INTRODUCTION TO SOCIO-SEMANTIC NETWORKS

Camille Roth
CNRS

Centre Marc Bloch Berlin e.V.
(BMBF / CNRS / Humboldt Universität / MAE)

ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

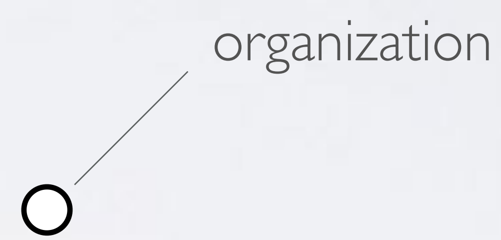
- **"actors"**
    -

ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**

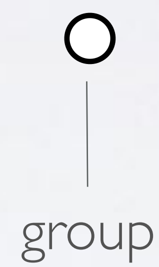
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**



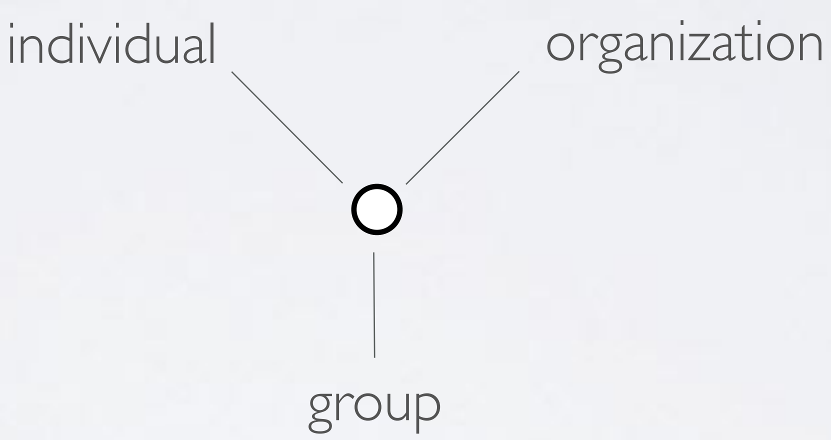
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**



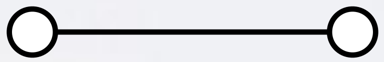
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**



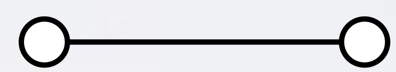
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**
  - inter-actor relationships



ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**
  - inter-actor relationships



"interaction"

'discussion'

'friendship'

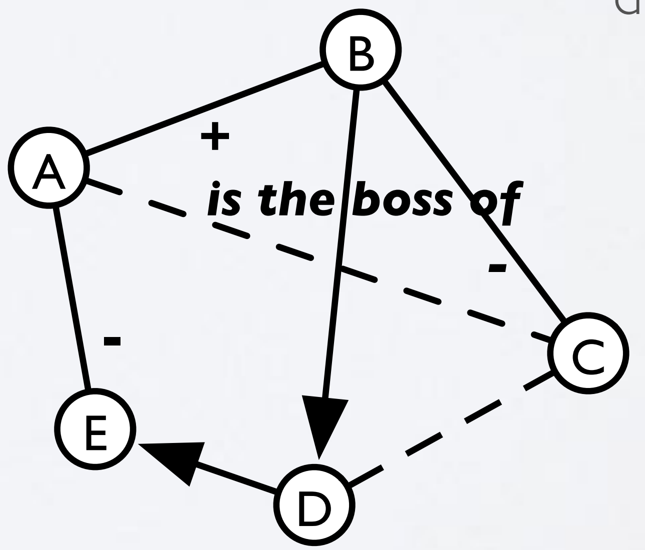
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**
    - inter-actor relationships

"interaction"

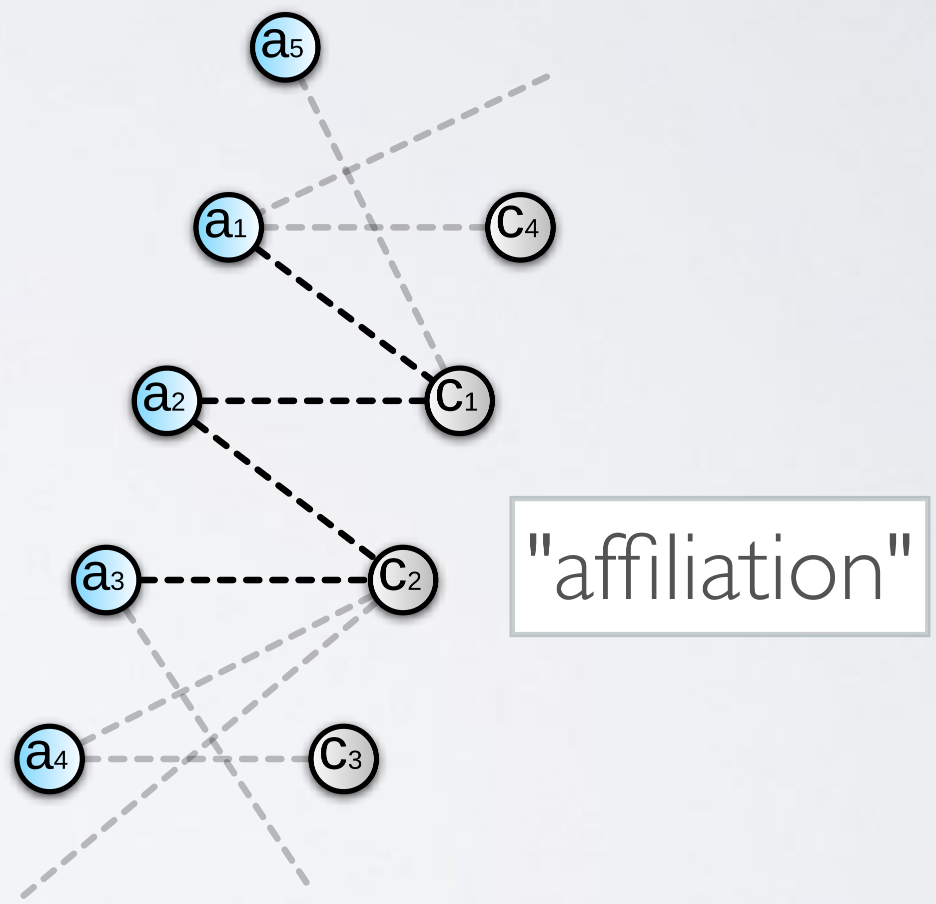
'discussion'

'friendship'



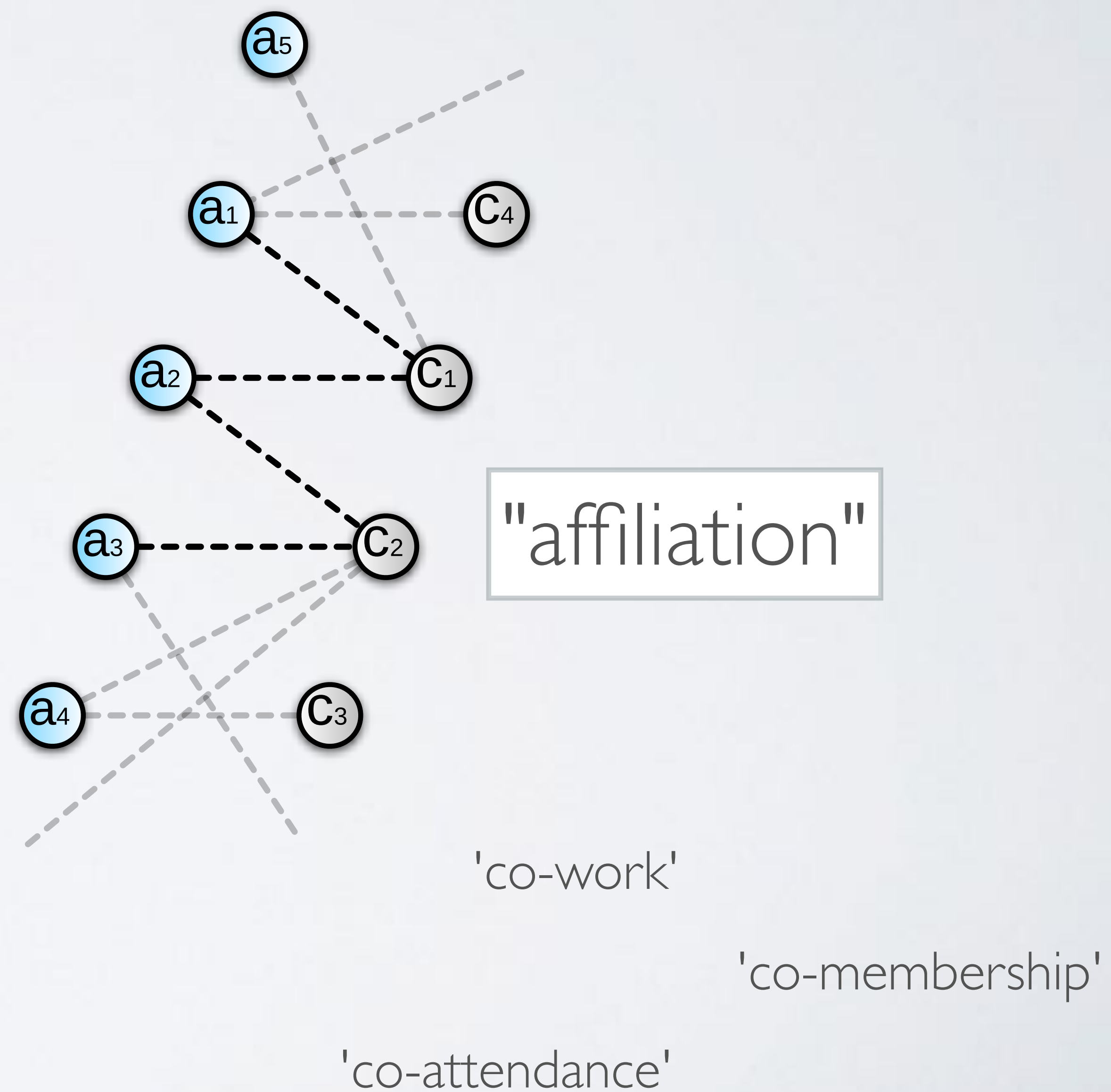
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**
  - inter-actor relationships
  - joint actor affiliations



ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **"actors"**
  - inter-actor relationships
  - joint actor affiliations



ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

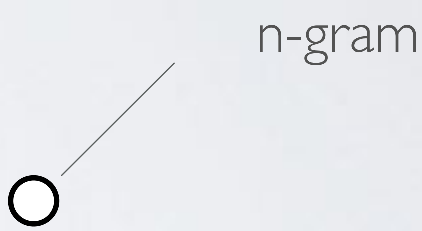
- **on the "semantic" side...**

ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **on the "semantic" side...**

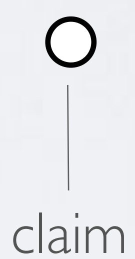
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **on the "semantic" side...**



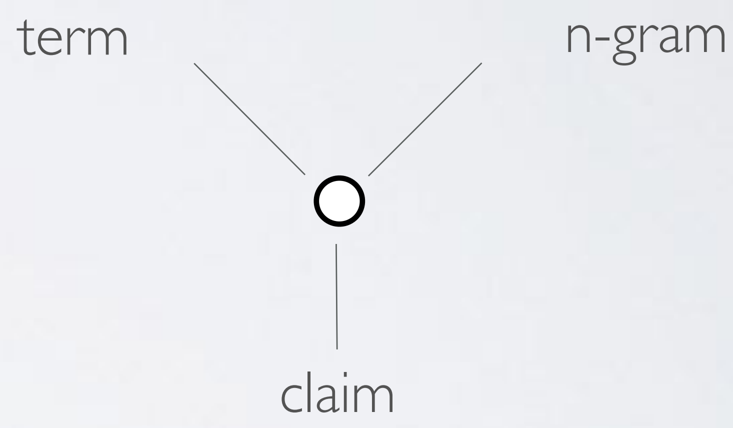
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **on the "semantic" side...**



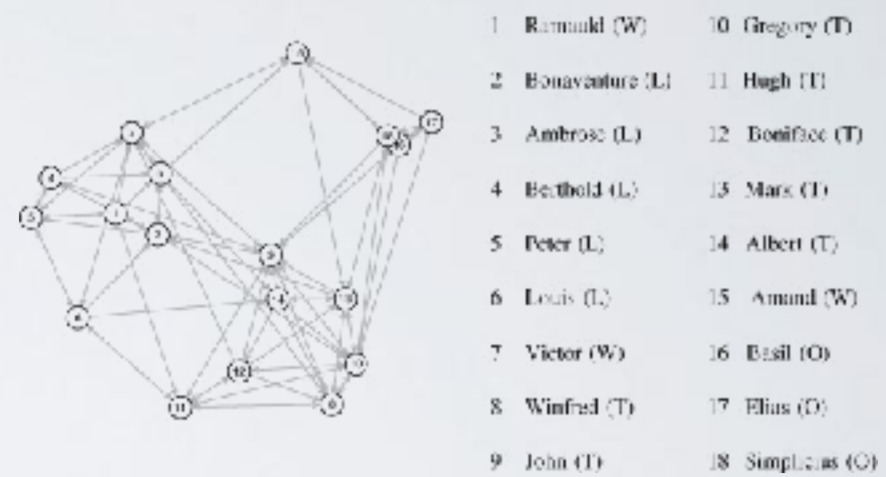
ONTOLOGY ISSUES: "SOCIO"-"SEMANTIC" ?

