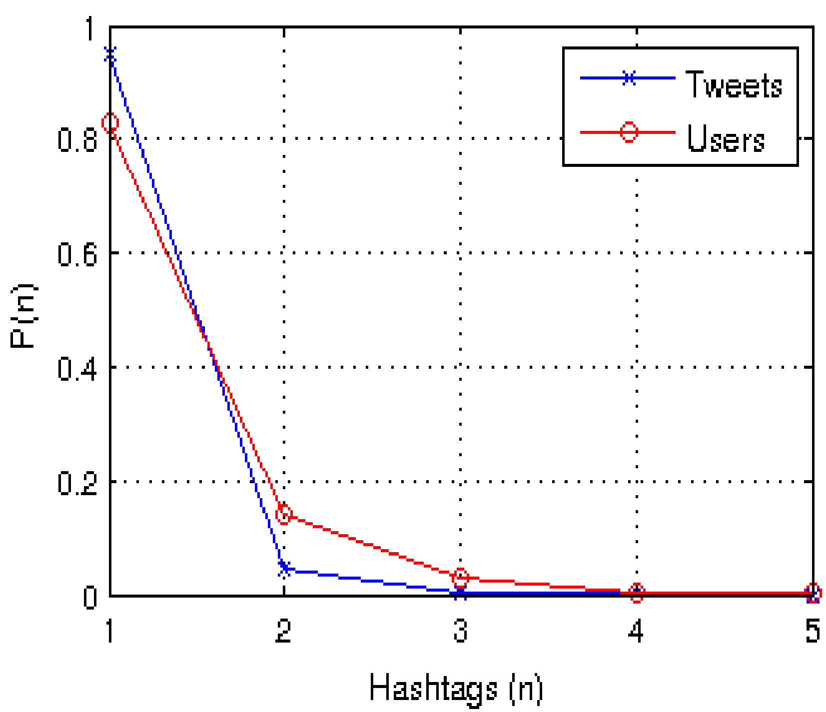
Source: Yale / George Mason University



# Twitter dataset collected Jan-May 2013

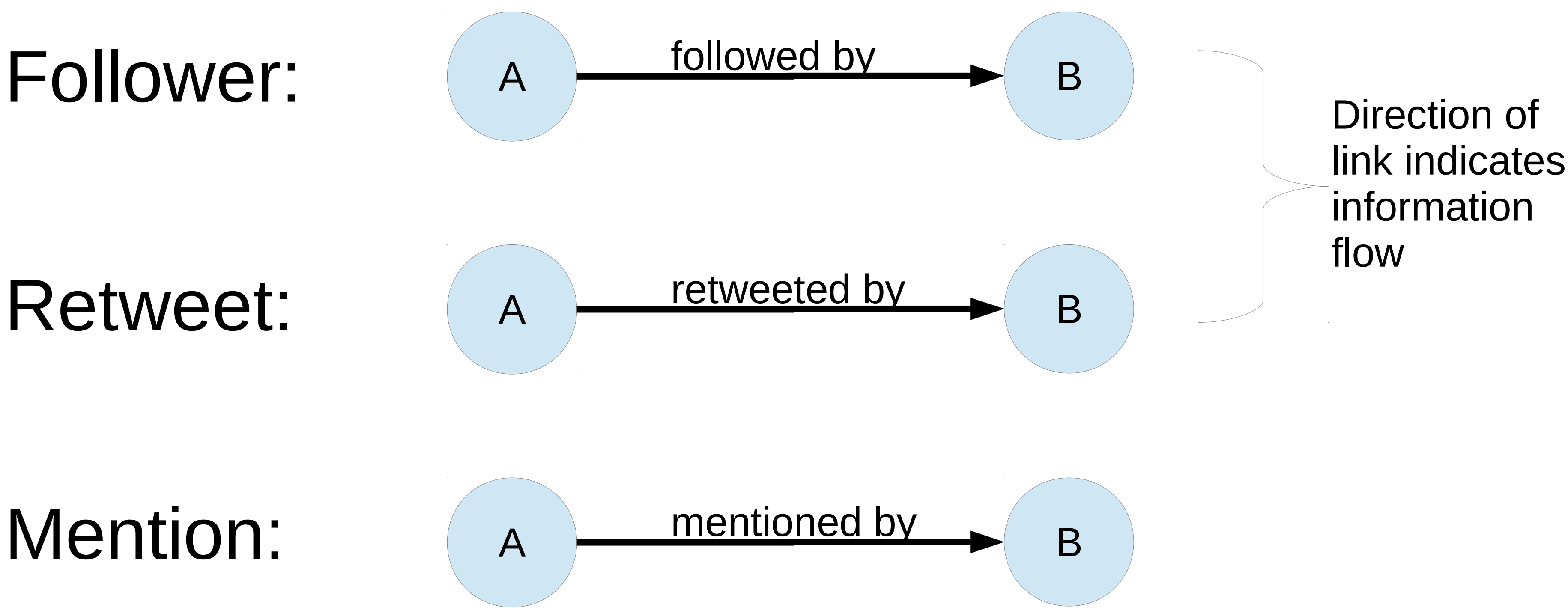
| Hashtag | Total tweets | Retweets | | Mentions | | Links | | Unique users |
|---|---|---|---|---|---|---|---|---|
| | | count | % total | count | % total | count | % total | |
| #globalwarming | 92190 | 21475 | 23.29 | 17226 | 18.69 | 43864 | 47.58 | 56517 |
| #climatechange | 230753 | 93618 | 40.57 | 52363 | 22.69 | 163069 | 70.67 | 86366 |
| #agw | 16987 | 4599 | 27.07 | 4238 | 24.95 | 13306 | 78.33 | 3115 |
| #climate | 280076 | 118008 | 42.13 | 61545 | 21.97 | 236200 | 84.33 | 70011 |
| #climaterealists | 1427 | 254 | 17.80 | 31 | 2.17 | 1411 | 98.88 | 208 |
| All hashtags | 590608 | 230120 | 38.96 | 129342 | 21.90 | 431340 | 73.03 | 179180 |

- Five climate-related “hashtags”
- Counted retweets, mentions, links, unique users



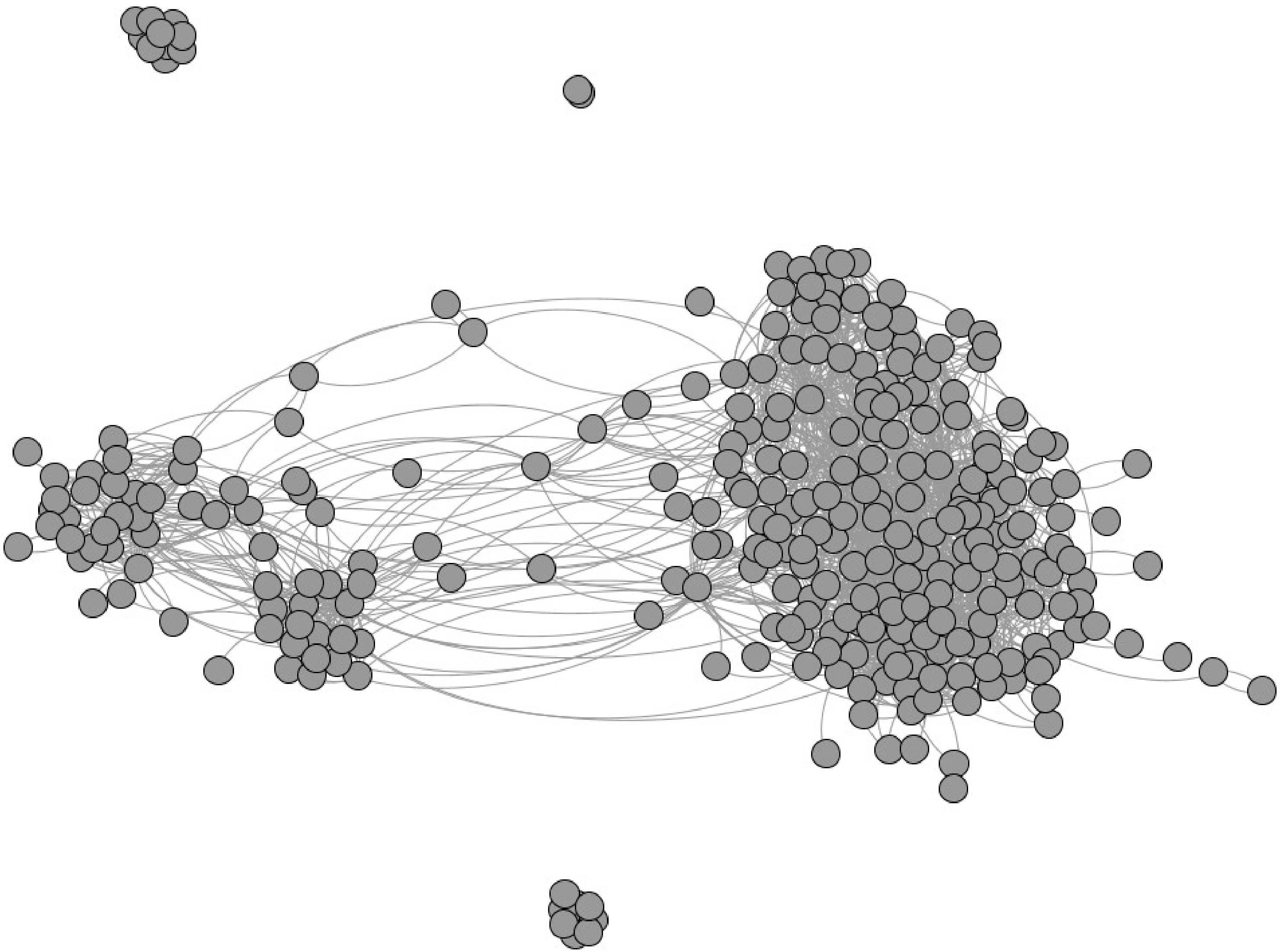

Is it big enough?

# Three forms of interaction

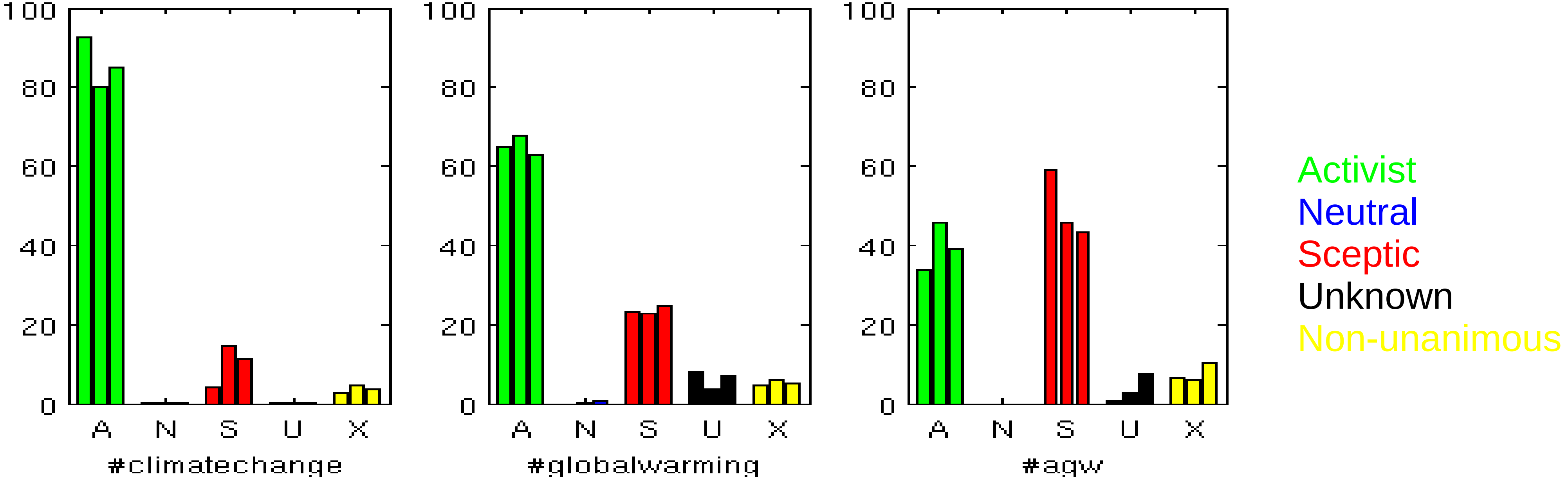


# Follower network: #globalwarming

(filtered by tweets/user, force-directed layout)