- **on the "semantic" side...**



SOCIAL NETWORK DYNAMICS

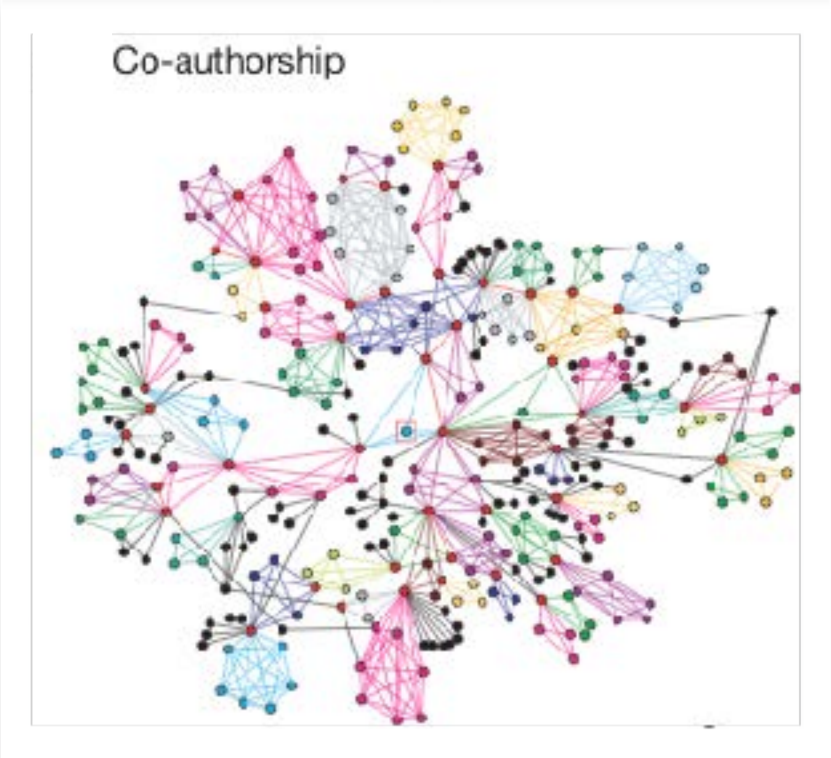
(Sampson, 1968)



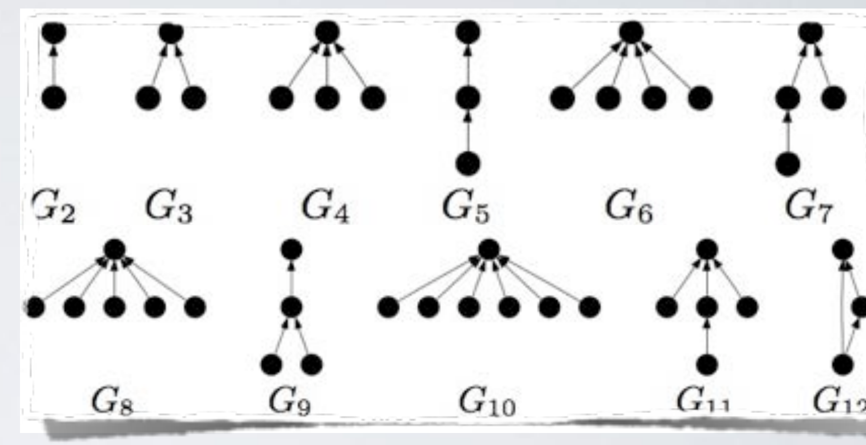
- **First period of development: 40s-70s**
  - social science, mathematical sociology
  - focused on "small" case-studies, algebraic definitions
- **Second period: the 'new science of networks', end of 90s-now**
  - large-scale datasets, complex systems standpoint
  - classical stylized facts: "power laws", communities, ...

SOCIAL NETWORK DYNAMICS

- **First period of development: 40s-70s**
  - social science, mathematical sociology
  - focused on "small" case-studies, algebraic definitions
- **Second period: the 'new science of networks', end of 90s-now**
  - large-scale datasets, complex systems standpoint
  - classical stylized facts: "power laws", communities, ...



SOCIAL NETWORK DYNAMICS



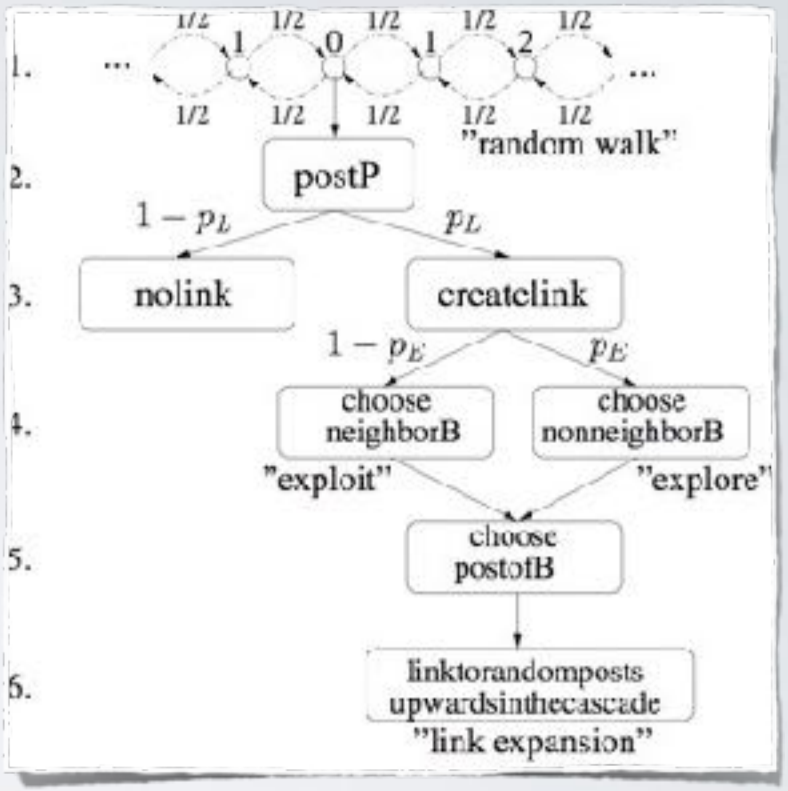
(McGlohon, Leskovec, Faloutsos, Hurst, Glance, 2007)

- **First period of development: 40s-70s**
  - social science, mathematical sociology
  - focused on "small" case-studies, algebraic definitions
- **Second period: the 'new science of networks', end of 90s-now**
  - large-scale datasets, complex systems standpoint
  - classical stylized facts: "power laws", communities, ...

SOCIAL NETWORK DYNAMICS

- **First period of development: 40s-70s**
  - social science, mathematical sociology
  - focused on "small" case-studies, algebraic definitions
- **Second period: the 'new science of networks', end of 90s-now**
  - large-scale datasets, complex systems standpoint
  - classical stylized facts: "power laws", communities, ...

(Gotz, Leskovec, McGlohon, Faloutsos, 2009)



SOCIAL NETWORK MODELS

reconstructing / using	dynamics	structure
dynamics	Preferential linking Scoring methods	Rewiring models Cost function optimization Agent-based models
structure	ERGMs, SAOMs	Prescribed structure Subgraph-based

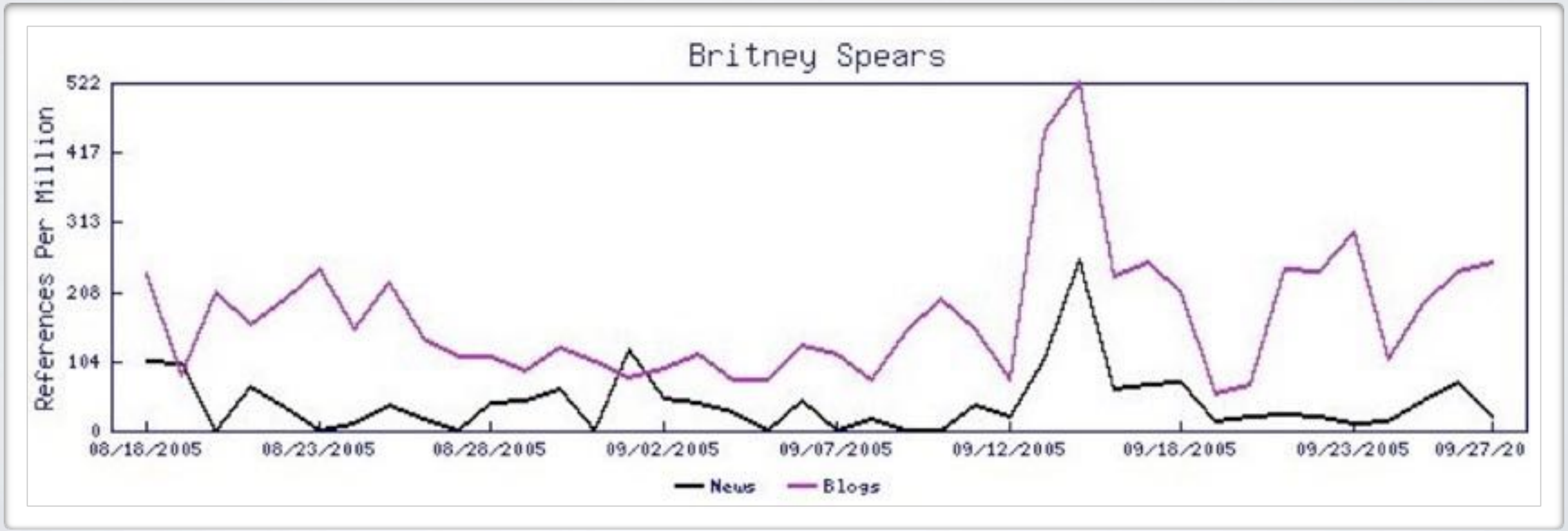
Symbolic regression

CONTENT DYNAMICS

Dynamics of term usage

- vs. source type
- vs. location
- predictive

(Lloyd, Kaulgud, Skiena, 2005)



CONTENT DYNAMICS

Dynamics of term usage

- vs. source type
- vs. location
- predictive

(Lloyd, Kaulgud, Skiena, 2005)

CONTENT DYNAMICS

Dynamics of term usage

- vs. source type
- vs. location
- predictive

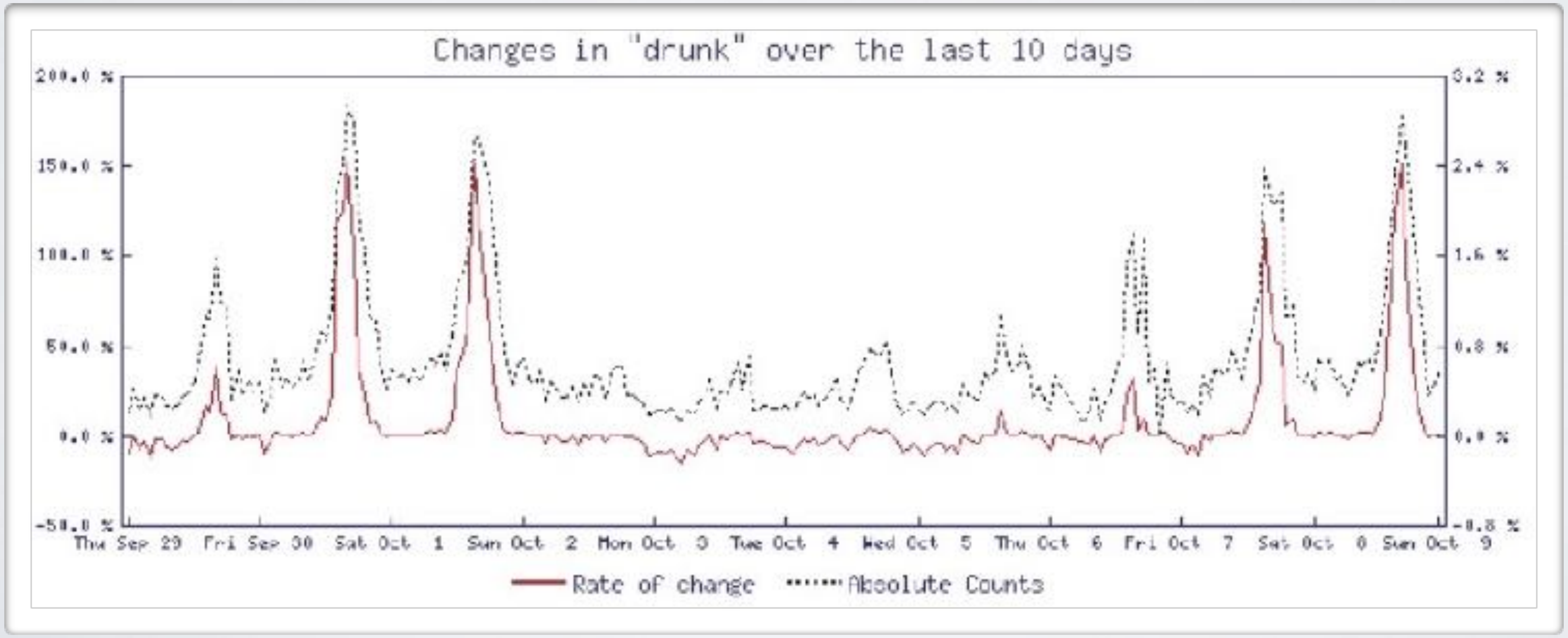
(Balog et al., 2004; Mishne et al., 2006)

CONTENT DYNAMICS

Dynamics of term usage

- vs. source type
- vs. location
- predictive

(Balog et al., 2004; Mishne et al., 2006)

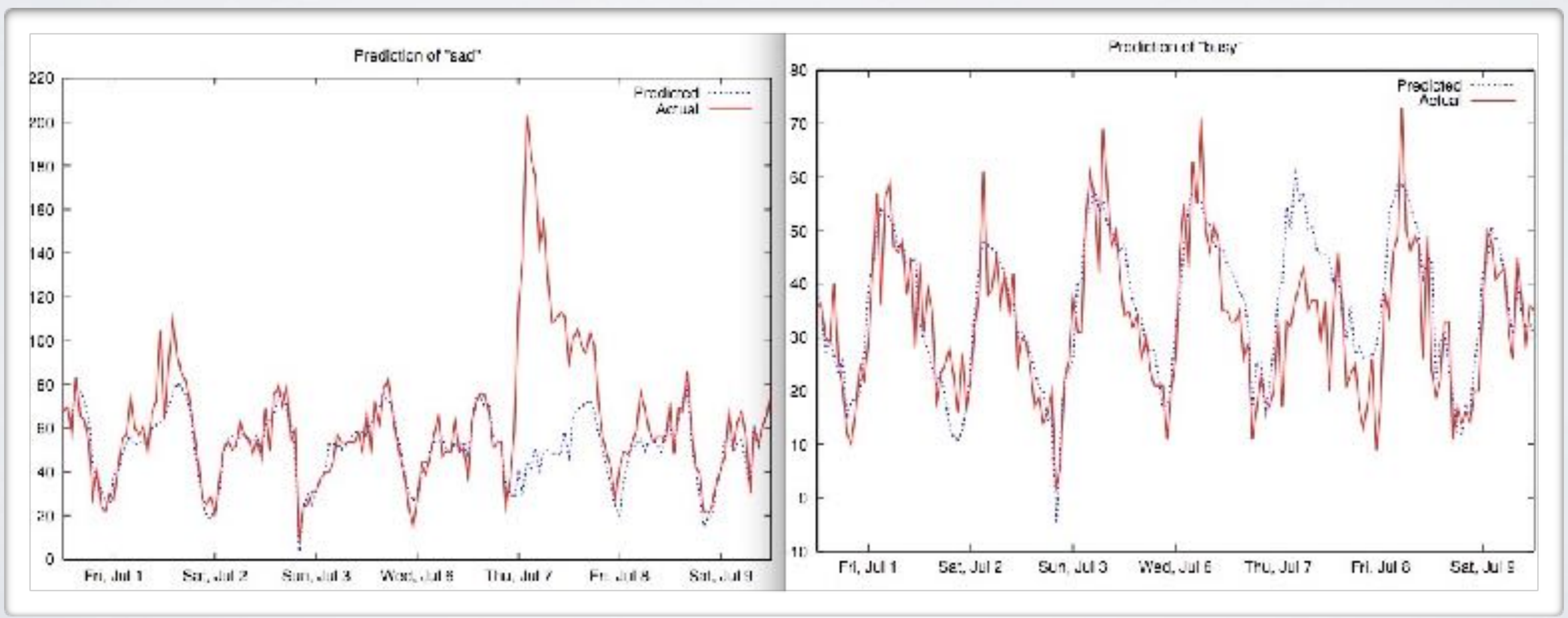


CONTENT DYNAMICS

Dynamics of term usage

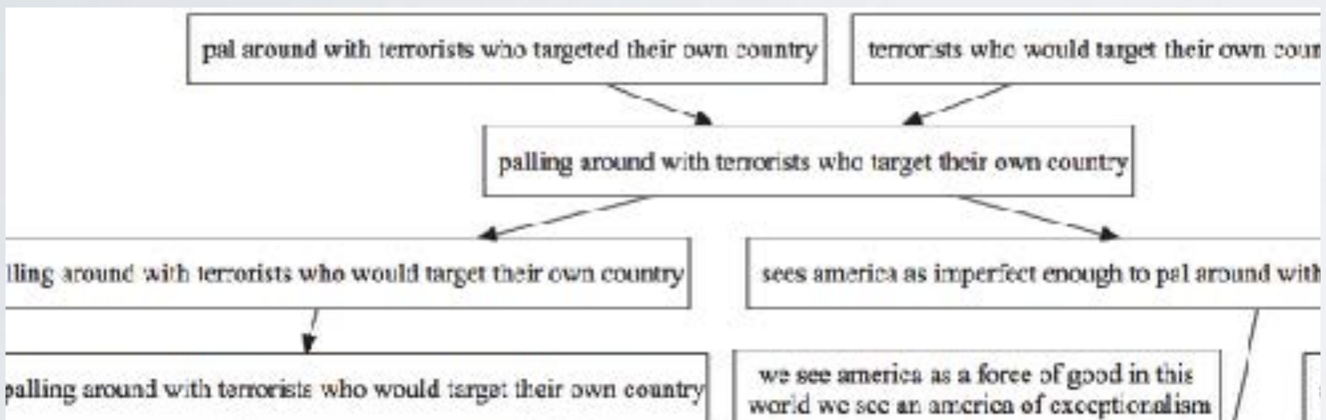
- vs. source type
- vs. location
- predictive

(Balog et al., 2004; Mishne et al., 2006)

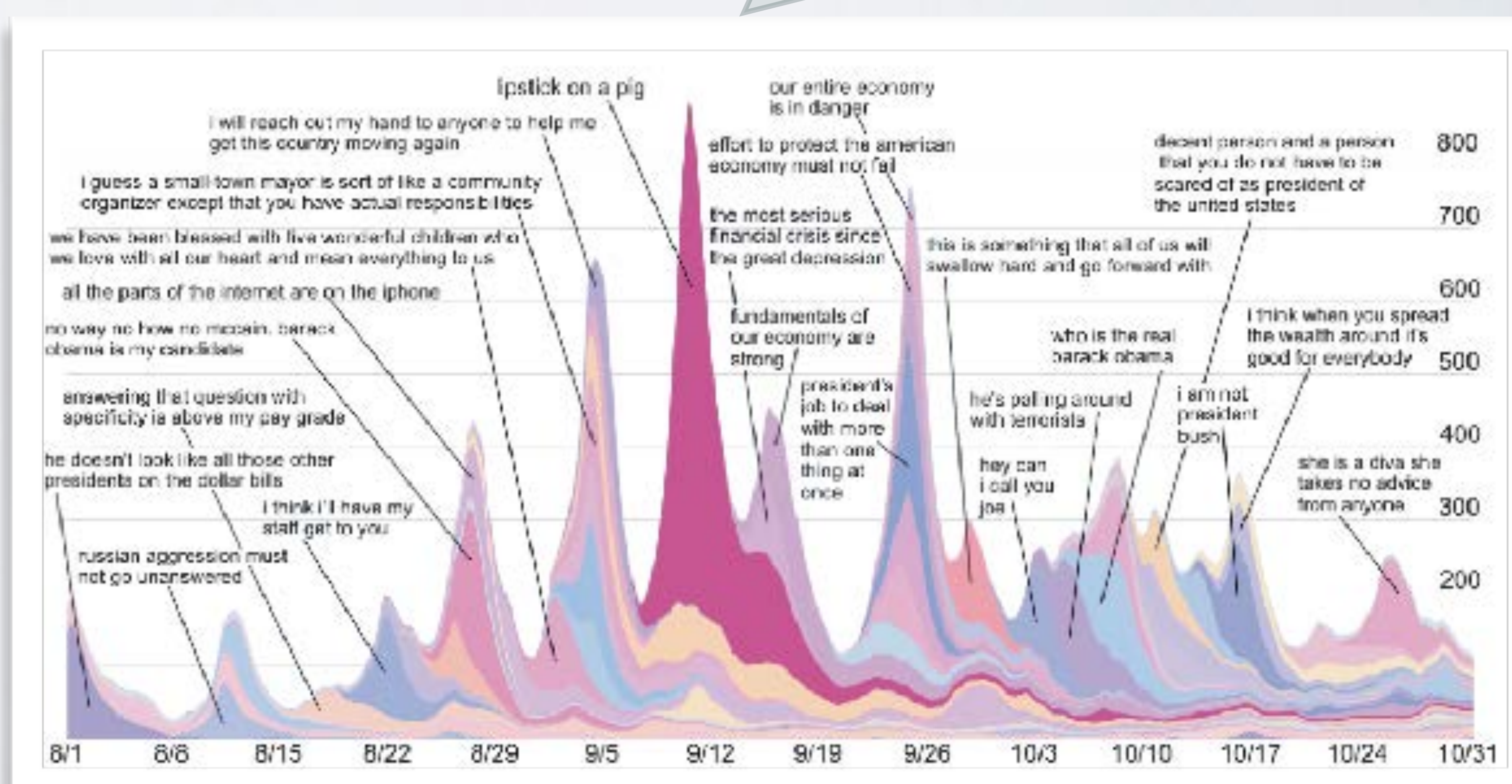


CONTENT DYNAMICS

e.g., dynamics of sentences / quotations, called "*memes*"



(Leskovec, Backstrom, Kleinberg, 2009)



CONTENT DYNAMICS

Observing conceptual mutation

(Simmons, Adamic, Adar, 2011)

abn.com

"i find that governor sarah palin abused her power by violating alaska statute 39 52 110 a of the alaska executive branch ethics act"

demconwatchdog.com

"i find that governor sarah palin abused her power by violating alaska statute 39"

lzydata.blogspot.com

“abused her power"

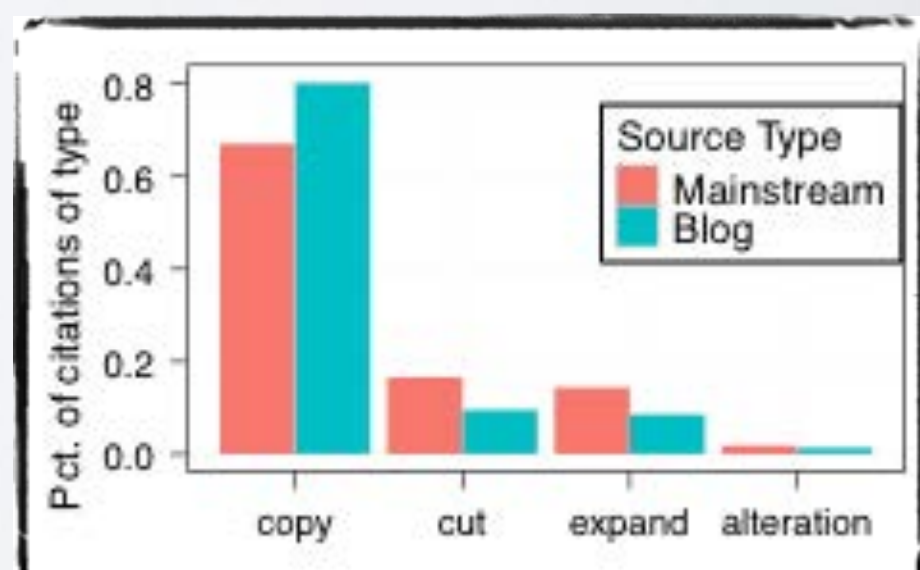
thenation.com

"i find that governor palin abused her power by violating alaska statute 39 52 110 a of the alaska executive branch ethics act *alaska statute 39 52 110 a provides ‘the legislature affirms that each public officer holds office as a public trust…*"

blogs.abcnews.com

"i find that governor sarah palin abused her power by violating alaska statute 39 52 110 a of the alaska *executive branch ethics act compliance with the code of ethics is not optional*"

Distinguishing between various node types



OPEN ACCESS Freely available online

PLOS | ONE

The Geography of Happiness: Connecting Twitter Sentiment and Expression, Demographics, and Objective Characteristics of Place

Lewis Mitchell[1]*, Morgan R. Frank[1], Kameron Decker Harris[1,2], Peter Sheridan Dodds[1], Christopher M. Danforth[1]