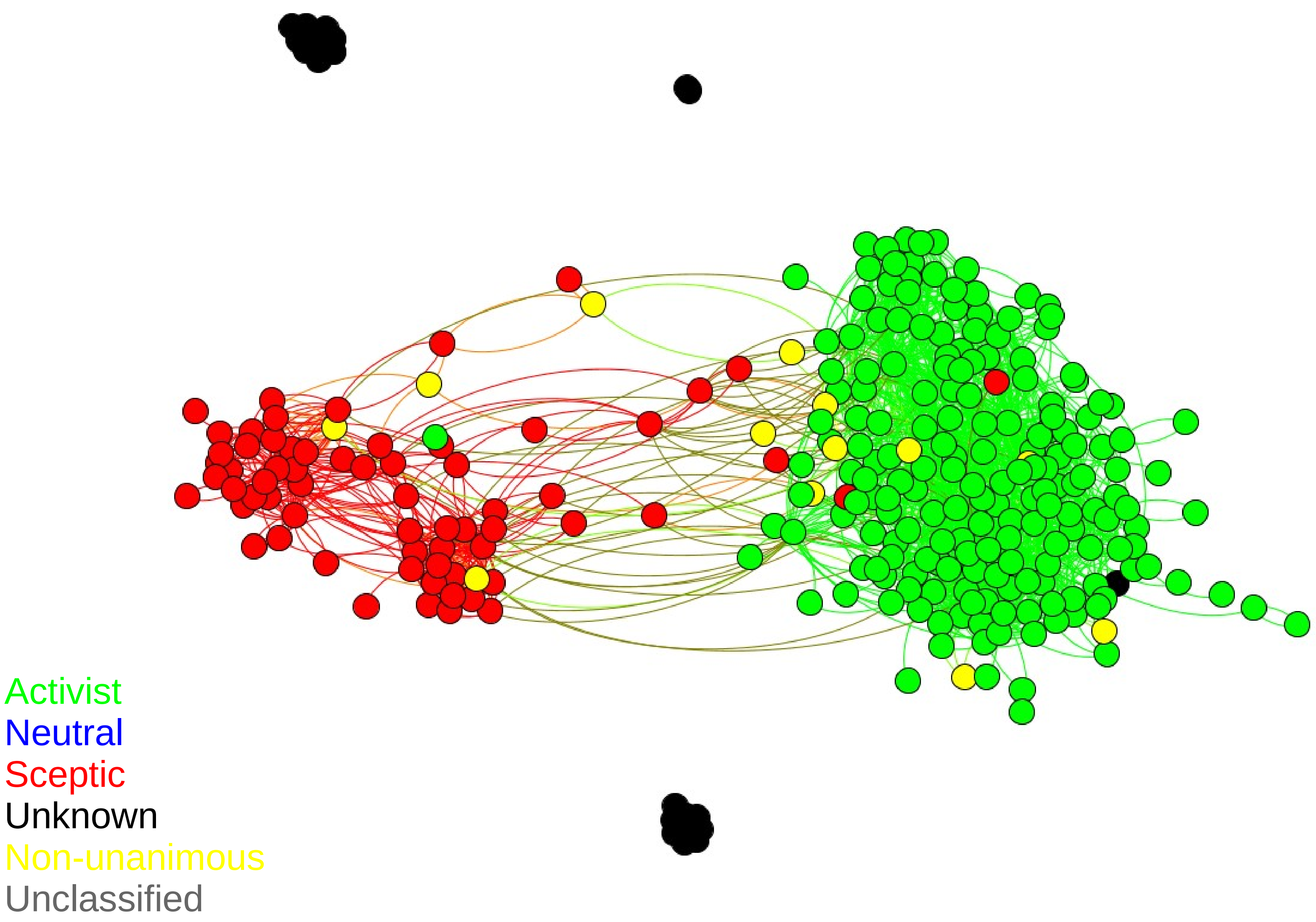


# Classification of user attitudes

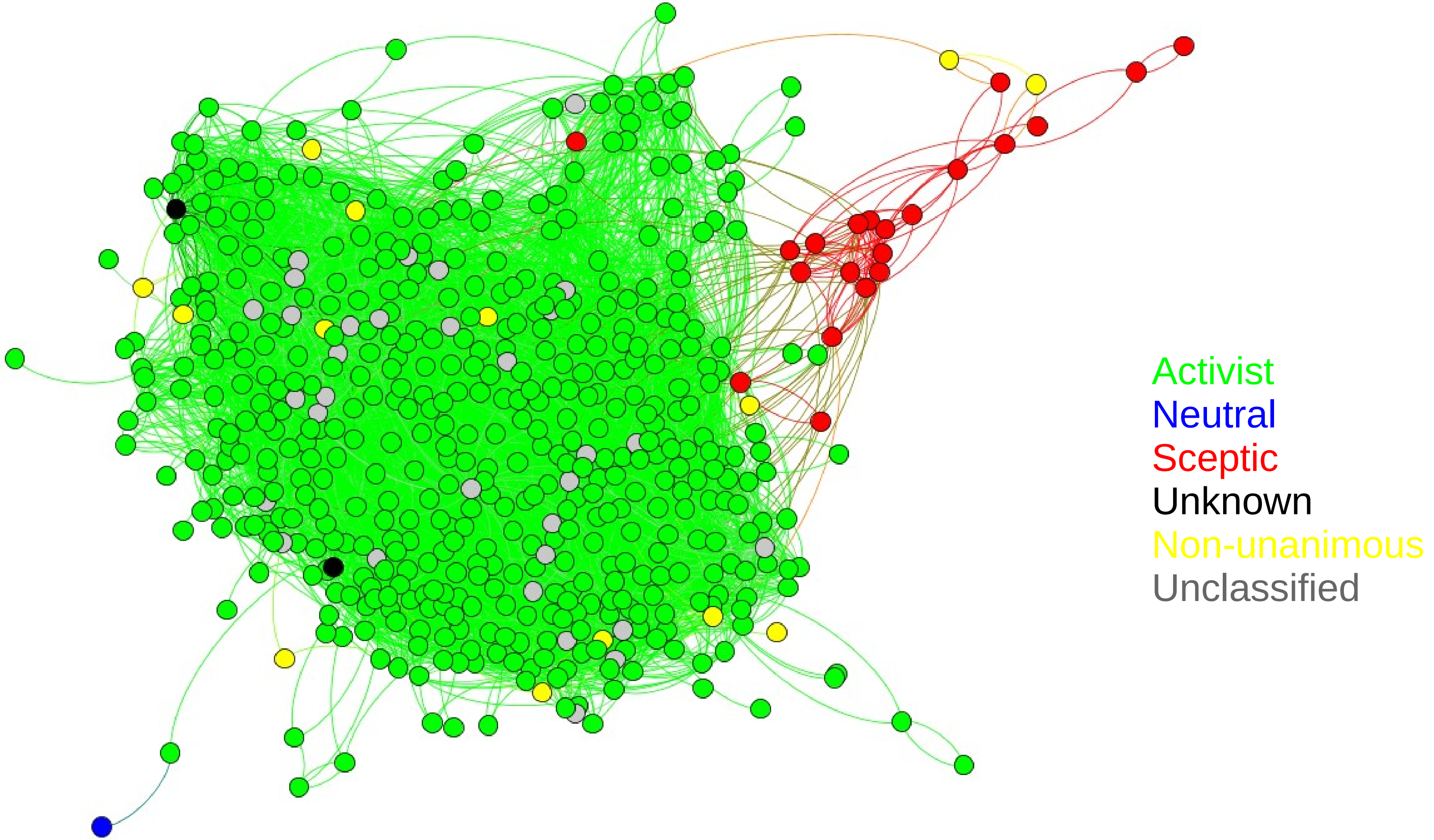



- Panel of 3 climate change PhD students classified users (n=1545) from profiles and tweets:
  - **Activist:** supporting mainstream climate science and/or promoting climate-friendly policies
  - **Neutral**: expressing a view on climate change, but not obviously activist or sceptic
  - **Sceptic:** contrarian view on climate science and/or critical of climate-friendly policies
  - **Unknown**: no view or attitude on climate change could be distinguished
- Only accept unanimous classifications

# Follower network: #globalwarming

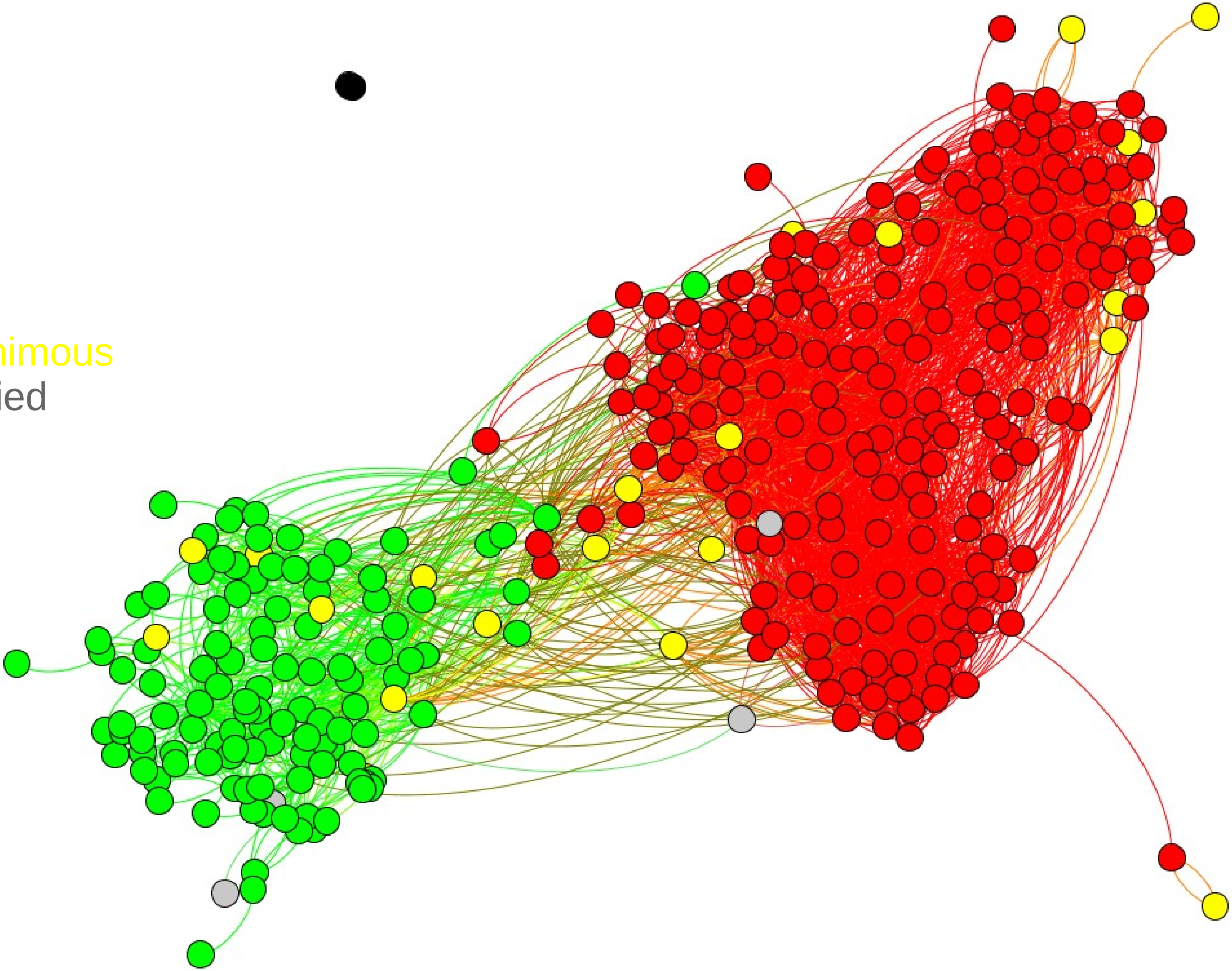


# Follower network: #climatechange



# Follower network: #agw

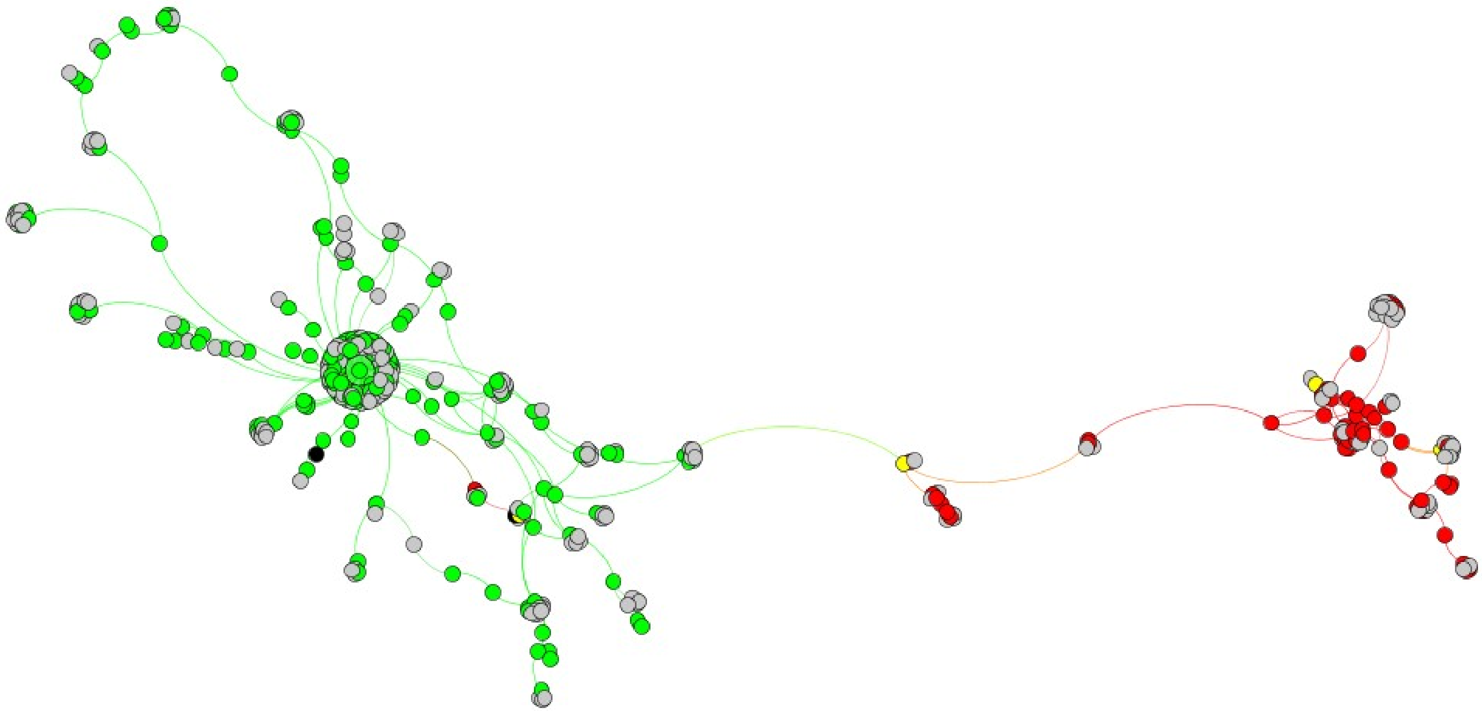
Activist
Neutral
Sceptic
Unknown
Non-unanimous
Unclassified