2013

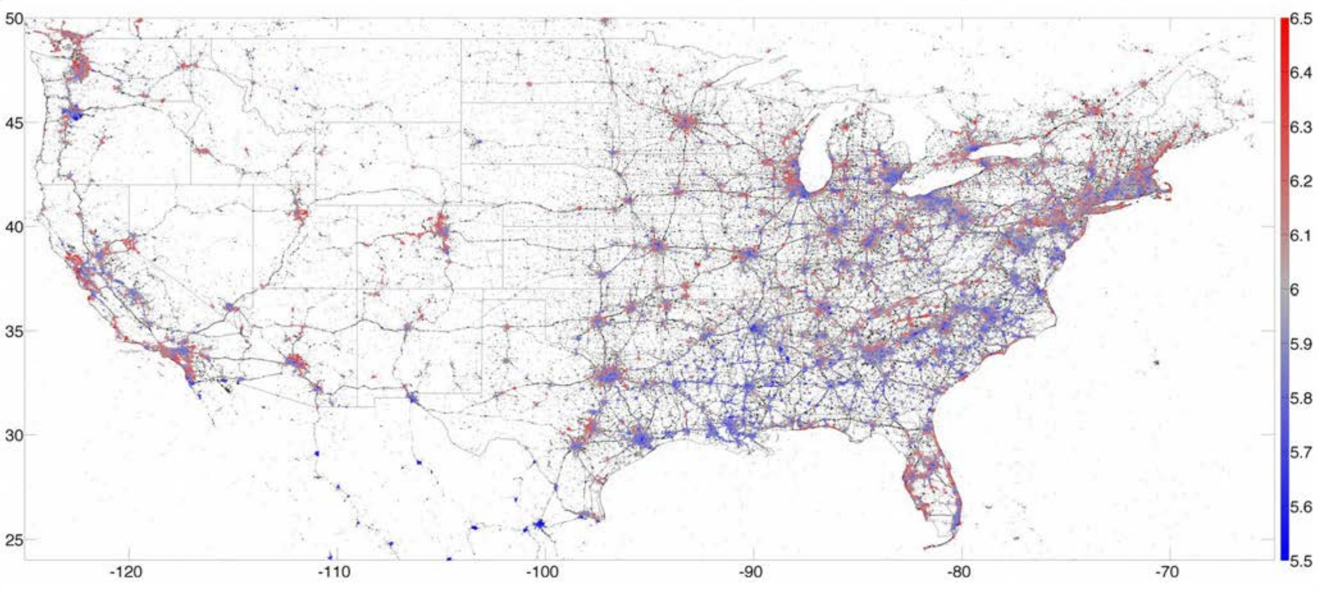


Figure 5. Map showing happiness of all tweets collected from the lower 48 US states during 2011. Points are colored as in figure 4,

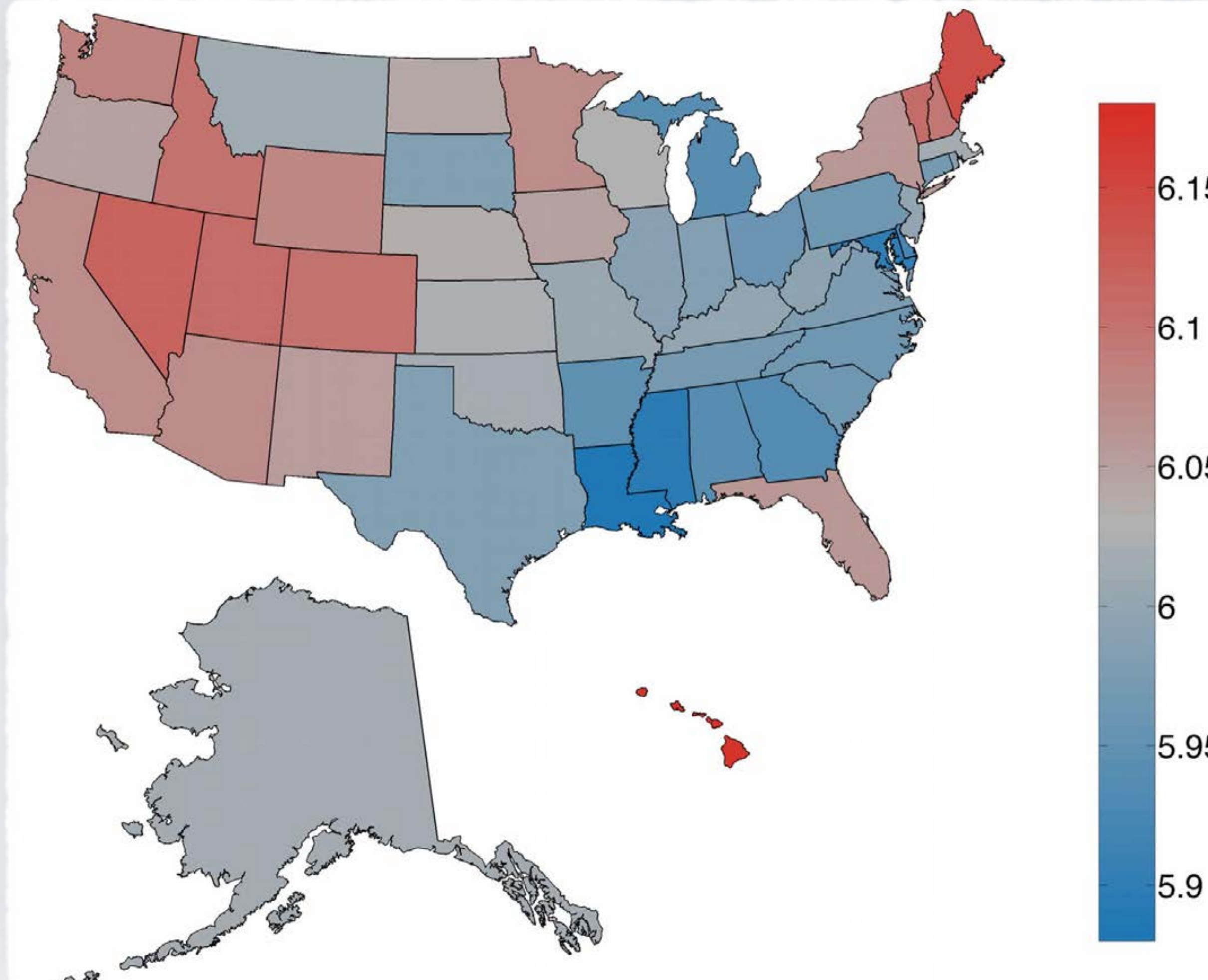
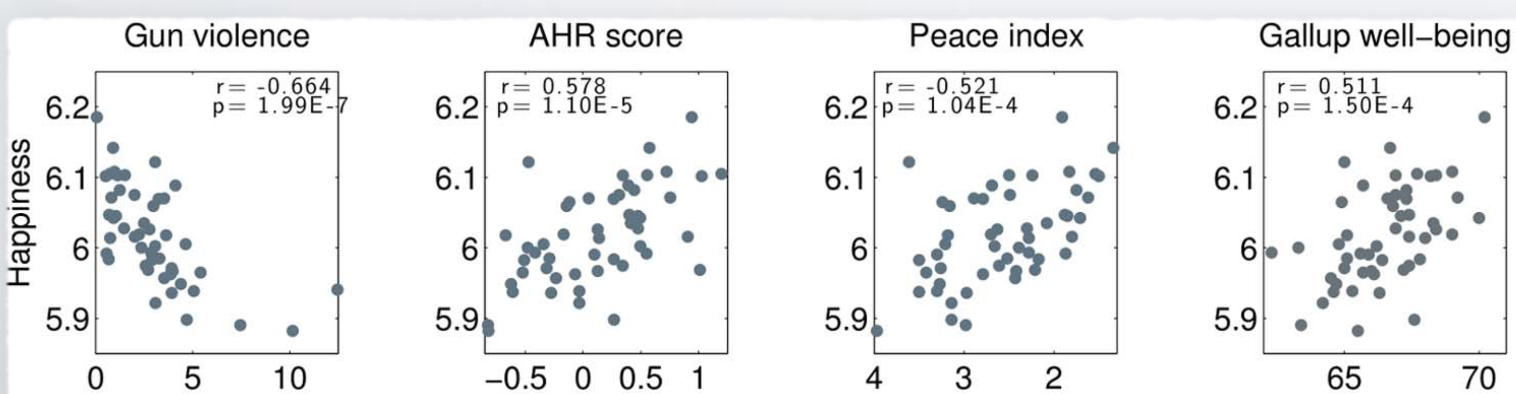


Figure 1. Average word happiness for geotagged tweets in all US states collected during calendar year 2011. The happiest 5 states, in order, are: Hawaii, Maine, Nevada, Utah and Vermont. The saddest 5 states, in order, are: Louisiana, Mississippi, Maryland, Delaware and Georgia. Word shift plots describing how differences in word usage contribute to variation in happiness between states are presented in Appendix B in Appendix S1 (online) [19].

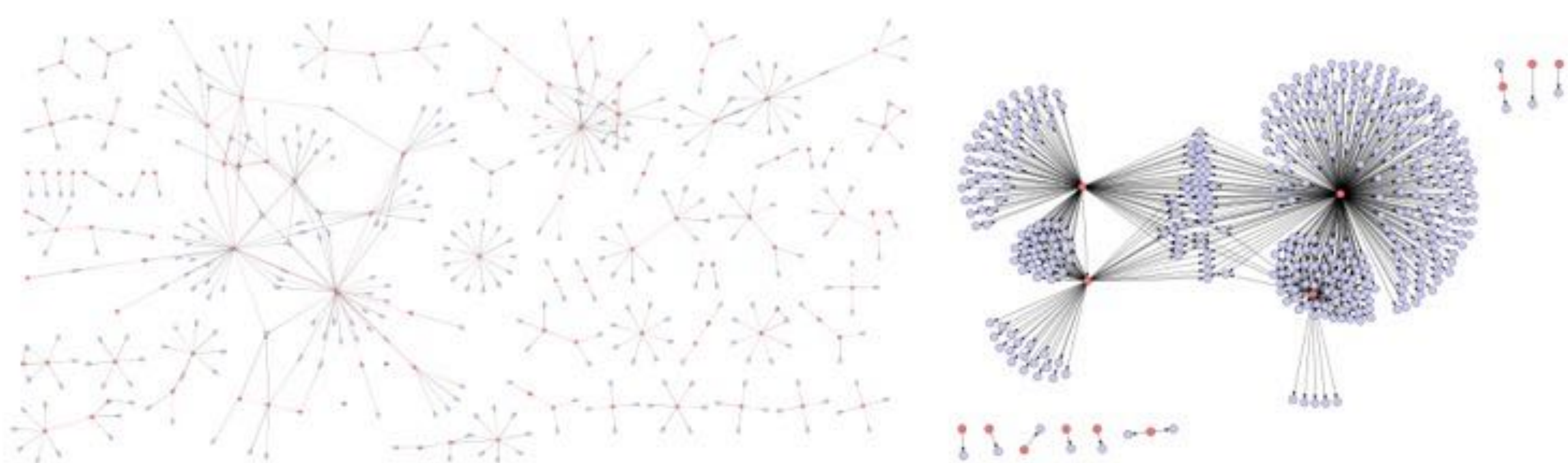


"Louisiana is revealed as the saddest state, with a significant factor being an abundance of profanity relative to the other states. This is in contrast with the findings of Oswald and Wu who determined Louisiana to be the state with highest well-being according to an alternate survey-based measure of life satisfaction"

CONTENT DYNAMICS AND DIFFUSION

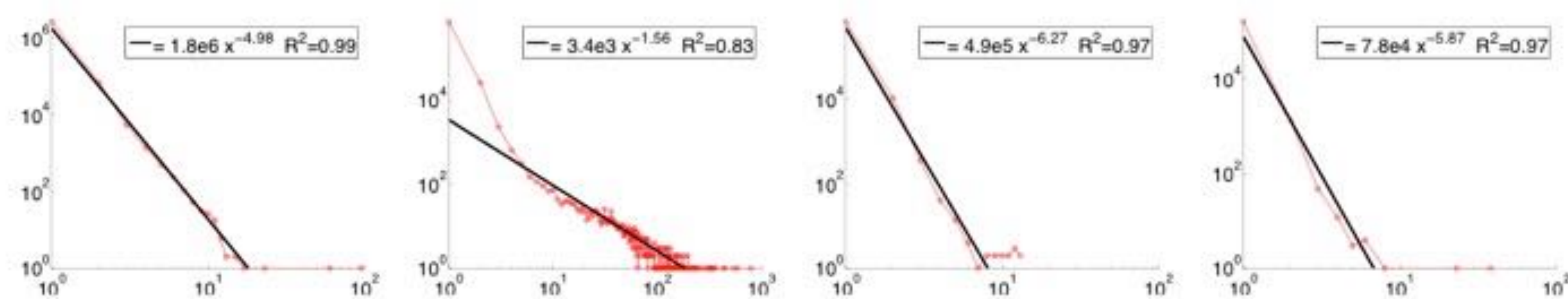
sometimes content type matters...