# Retweet network: #climatechange

(filtered by edge-weight, force-directed layout)

# Retweet network: #globalwarming



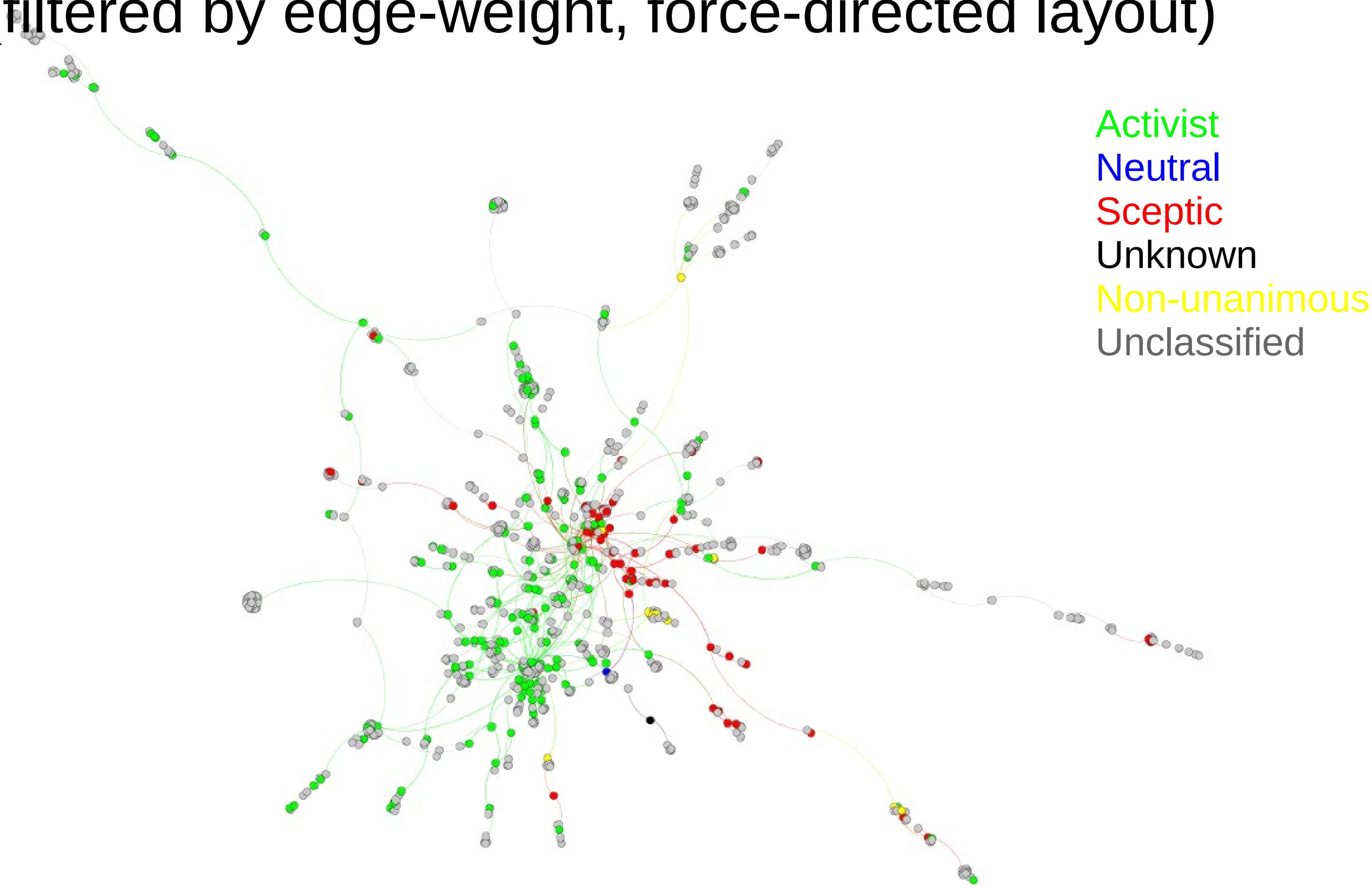
Activist
Neutral
Sceptic
Unknown
Non-unanimous
Unclassified

# Retweet network: #agw

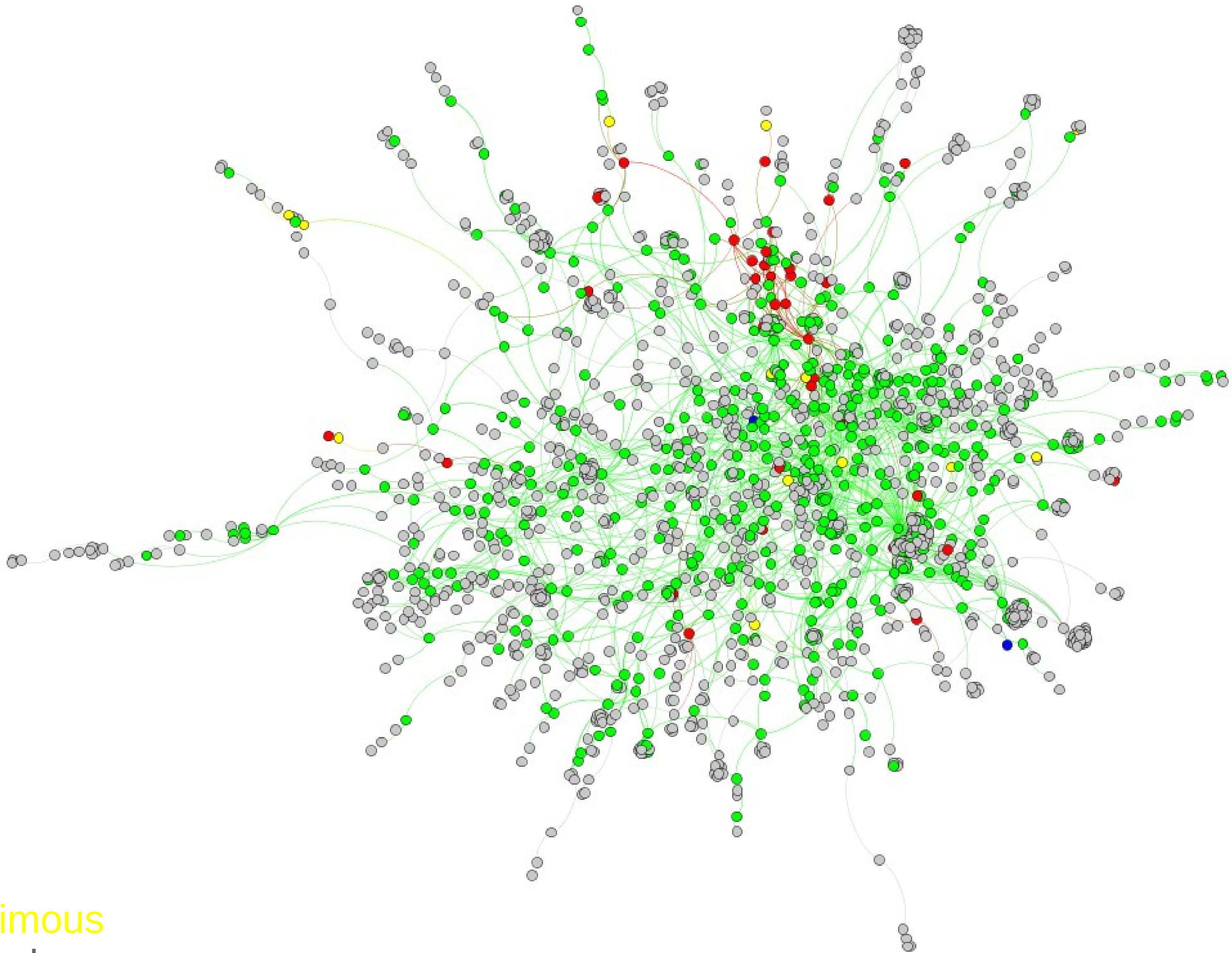
Activist
Neutral
Sceptic
Unknown
Non-unanimous
Unclassified

# Mention network: #globalwarming

(filtered by edge-weight, force-directed layout)