(Leskovec, Adamic, Huberman, 2006)



(a) Medical book (b) Japanese graphic novel

Figure 1: Examples of two product recommendation networks: (a) First aid study guide *First Aid for the USMLE Step*, (b) Japanese graphic novel (manga) *Oh My Goddess!: Mara Strikes Back*.



(a) Book (b) DVD (c) Music (d) Video

Figure 2: Size distribution of cascades (size of cascade vs. count). Bold line presents a power-fit.

The dynamics of viral marketing

CONTENT DYNAMICS AND DIFFUSION

sometimes content type matters...

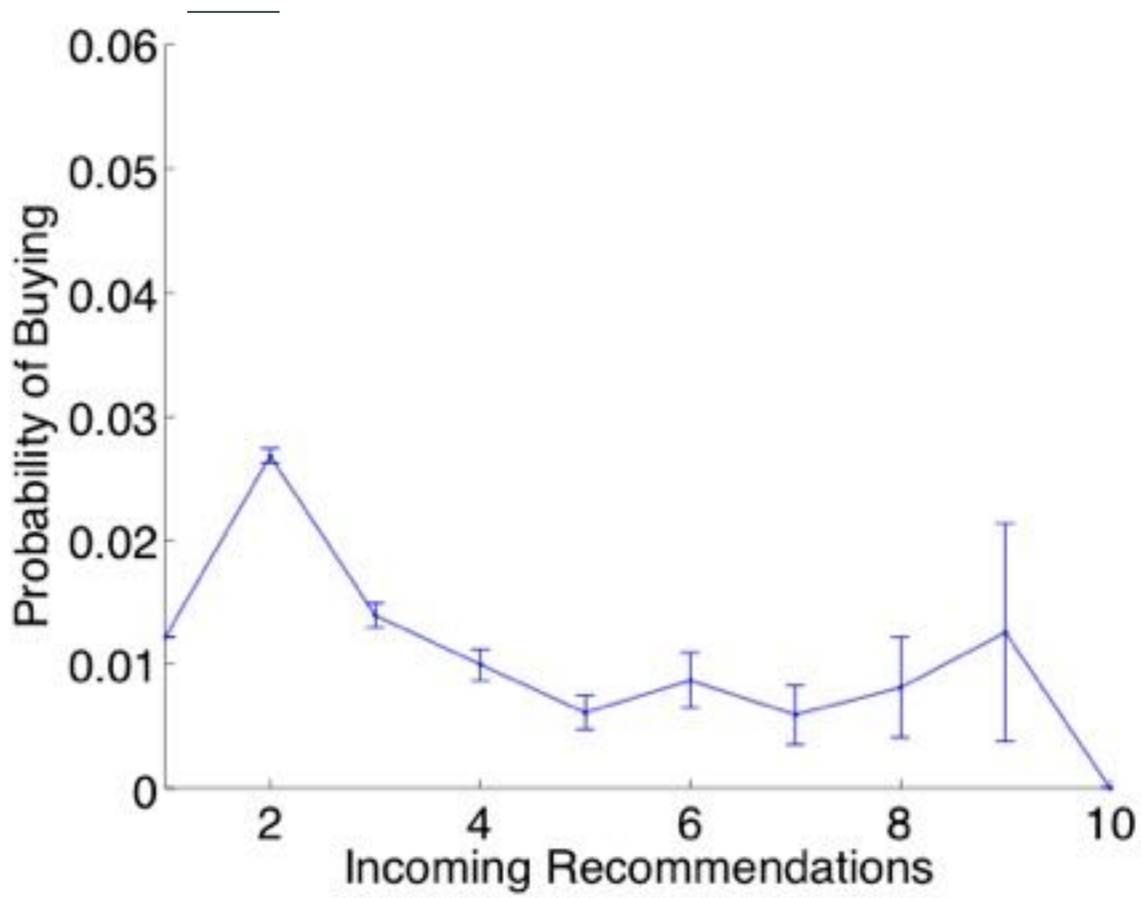


Figure 4: Probability of buying a book given a number of incoming recommendations.

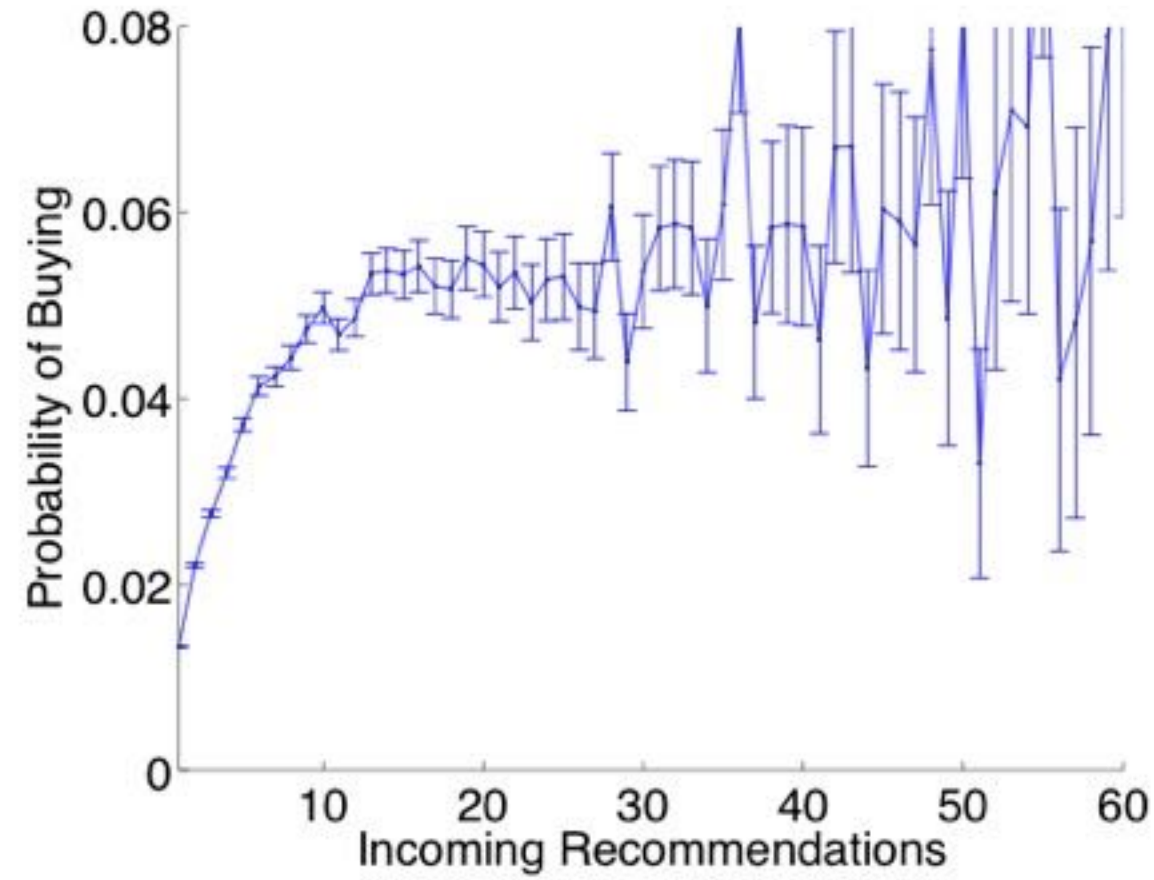


Figure 5: Probability of buying a DVD given a number of incoming recommendations.

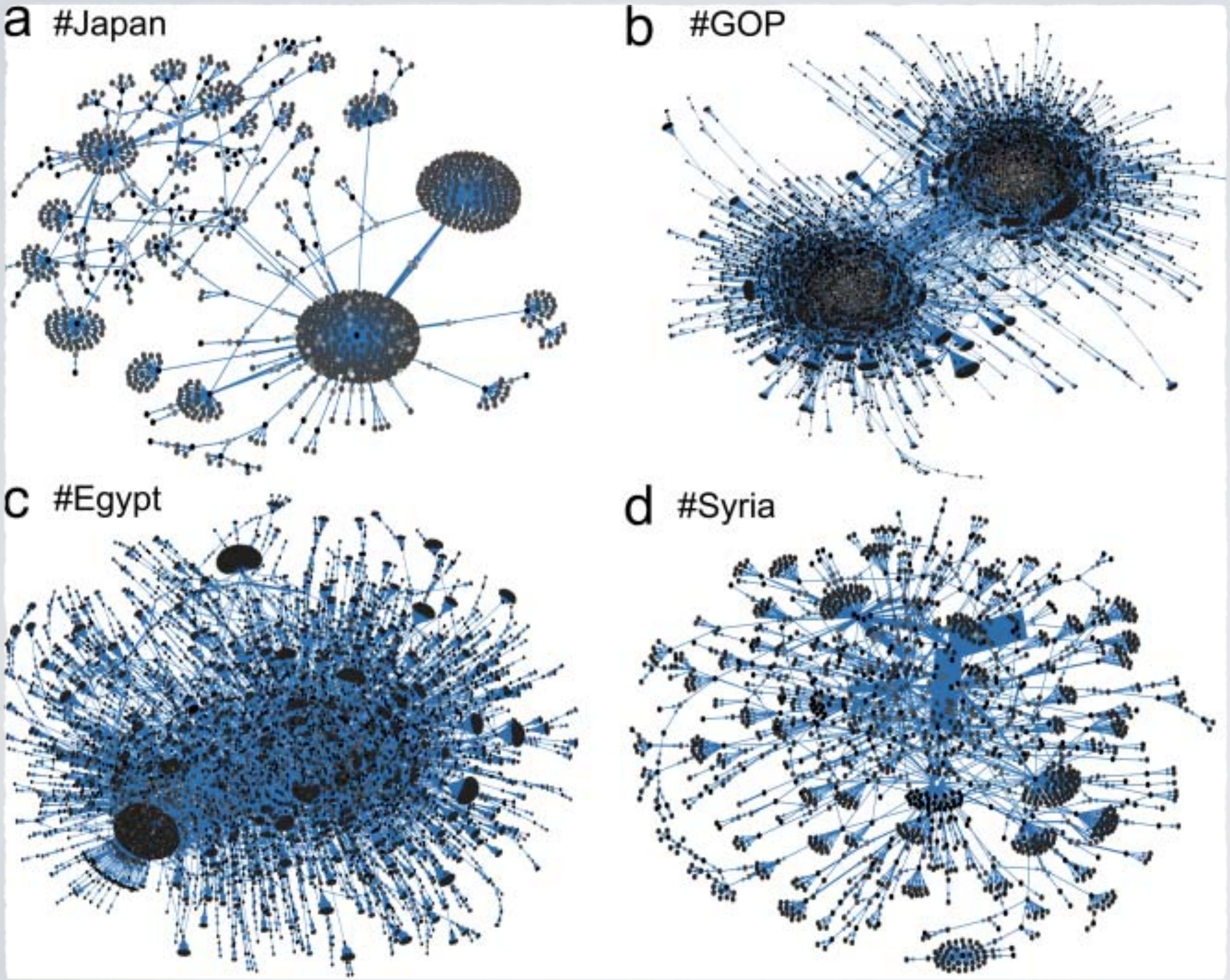
The dynamics of viral marketing

(Leskovec, Adamic, Huberman, 2006)

Competition among memes in a world with limited attention

L. Weng[1], A. Flammini[1], A. Vespignani[2,3,4] & F. Menczer[1]

Published
29 March 2012



(a) The #Japan meme shows how news about the March 2011 earthquake propagated.

(b) The #GOP tag stands for the US Republican Party and as many political memes, displays a strong polarization between people with opposing views.

(c) Memes related to the ‘‘Arab Spring’’ and in particular the 2011 uprisings in #Egypt

(d) and in #Syria These memes display characteristic hub users and strong connections, respectively.

CONTENT DYNAMICS AND DIFFUSION

same on Twitter with *hashtags*...

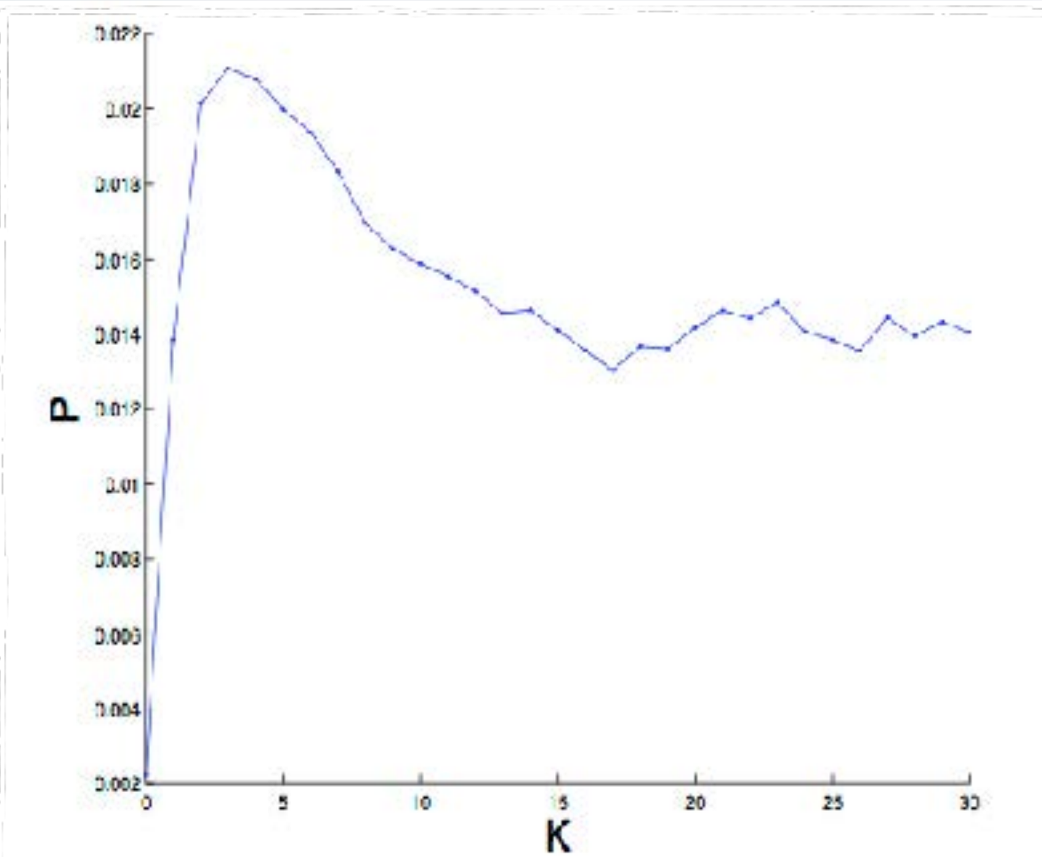
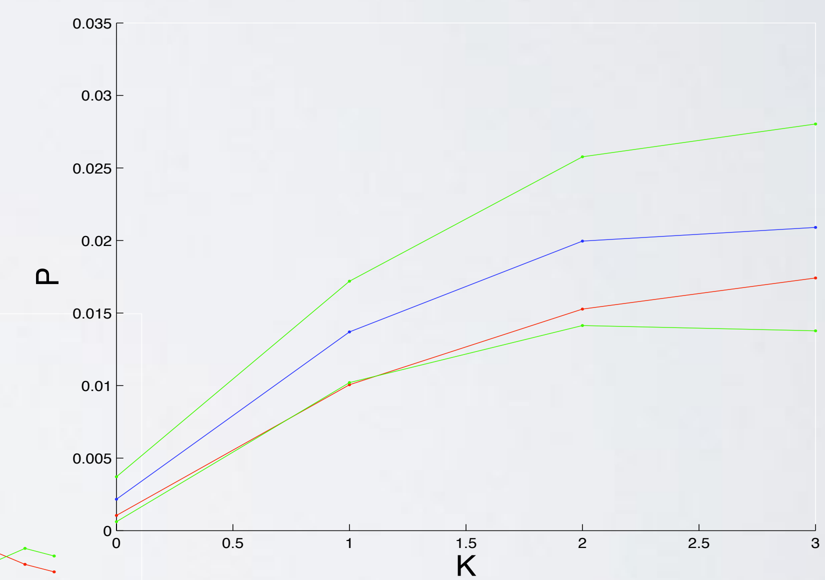


Figure 1: Average exposure curve for the top 500 hashtags. $P(K)$ is the fraction of users who adopt the hashtag directly after their k^{th} exposure to it, given that they had not yet adopted it

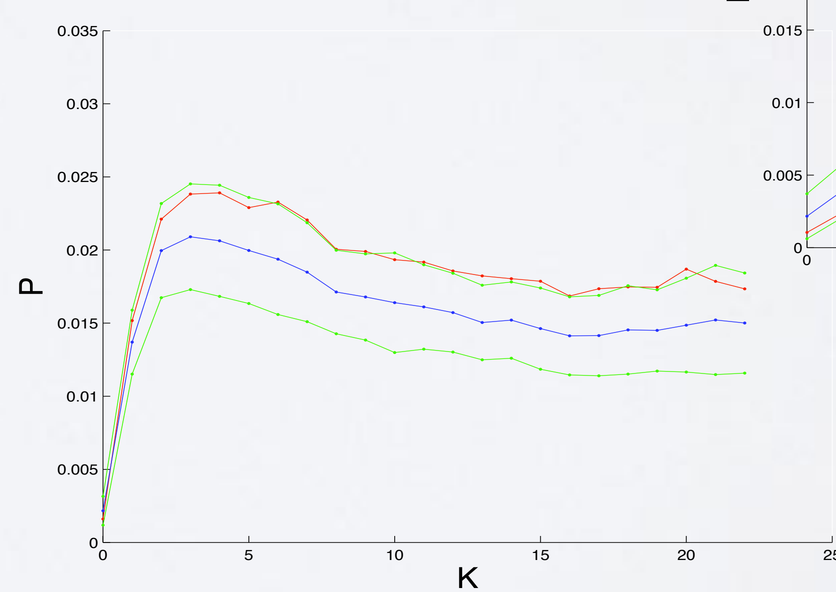
Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter

WWW 2011, March 28–April 1, 2011, Hyderabad, India.

Daniel M. Romero Brendan Meeder Jon Kleinberg



(h) Games



(f) Political

Diffusion observation

...with respect to community structure

- effect of **clustered communities**?
  - contradictory influence of:
    - homophily (communities reinforcing contagion through multiple exposures)
    - clustering (slowing random information flows).

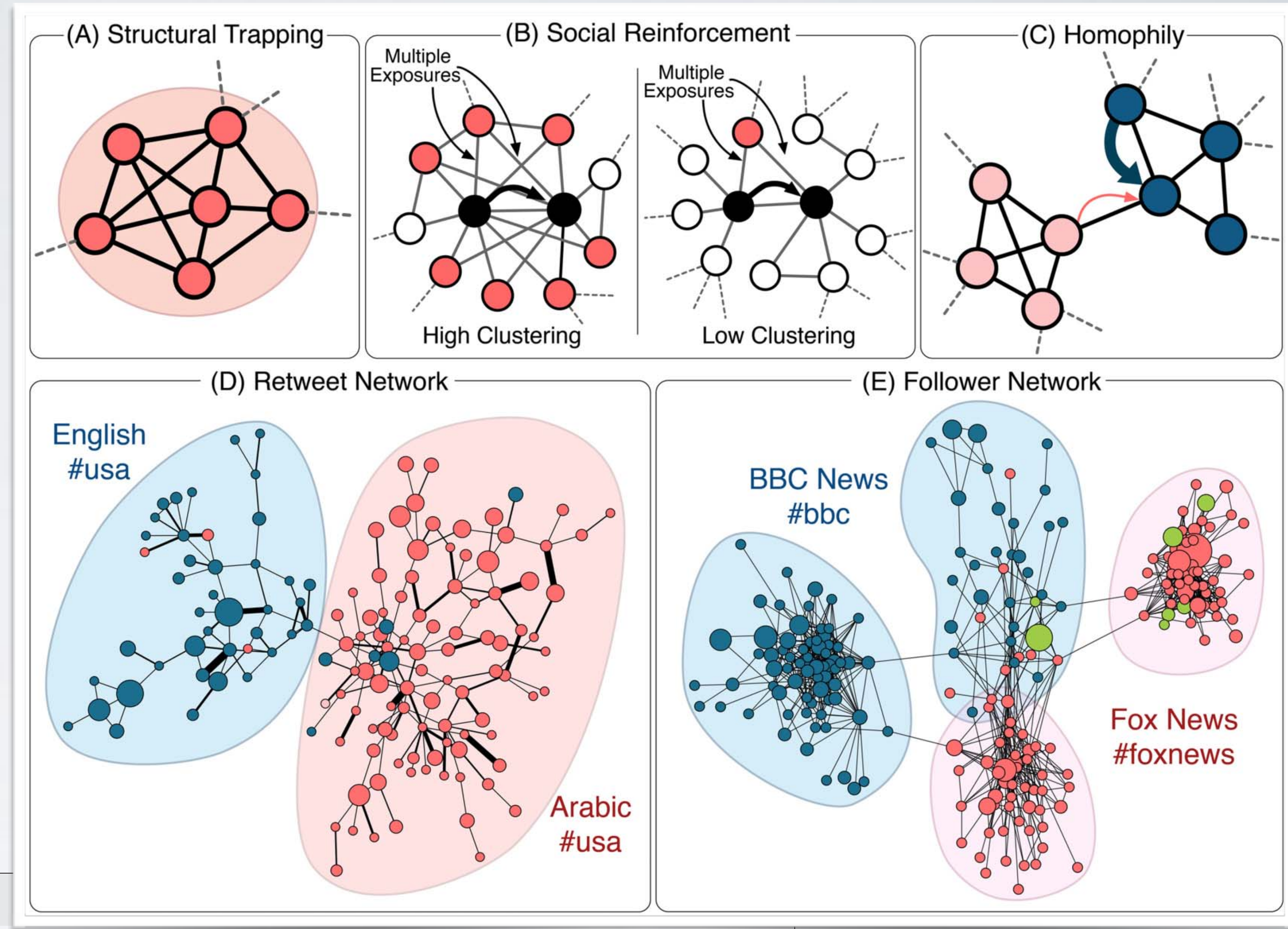


Table 1 | Baseline models for information diffusion

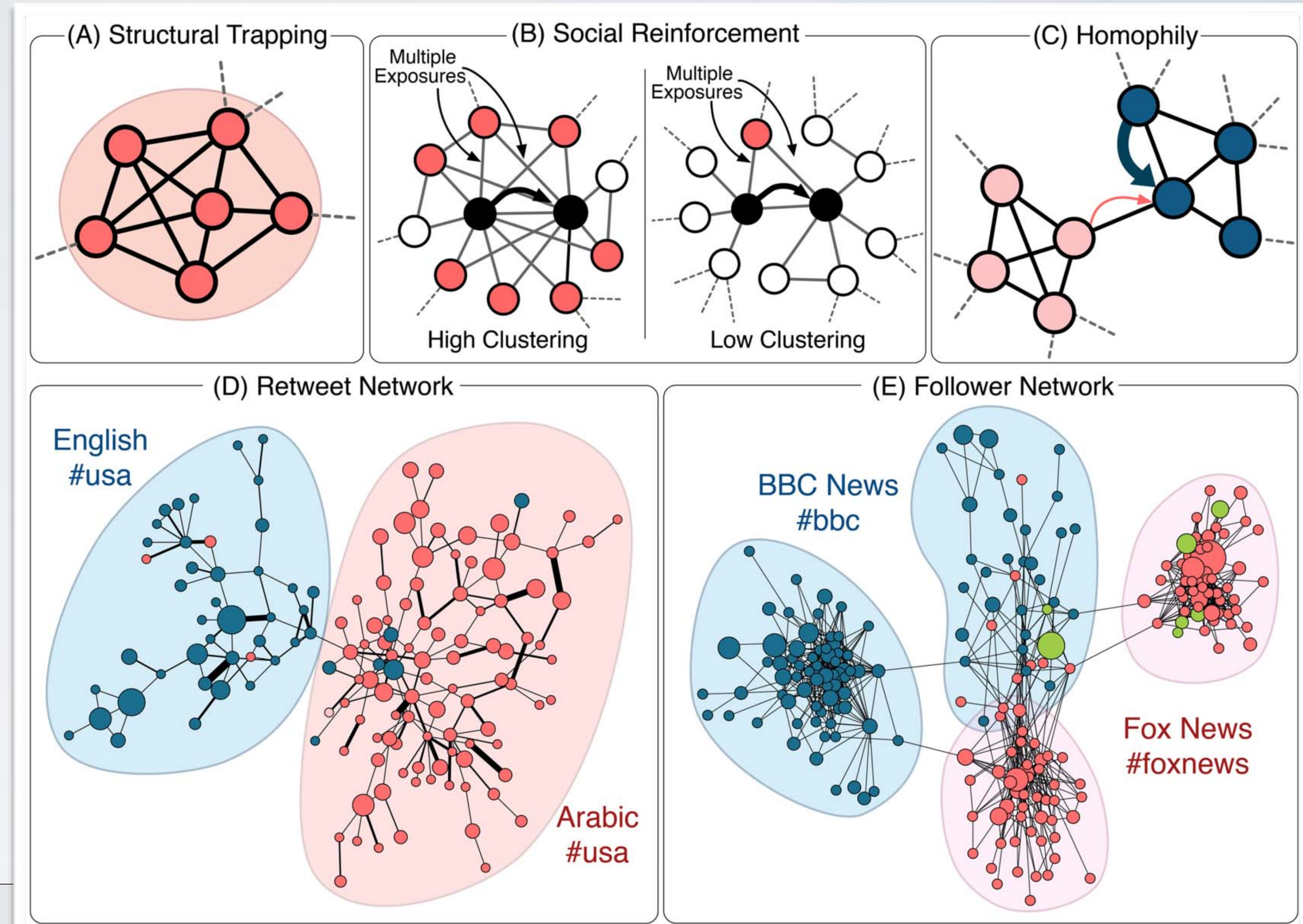
	Community effects			
	Network	Reinforcement	Homophily	Simulation implementation
M_1				For a given hashtag h, M_1 randomly samples the same number of tweets or users as in the real data.
M_2	✓			M_2 takes the network structure into account while neglecting social reinforcement and homophily. M_2 starts with a random seed user. At each step, with probability p, an infected node is randomly selected and one of its neighbors adopts the meme, or with probability $1 - p$, the process restarts from a new seed user ($p = 0.85$).
M_3	✓	✓		The cascade in M_3 is generated similarly to M_2 but at each step the user with the maximum number of infected neighbors adopts the meme.
M_4	✓		✓	In M_4, the simple cascading process is simulated in the same way as in M_2 but subject to the constraint that at each step, only neighbors in the same community have a chance to adopt the meme.