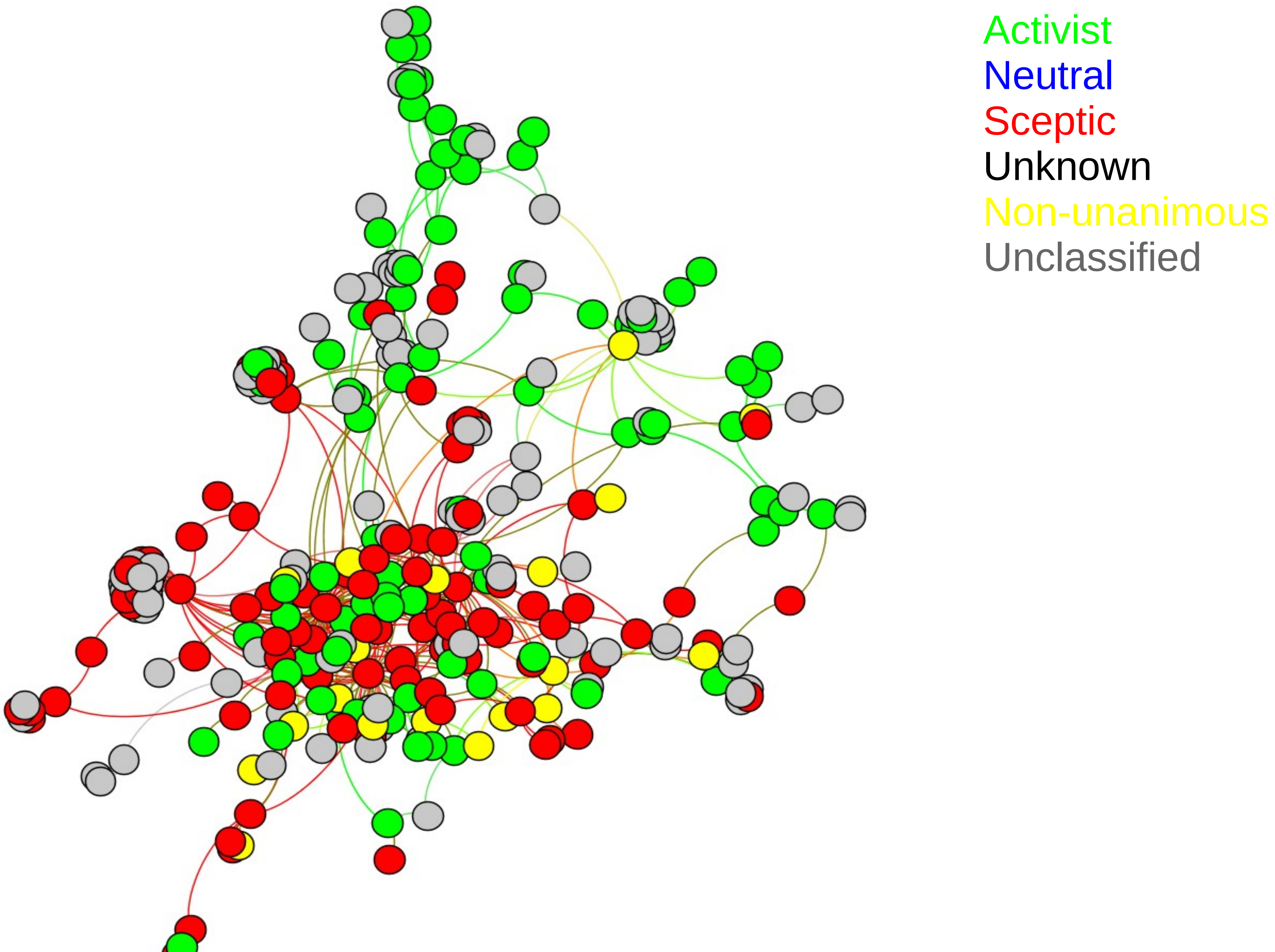


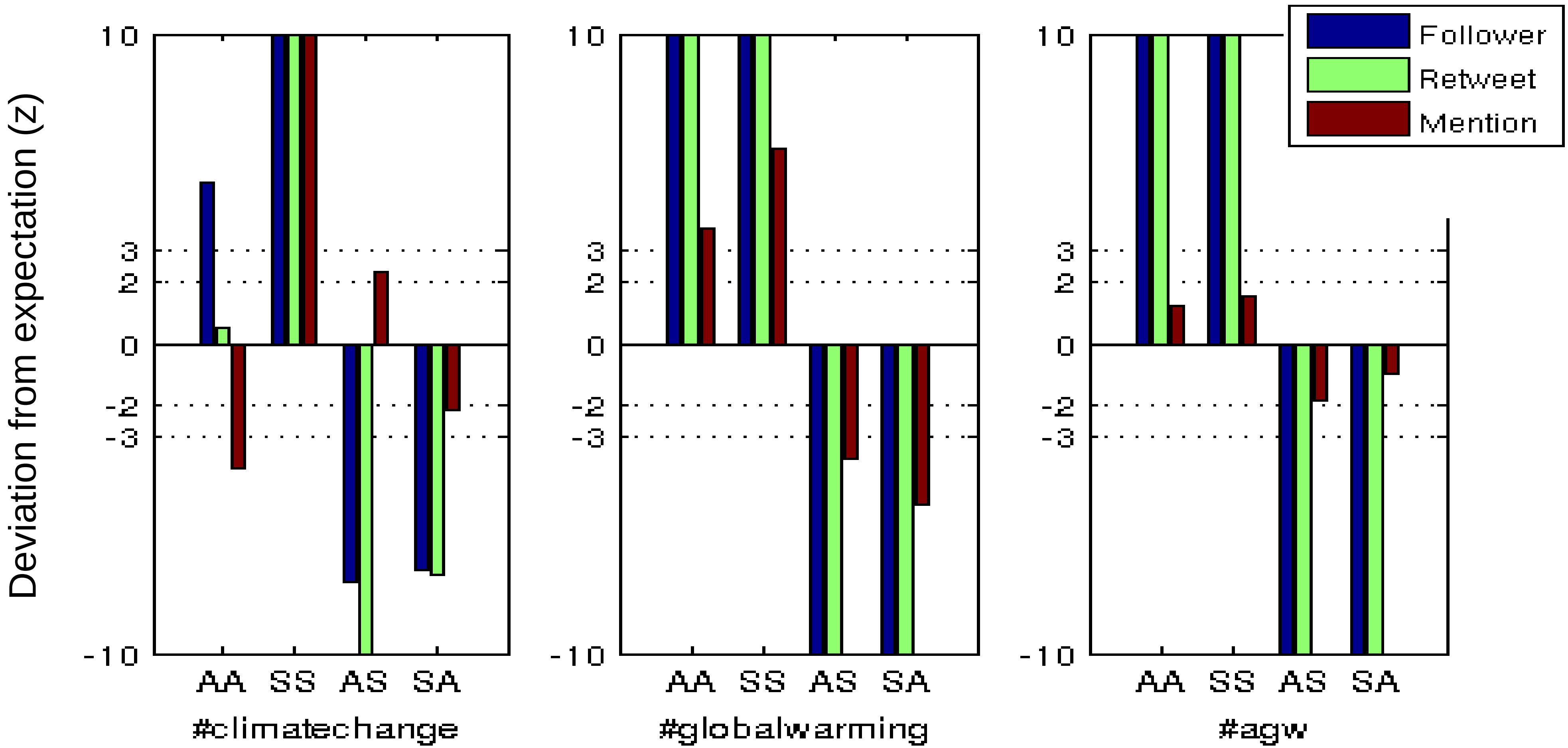
# Mention network: #climatechange



Activist
Neutral
Sceptic
Unknown
Non-unanimous
Unclassified

# Mention network: #agw





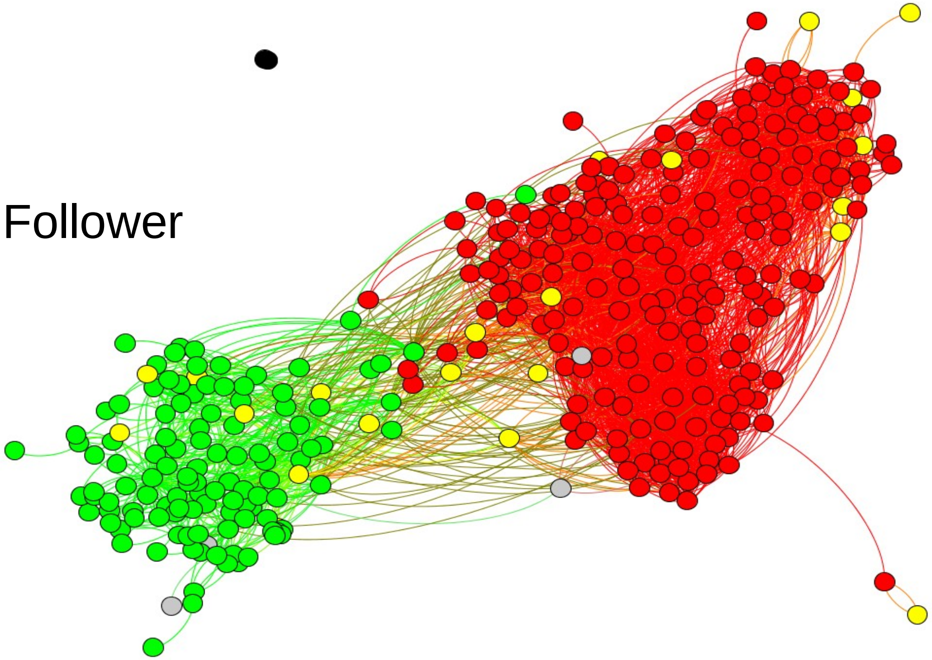

Bootstrap method:

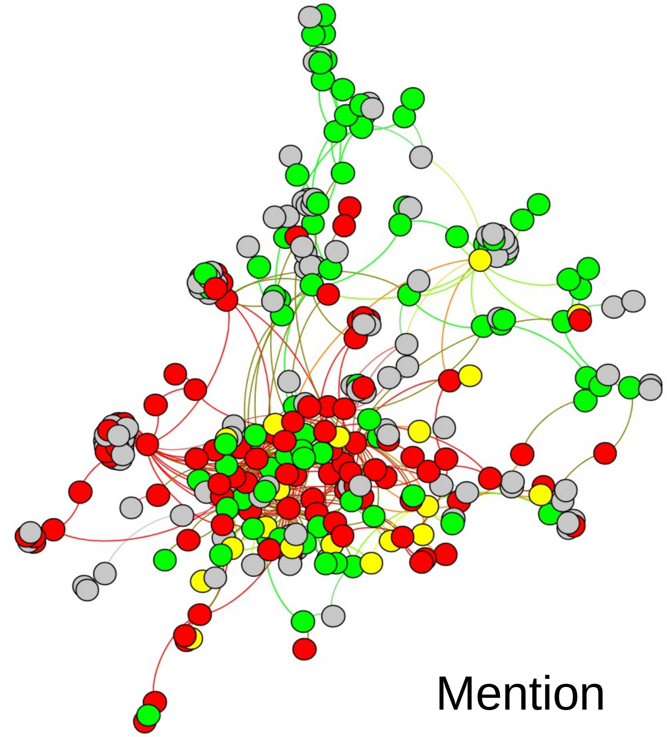
Homophily: Positive z-scores for AA & SS, negative AS & SA.

Significance: $|z|>2 \rightarrow p<0.05$, $|z|>3 \rightarrow p<0.003$ .

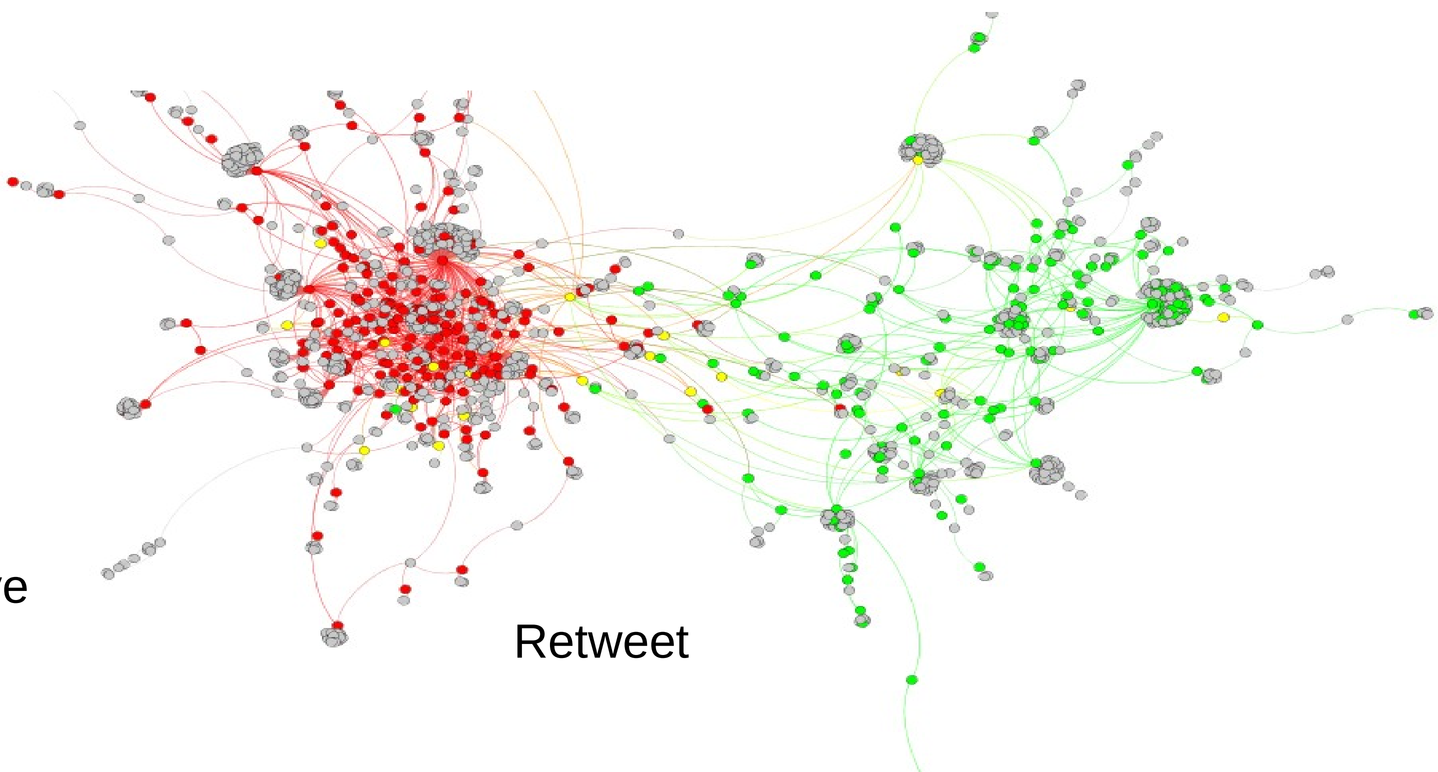
Strong significant homophily for followership and retweets. Mixed signal for mentions.

Follower





Mention



Retweet

Different interactions have different motivations?

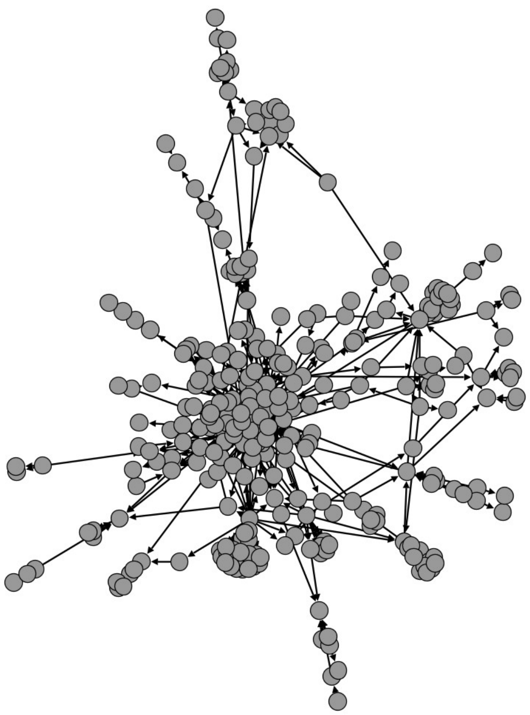

British youth subcultures.
The Guardian,
20th March 2014.