- "**viral**" (irrespective of community structure, disease-like spreading: clustering slows diffusion)
- vs. "**non-viral**" (community structure-dependent, clustering facilitates diffusion)

Diffusion observation

(Weng, Menczer, Ahn, 2013)

...with respect to community structure



- effect of **clustered communities**?
    - contradictory influence of:
        - homophily (communities reinforcing contagion through multiple exposures)
        - clustering (slowing random information flows).

Table 1 | Baseline models for information diffusion

	Community effects: Network	Reinforcement	Homophily	Simulation implementation
M_1				For a given hashtag h, M_1 randomly samples the same number of tweets or users as in the real data.
M_2	✓			M_2 takes the network structure into account while neglecting social reinforcement and homophily. M_2 starts with a random seed user. At each step, with probability p, an infected node is randomly selected and one of its neighbors adopts the meme, or with probability $1 - p$, the process restarts from a new seed user ($p = 0.85$).
M_3	✓	✓		The cascade in M_3 is generated similarly to M_2 but at each step the user with the maximum number of infected neighbors adopts the meme.
M_4	✓		✓	In M_4, the simple cascading process is simulated in the same way as in M_2 but subject to the constraint that at each step, only neighbors in the same community have a chance to adopt the meme.

- "**viral**" (irrespective of community structure, disease-like spreading: clustering slows diffusion)
- vs. "**non-viral**" (community structure-dependent, clustering facilitates diffusion)

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

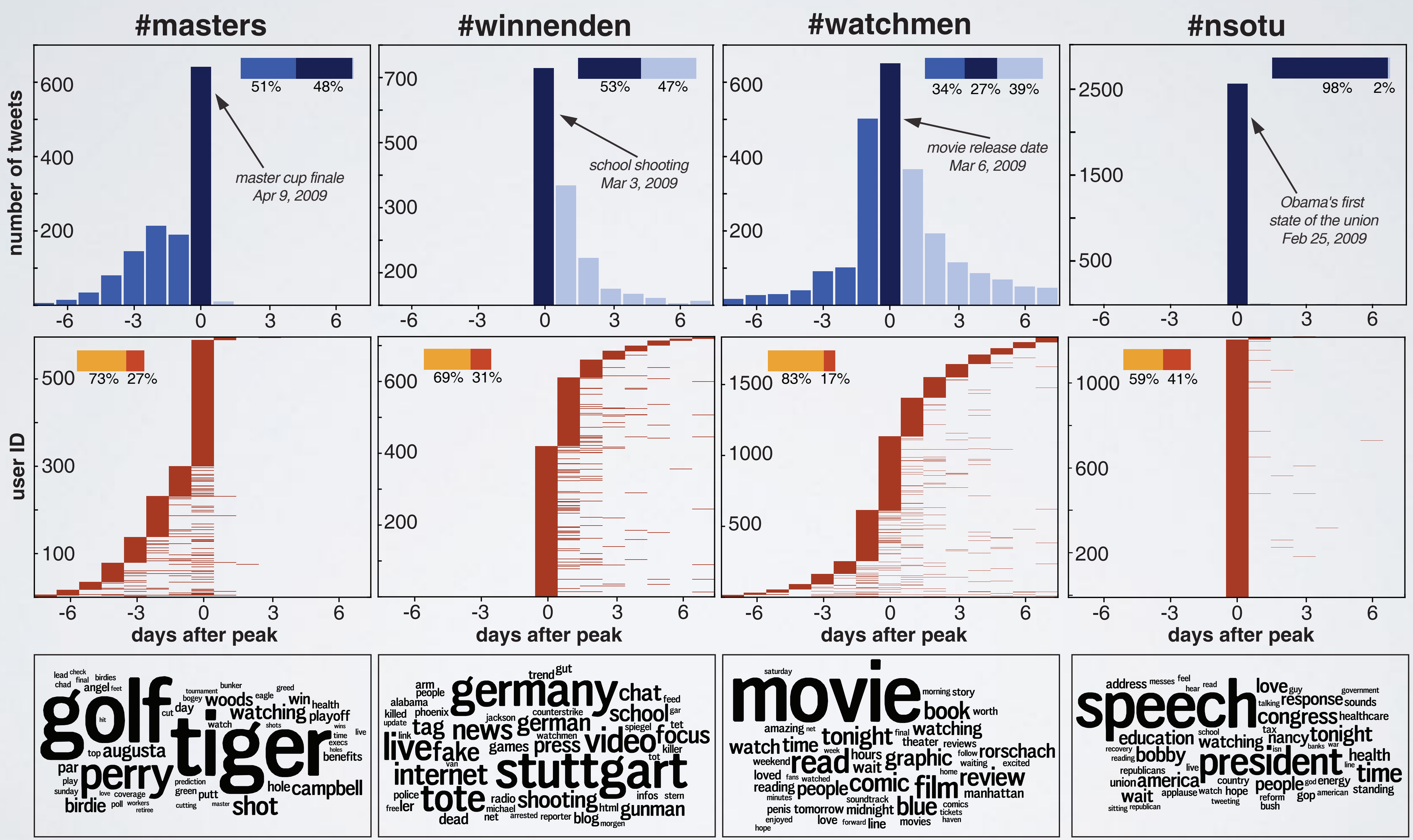


Figure 1: Activity associated with four hashtags that exhibit a popularity peak: daily activity over time (top row), individual user activity (middle) row, and word clouds of tweet content (bottom row).

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

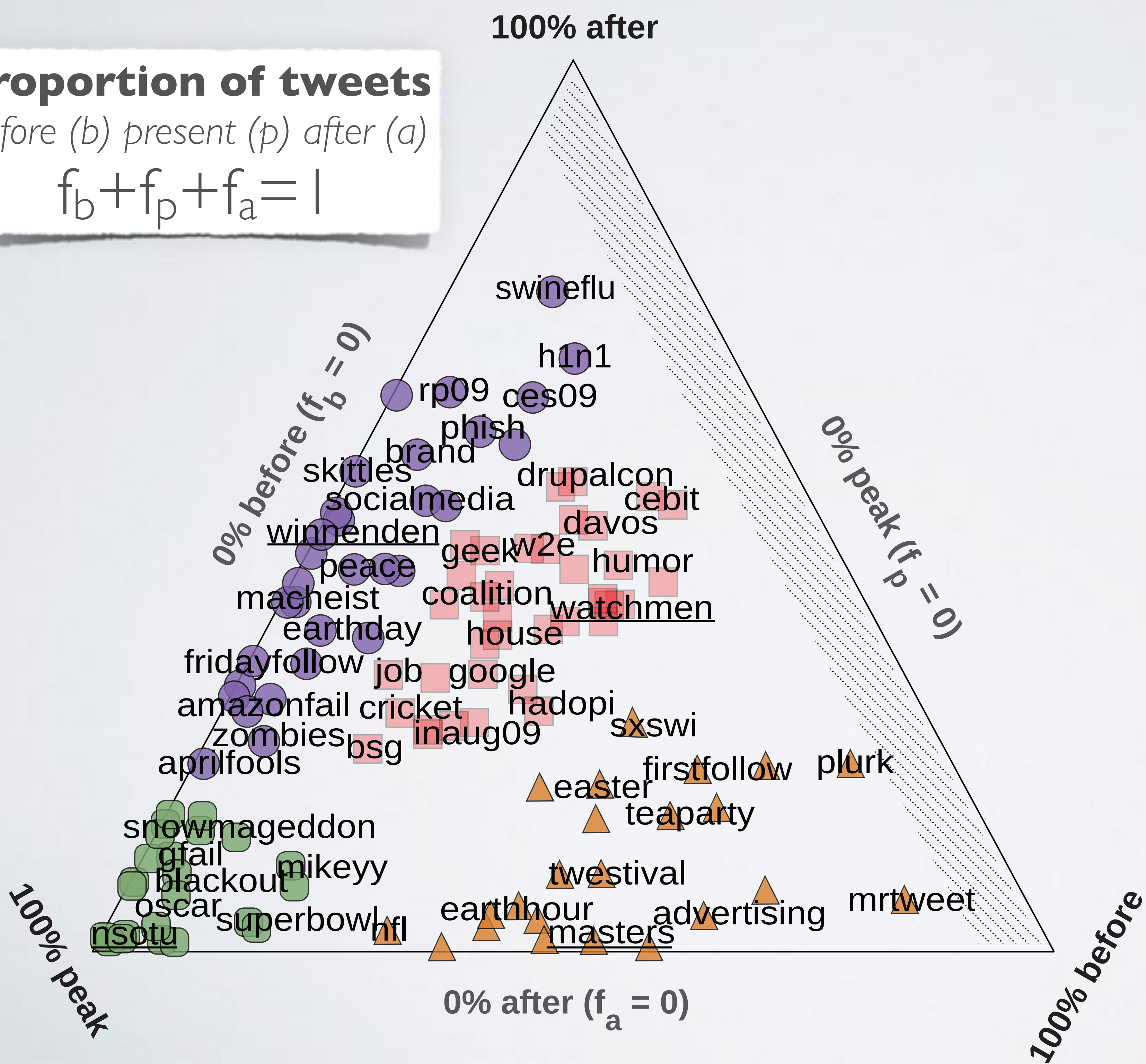
Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

proportion of tweets

before (b) present (p) after (a)

$$f_b+f_p+f_a=1$$



Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

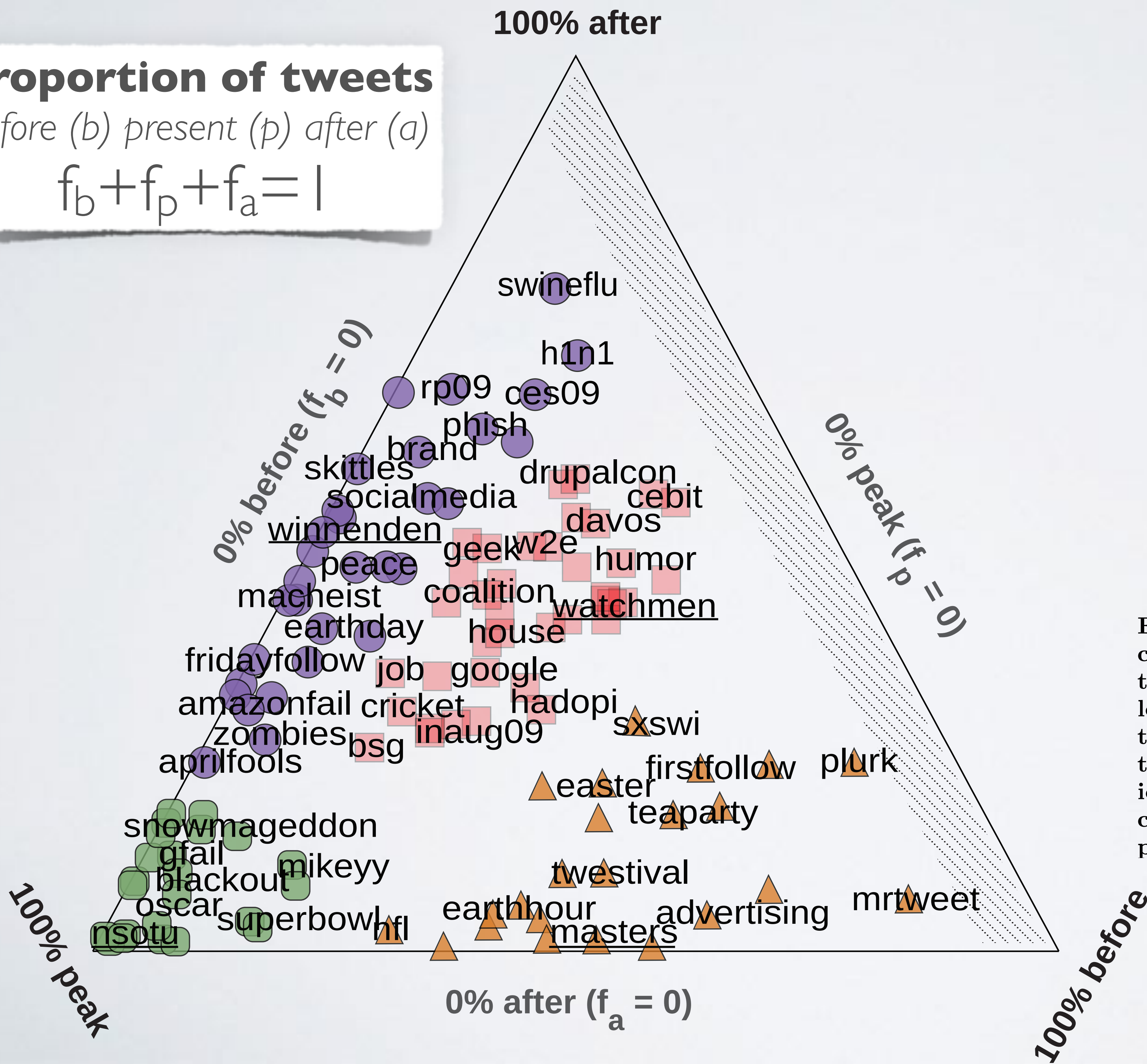
Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