Metallers, goth, Molly Soda, haul girl, seapunks.

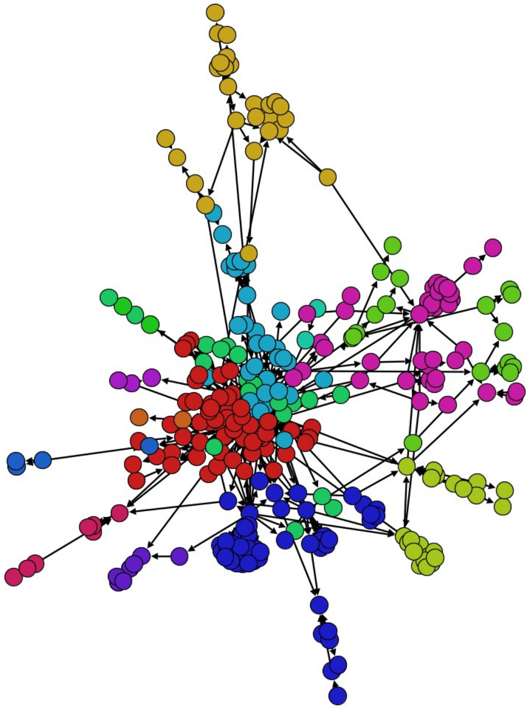
Followership, retweeting → approval/endorsement?

Mention → context dependent

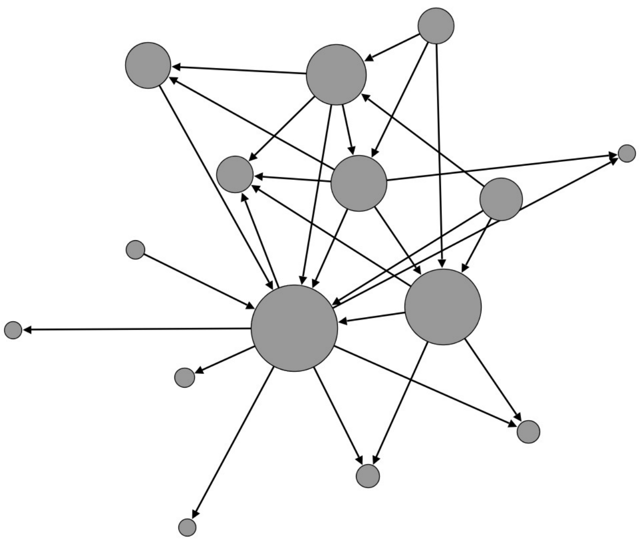
# Community detection



Raw network



Communities

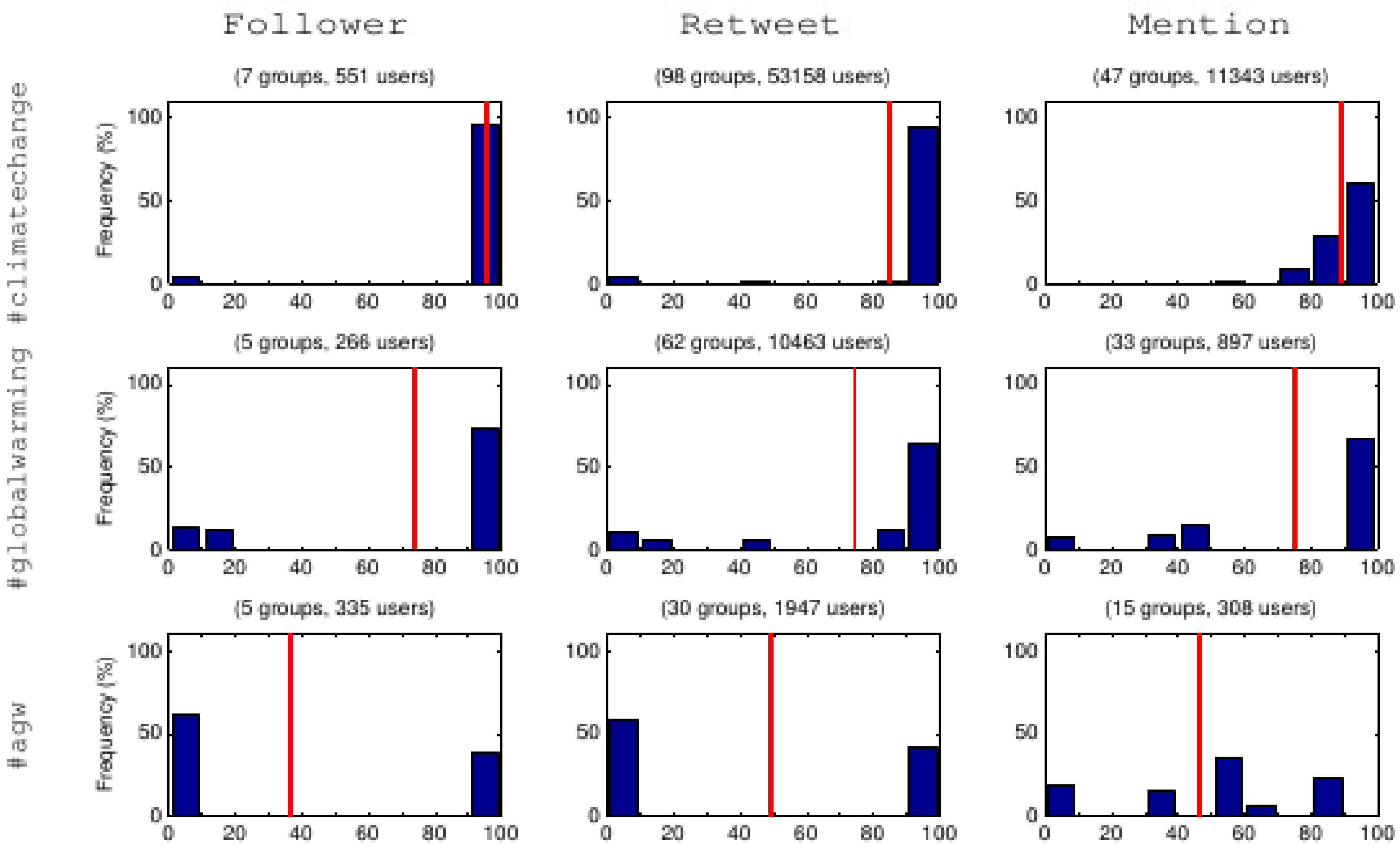


Induced graph

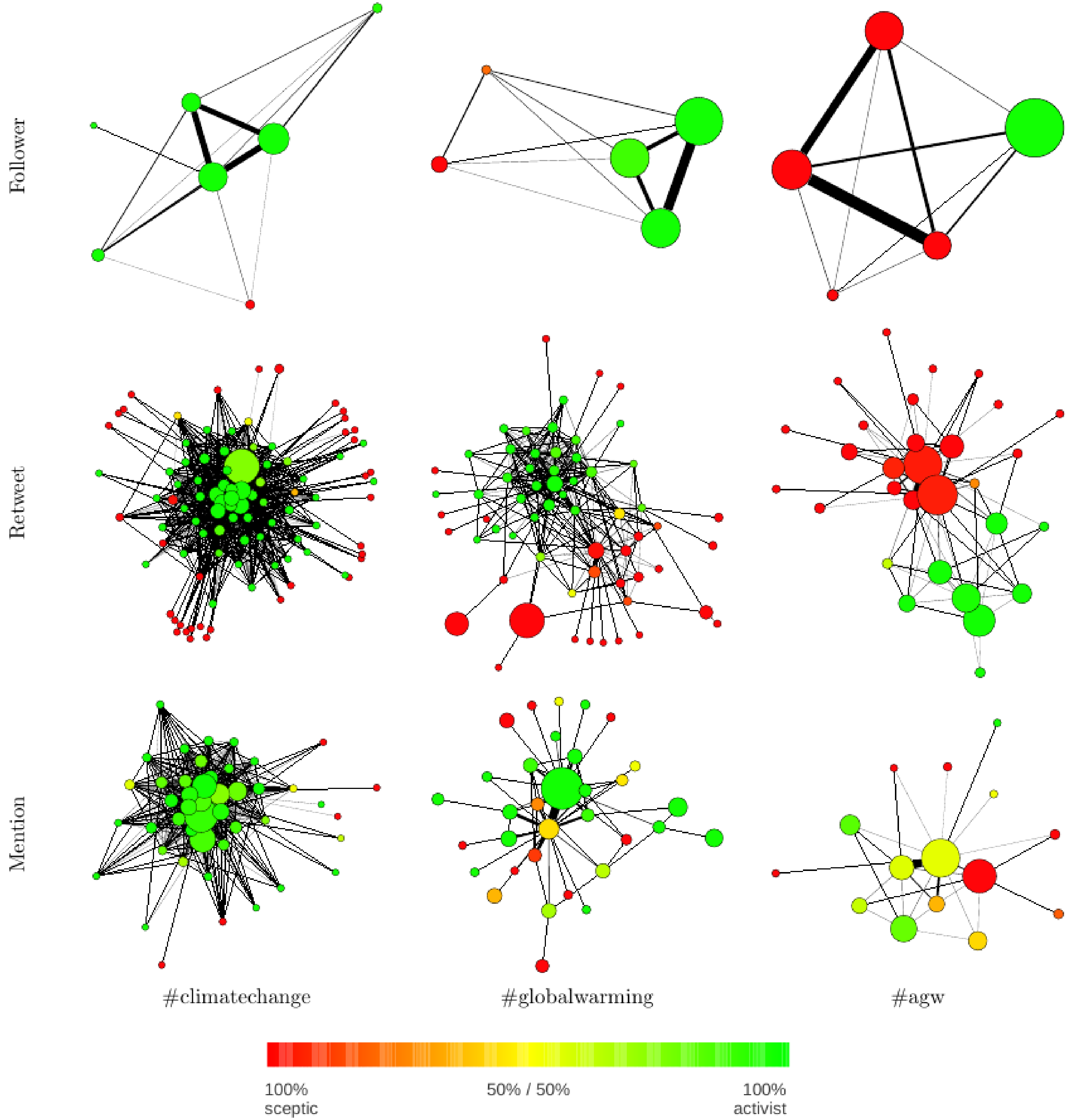
Use Louvain method (*) to **partition** nodes into distinct communities such that **modularity** (**) is maximised. Draw **induced graph** representing each community as a single node.

(*) Initialise with each node in its own community. Sequentially merge neighbouring communities if doing so will increase overall modularity, until no further increase is possible.
(**) Modularity: high when all edges fall within groups, low when edges fall between groups.

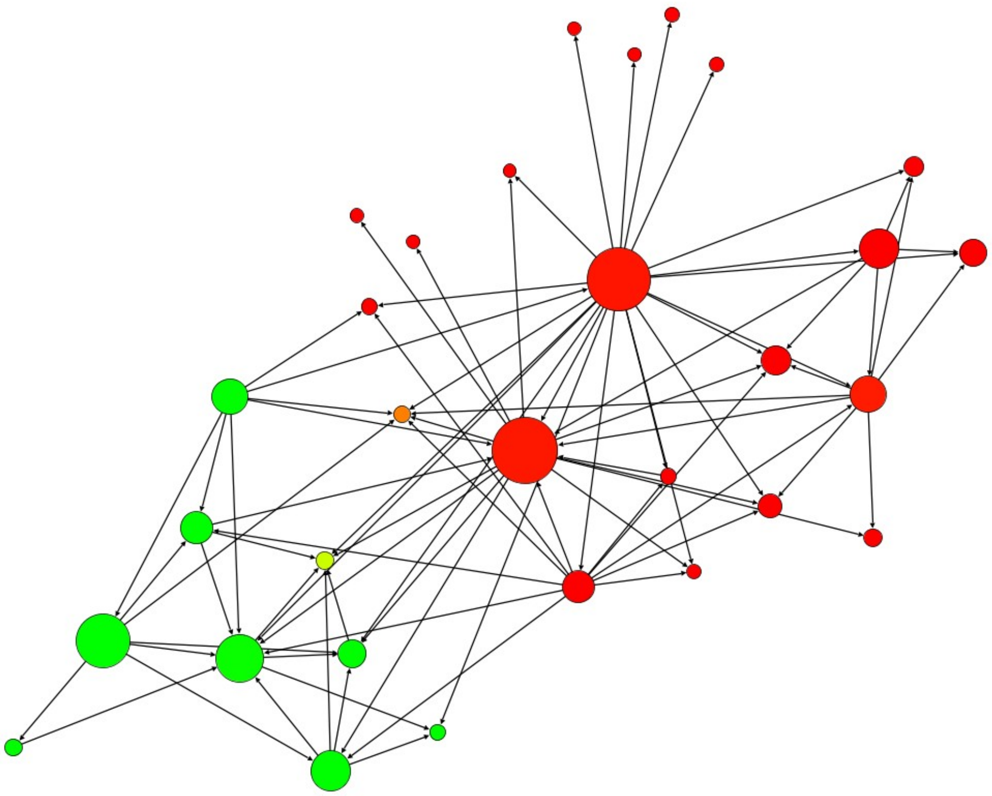
# Composition of user communities



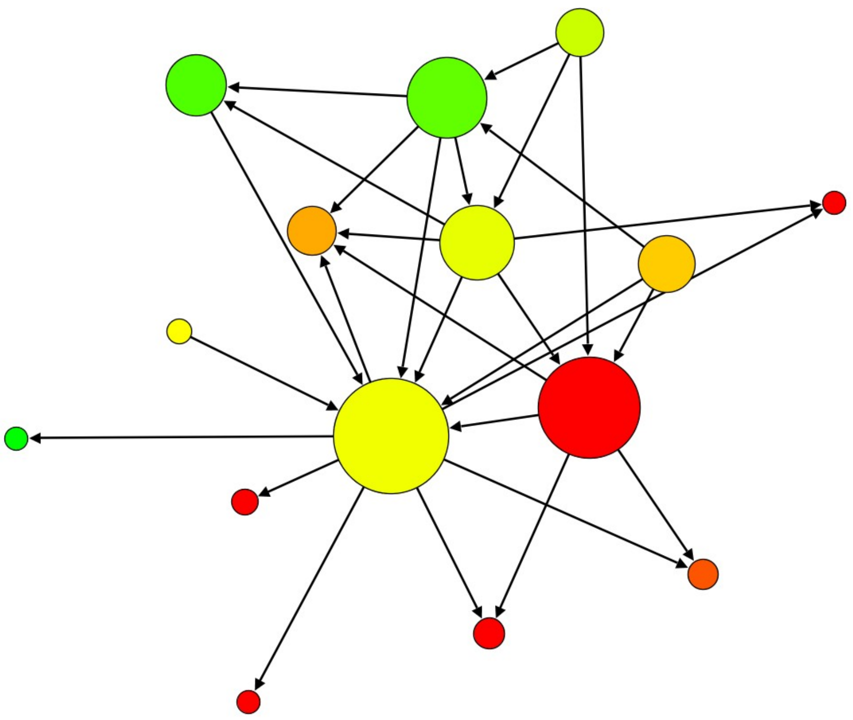

Apply community detection. For each group with >10 tagged users, calculate composition as frequency of activists. Back-calculate distribution of group compositions for each user. Compare to expected value for homogeneous distribution of user types.


Follower
Retweet
Mention
#climatechange
#globalwarming
#agw
100%
sceptic
50% / 50%
100%
activist

# “Echo-chambers” and “open forums”



#agw (retweets)



#agw (mentions)

0% Proportion of activist members 100%

# Does group composition matter?


Homogeneous groups → polarisation, fragmentation, filter bubbles, extreme views

Heterogeneous groups → information spreading, diversity of choice, moderate views

# Sentiment in mentions

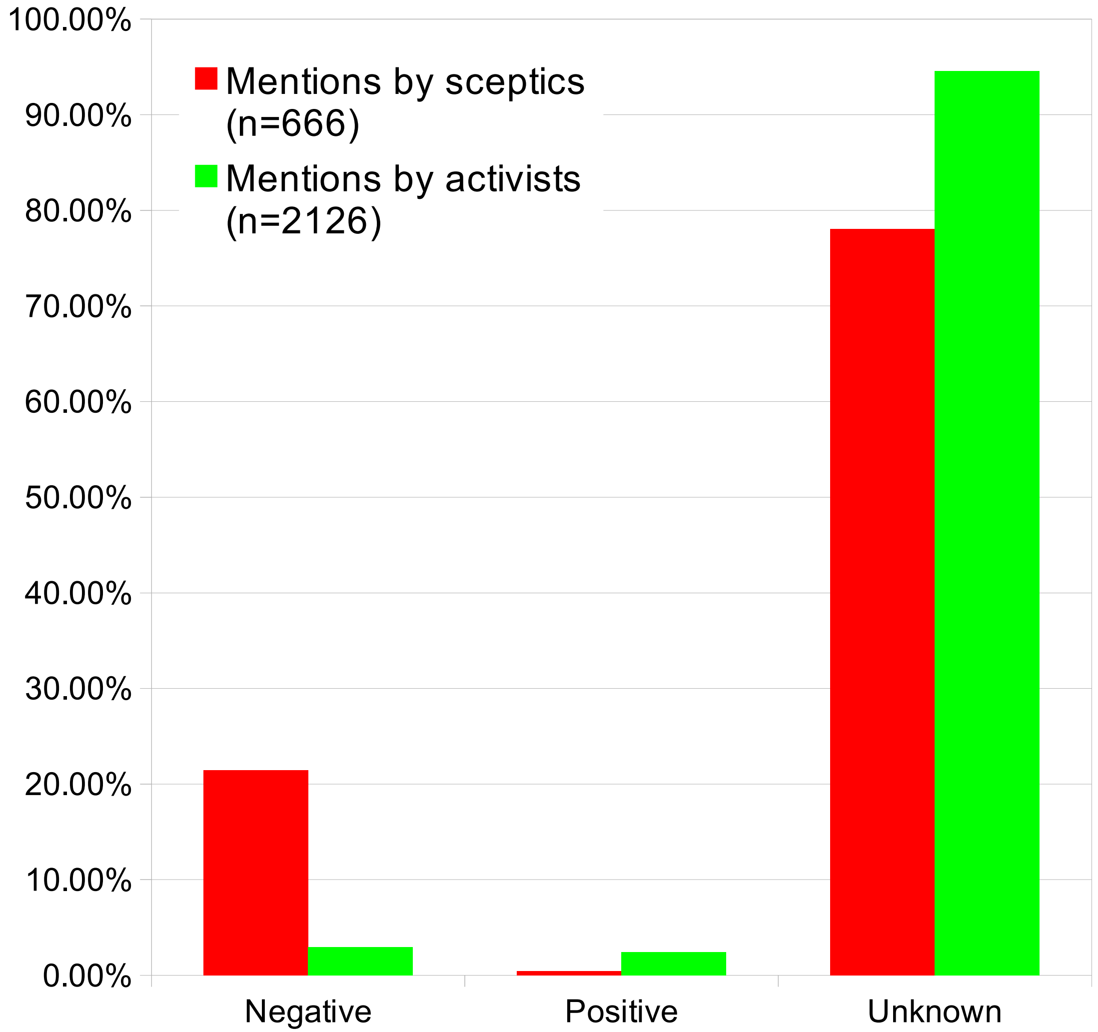
**Positive:** support, agreement, praise, confirmation

**Neutral**: neither positive nor negative

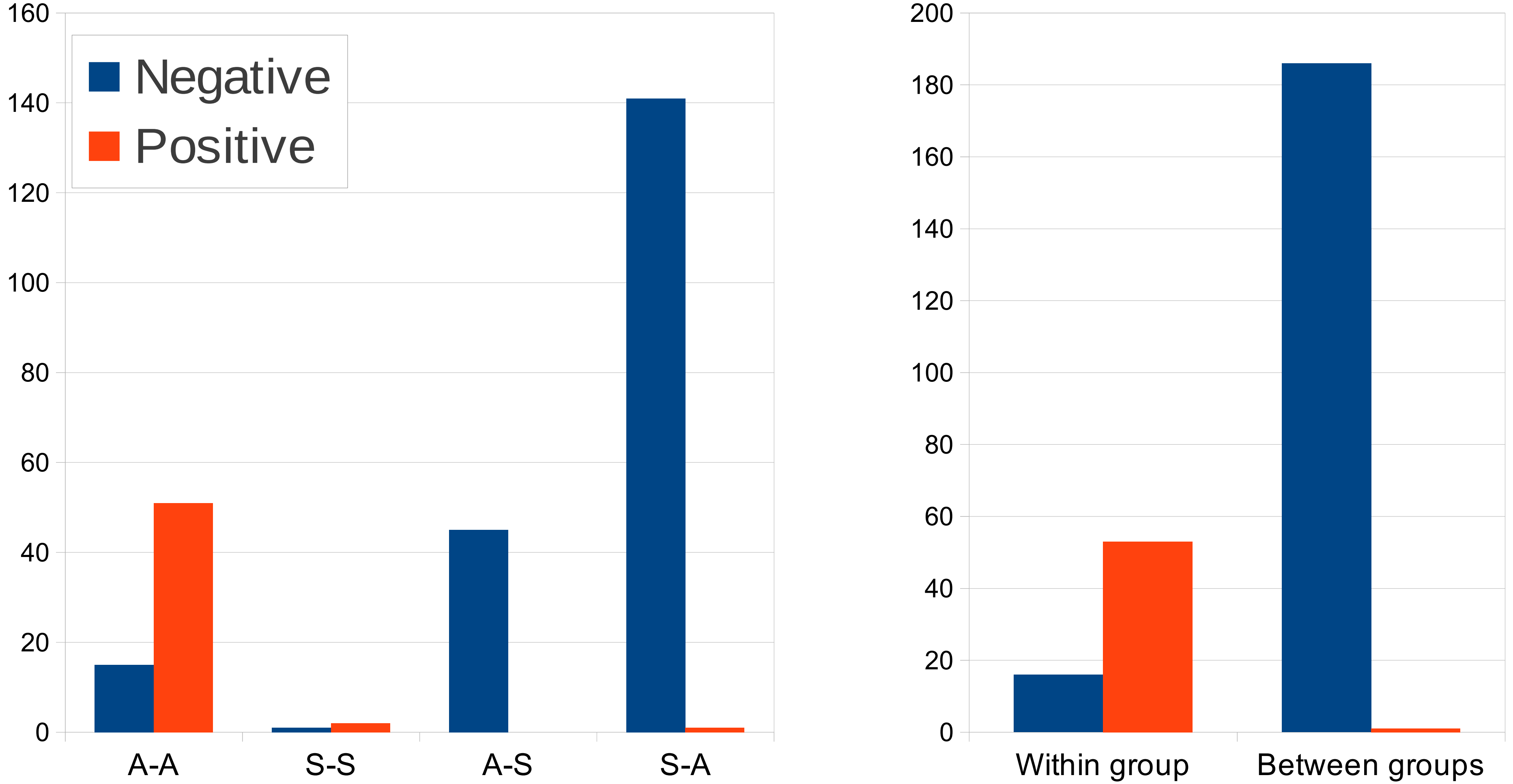
**Negative**: criticism, disagreement, abuse, undermining

**Unknown**: no sentiment could be distinguished

(Panel of 3 researchers, only unanimous decisions accepted.)



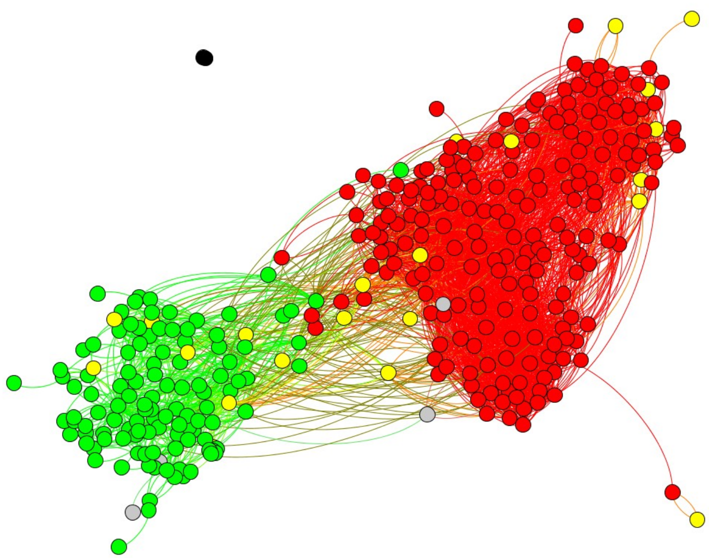
# Sentiment within and between groups

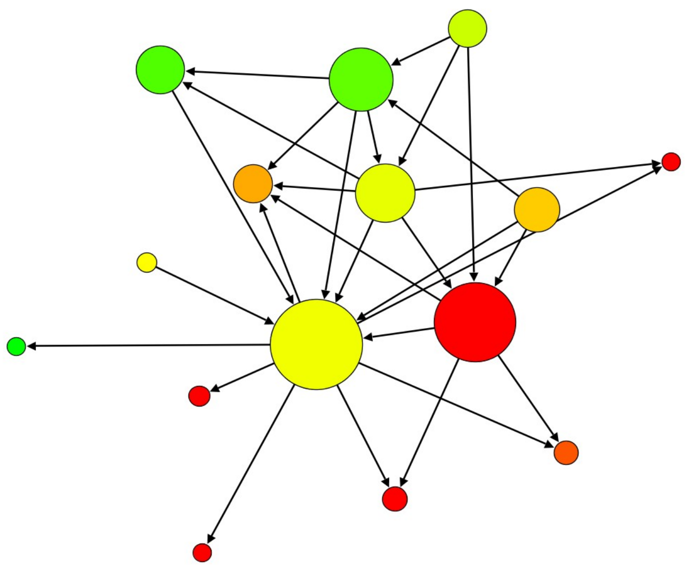



X-Y indicates mentions of group X by group Y.

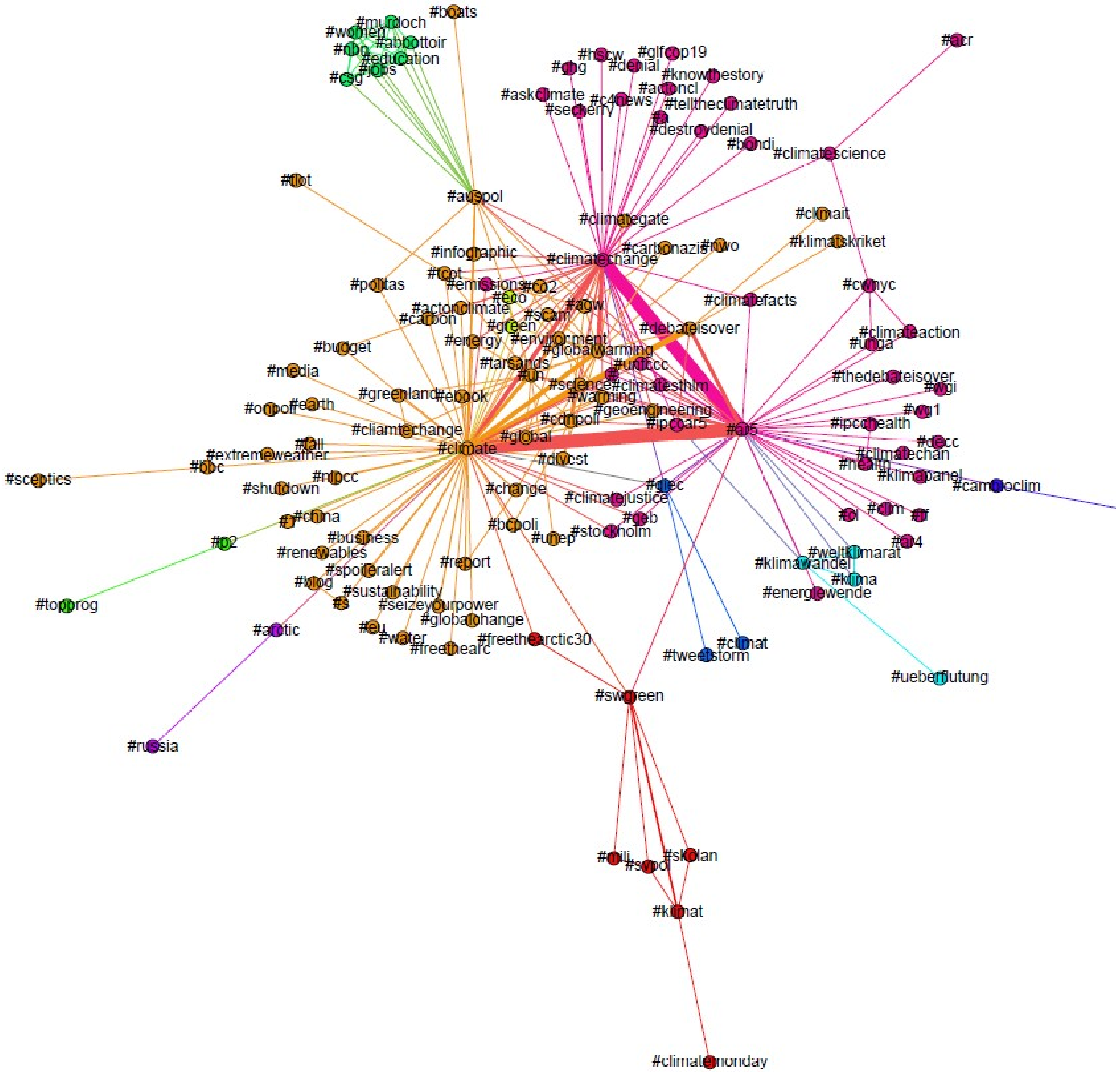
# Recap: Social network analysis

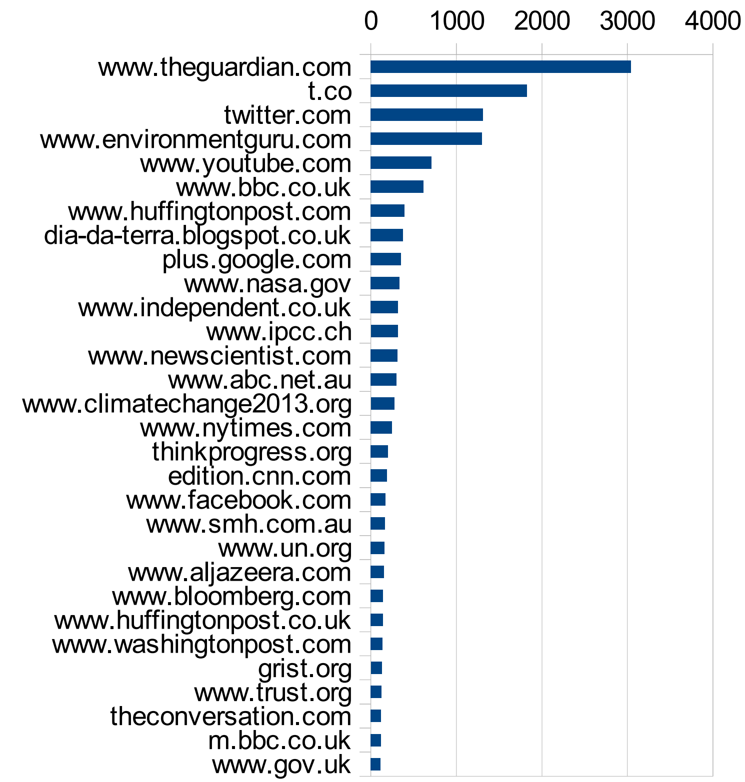
- Climate debate on Twitter is mostly characterised by homophily and polarised communities
- But some communities show diversity of viewpoints (albeit often with negative tone)





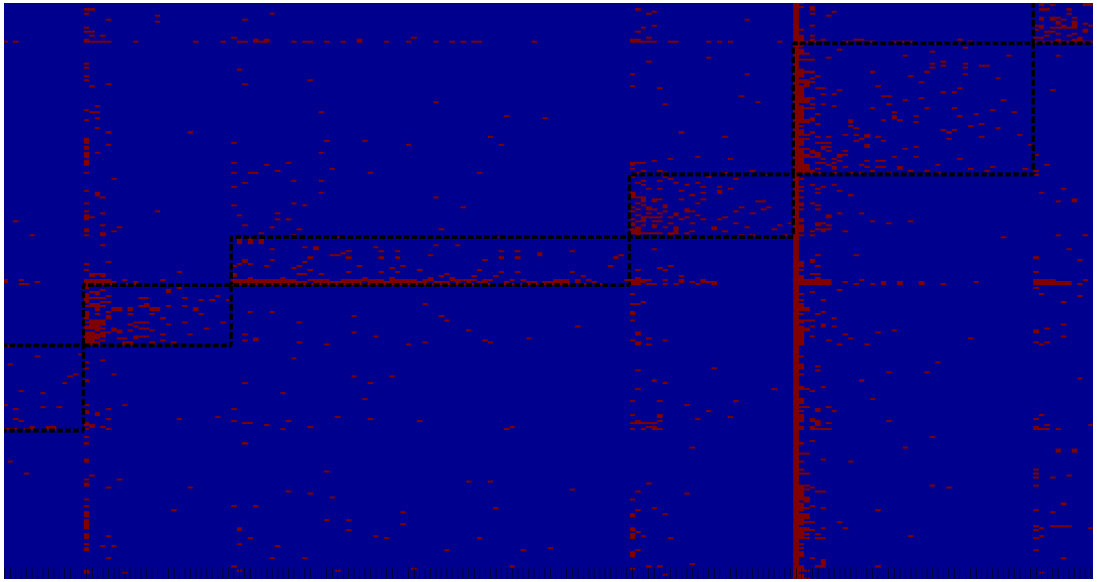
# Content analysis

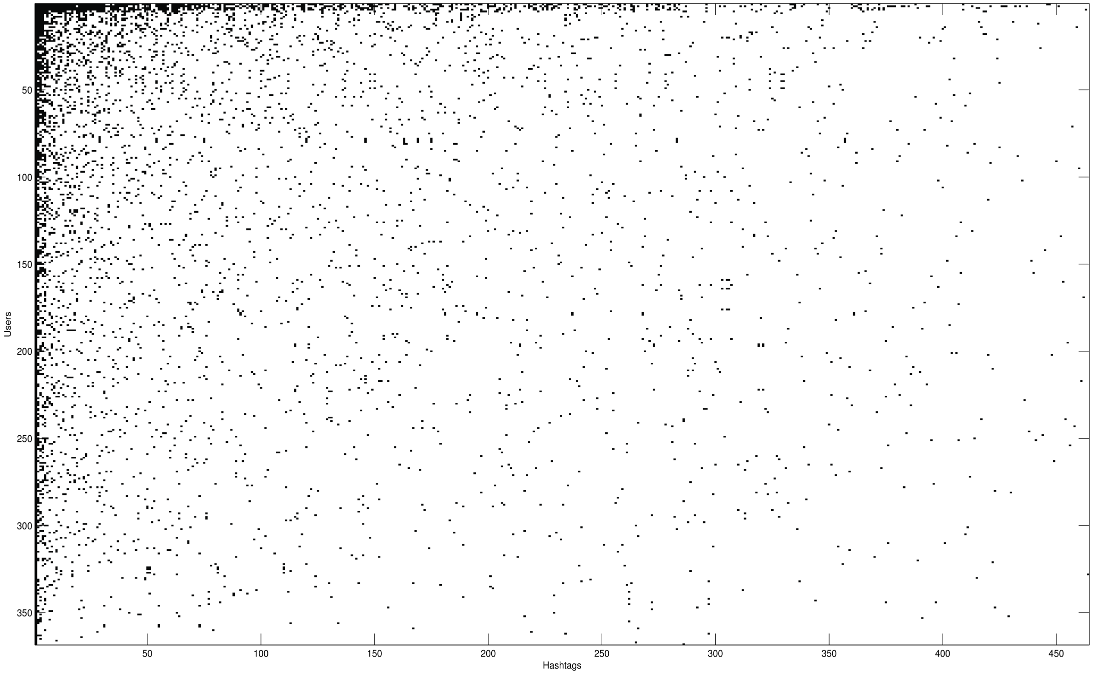


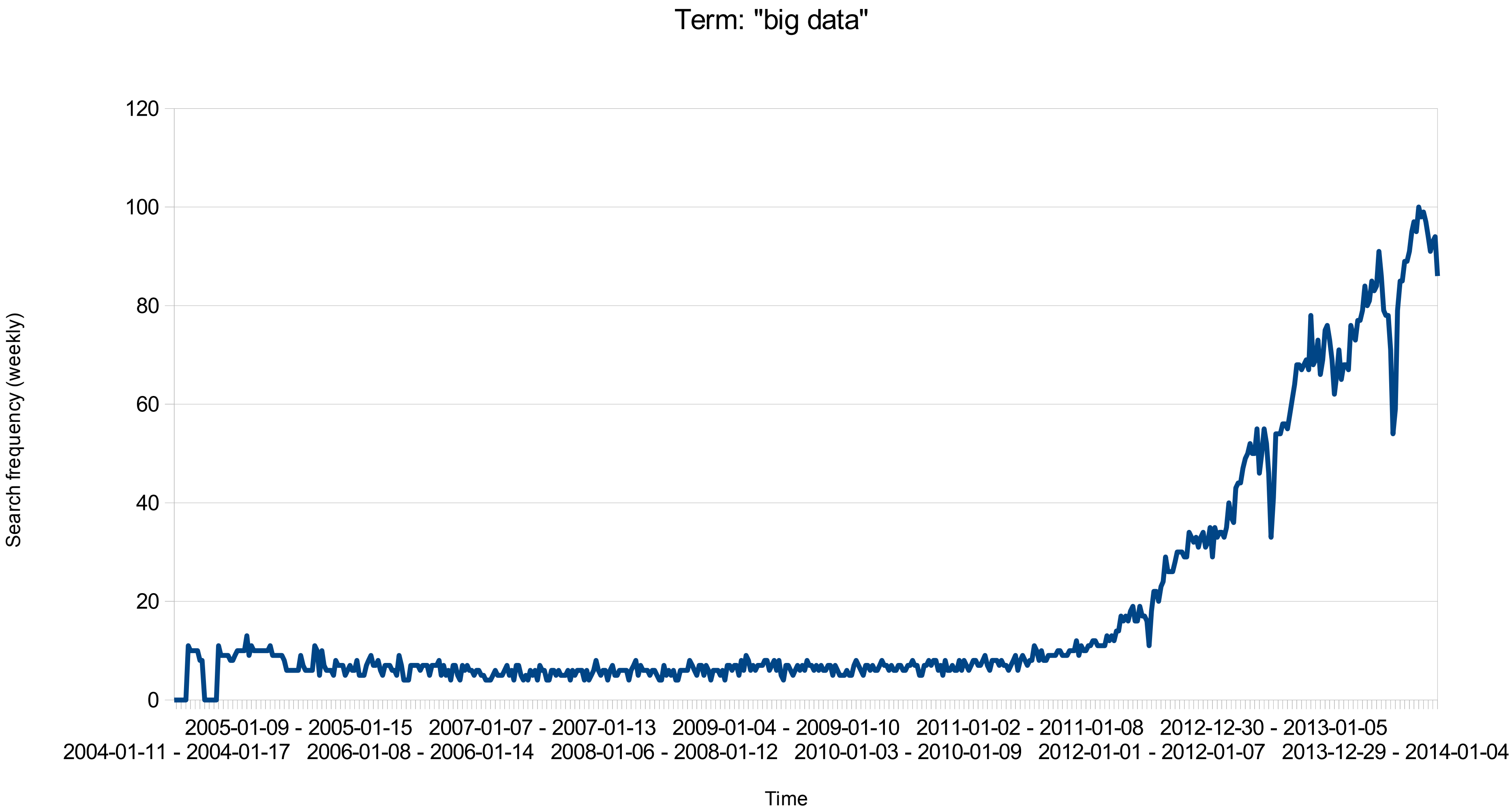

Hashtag associations and most-shared domains from tweets including #IPCC around AR5.

# Bipartite hashtag-user association graphs





# Big data analysis of “big data”




Source: Google Trends

# Big data claims

- Produces very accurate predictions
- All data can be captured (*n=all*) → no need for sampling
- Causality/models/theory not needed


# ...and some counter-claims

- Big data produces many false positives
- Predictions can fail when environment changes
- In reality *n=all* is rarely possible
- Sample volume does not compensate for sample bias
- Causality and mechanism may not be needed for engineering, but are essential for science


FIGURE 13-2A: COMPETING SIGNALS WITH ONE SIGNAL HIGHLIGHTED

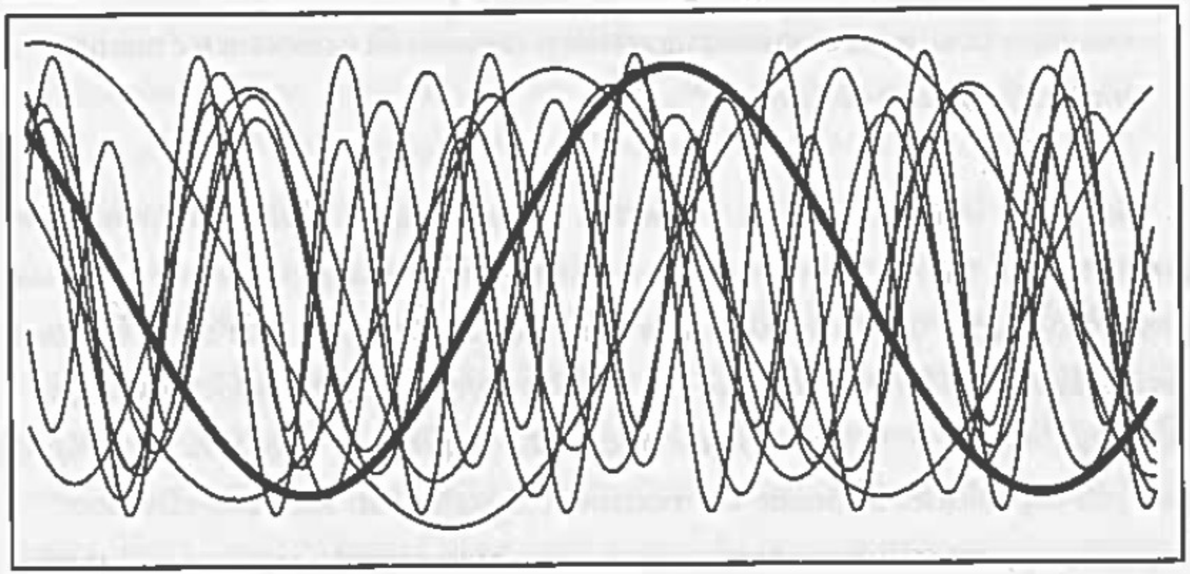
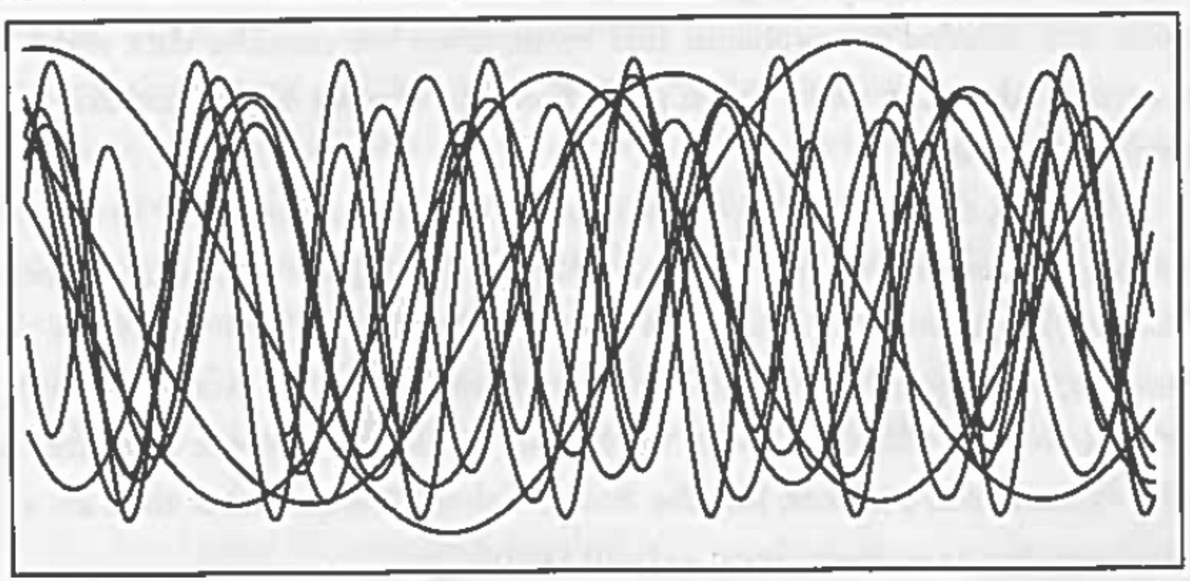


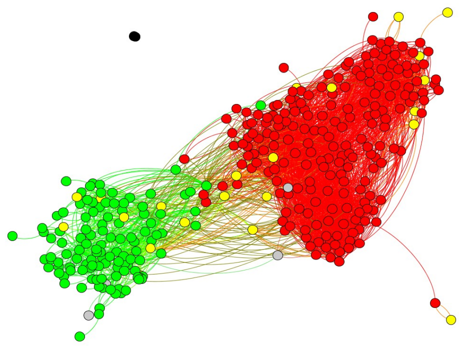
FIGURE 13-2B: COMPETING SIGNALS, UNDIFFERENTIATED



Big data makes theory more important, not less.

# Discussion

- Large and novel datasets are a by-product of the digital revolution
- Creates scientific opportunities – e.g. for understanding large-scale collective behaviour
- Example: Social network analysis of online climate change debate
- Next: Dynamics (networks and groups), other datasets (Tumblr, blogs), cross-media analysis (multiplex networks), machine learning (user categorisation).



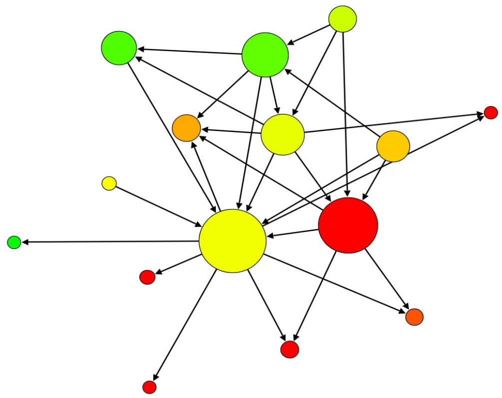
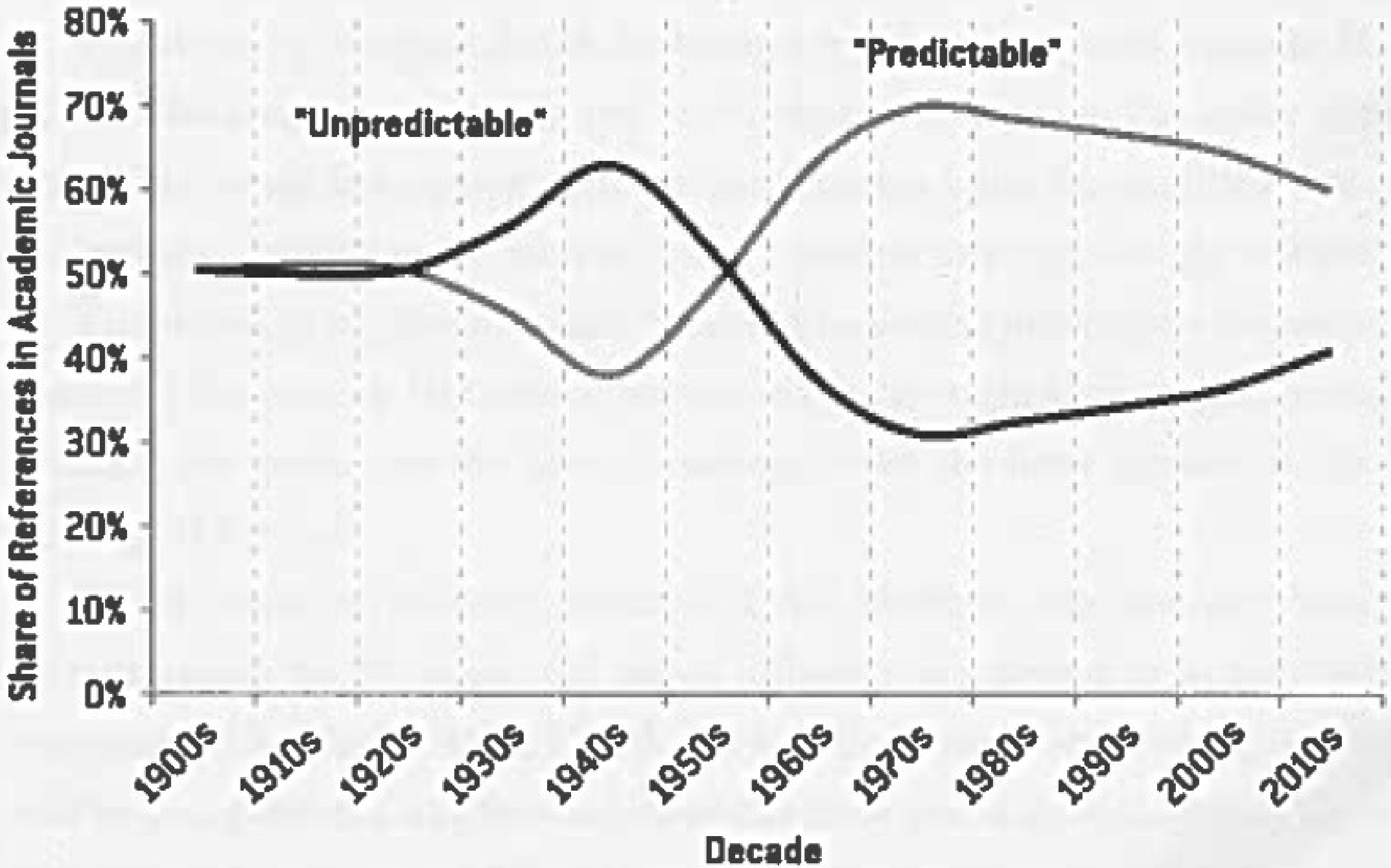


FIGURE C-2: THE PERCEPTION OF PREDICTABILITY, 1900–2012



# Quantifying homophily

- Homophily: prevalence of same-same interactions
- Bootstrap method:
  - Create ensemble of 10000 “random” networks
    - Preserve total number of edges
    - Preserve number of users in each class
    - Preserve in/out degree distributions of each user class
  - Compare observed network statistics to expectation from bootstrap ensemble
- Calculate z-score and p-value (null distribution is normal)

$$z = \frac{obs - mn}{sd}$$

# Modularity

Modularity, Q, is a quality function for a network partition:

$$Q = \sum_{ij} \left[ \frac{A_{ij}}{2m} - \frac{k_i k_j}{(2m)^2} \right] \delta(c_i, c_j)$$

where:

- $A_{ij}$=1 if there is an edge connecting nodes i and j, $A_{ij}$=0 otherwise
- $k_i$ is the degree of node i
- m is the total number of edges
- δ is the Kronecker delta (returns 1 if arguments are equal, 0 otherwise)
- $c_i$ is the assigned community for node i

Modularity is high when edges fall within (and not between) communities.