proportion of tweets

before (b) present (p) after (a)

$$f_b+f_p+f_a=1$$



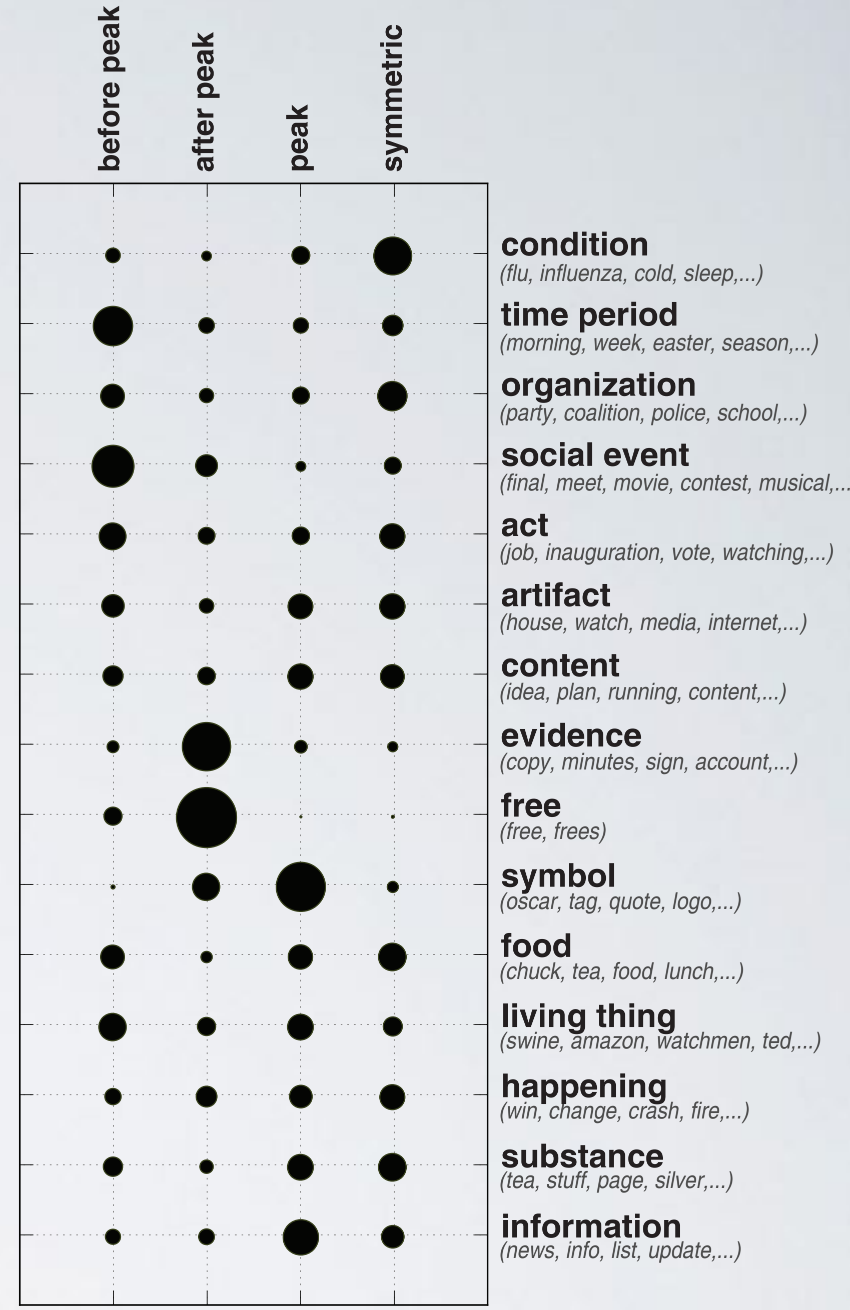


Figure 4: Semantic makeup of the hashtag classes: columns represent peak types and rows correspond to topics, i.e., concepts in the WordNet semantic lexicon. The radius of a circle is proportional to the average normalized frequency of the topic in the corresponding hashtag class. The displayed topics represent the most frequently observed generic concepts. Sample terms subsumed by them are reported in parenthesis.

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

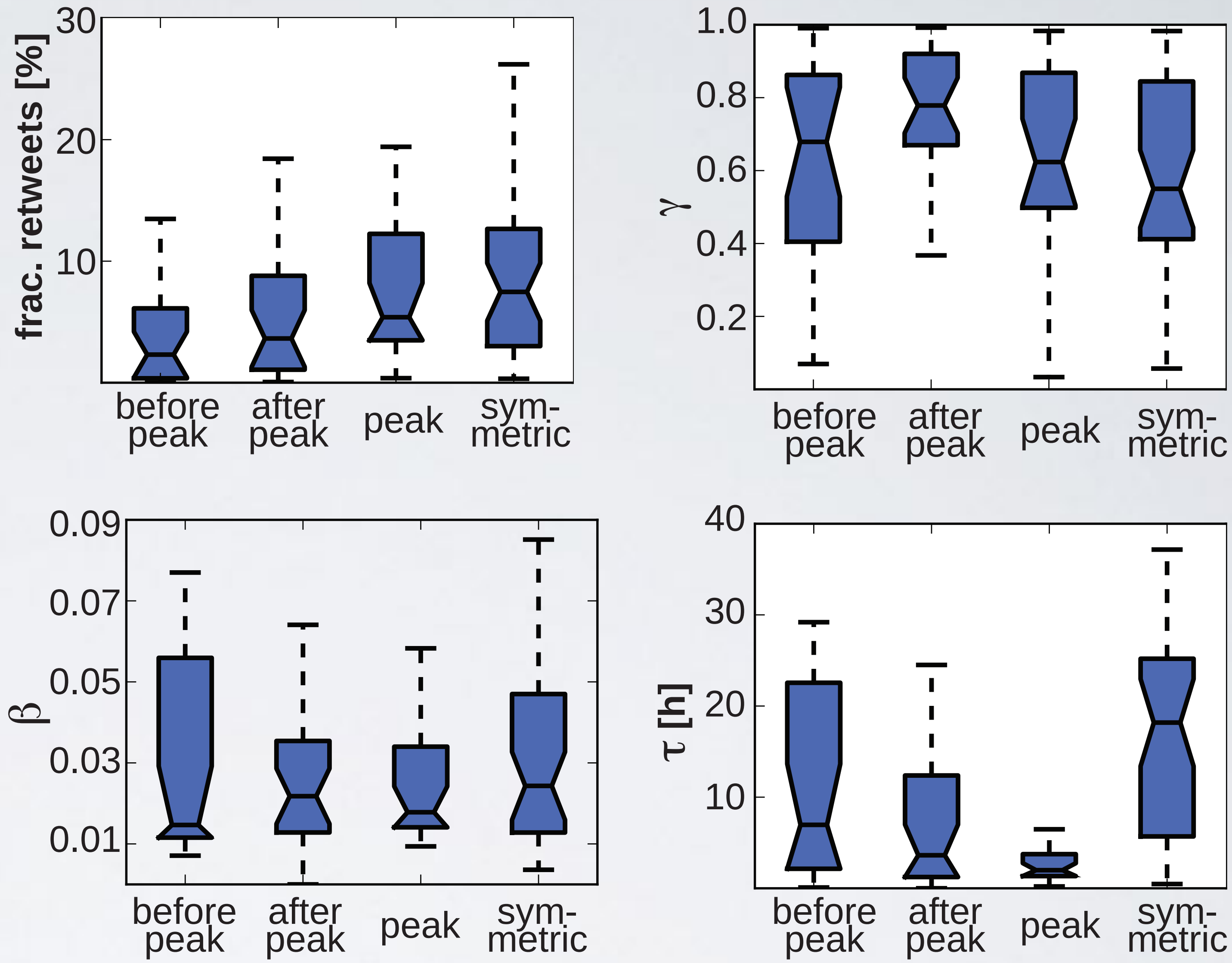


Figure 5: Parameters controlling the spreading of hashtags, broken down by hashtag class. Top left: fraction of retweets to regular tweets. Top right: fraction of seeders γ. Bottom left: fraction β of followers that adopt the hashtag after seeing it. Bottom right: average time τ between the first tweet with the hashtag and the last one.

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

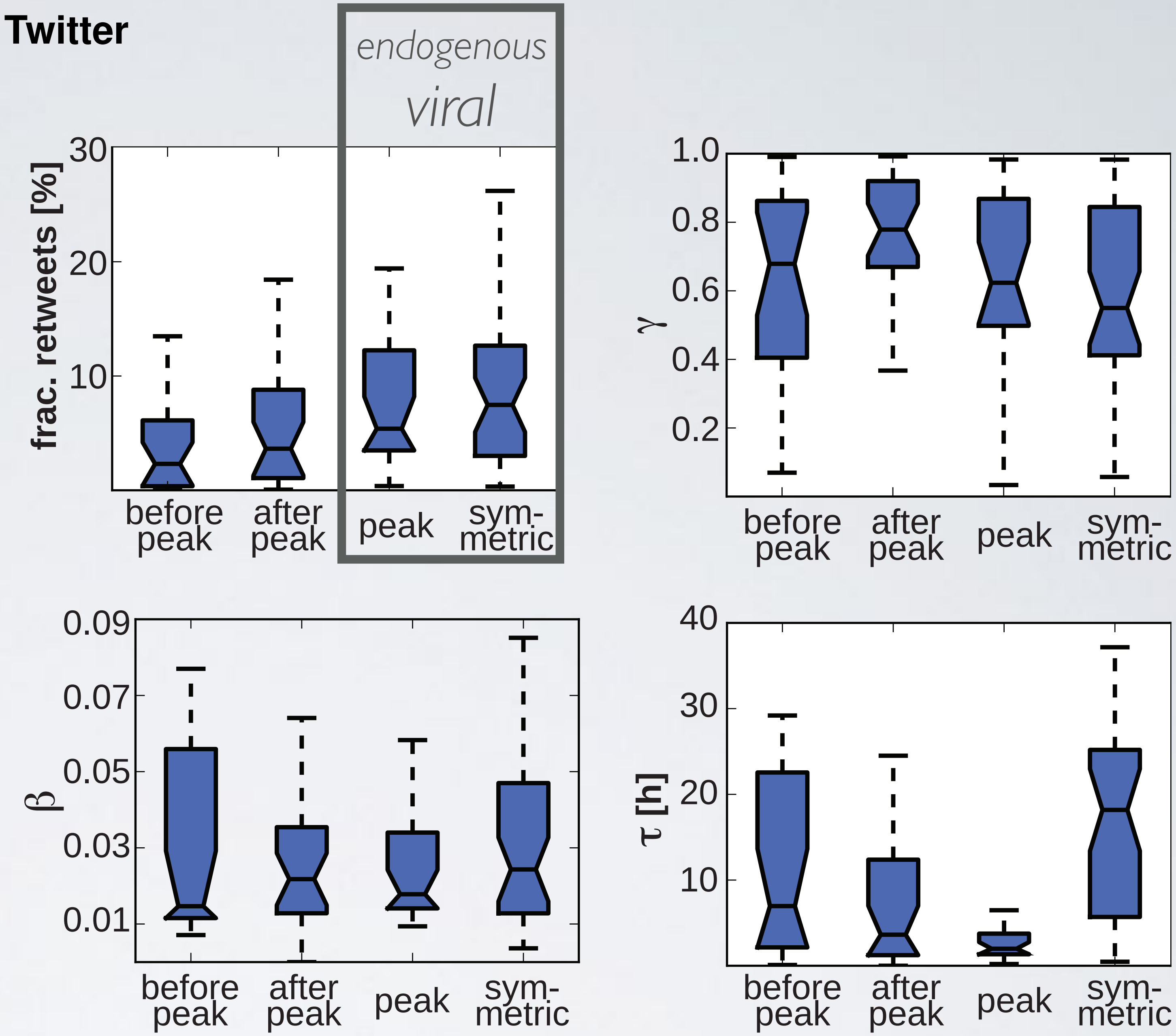


Figure 5: Parameters controlling the spreading of hashtags, broken down by hashtag class. Top left: fraction of retweets to regular tweets. Top right: fraction of seeders γ. Bottom left: fraction β of followers that adopt the hashtag after seeing it. Bottom right: average time τ between the first tweet with the hashtag and the last one.

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

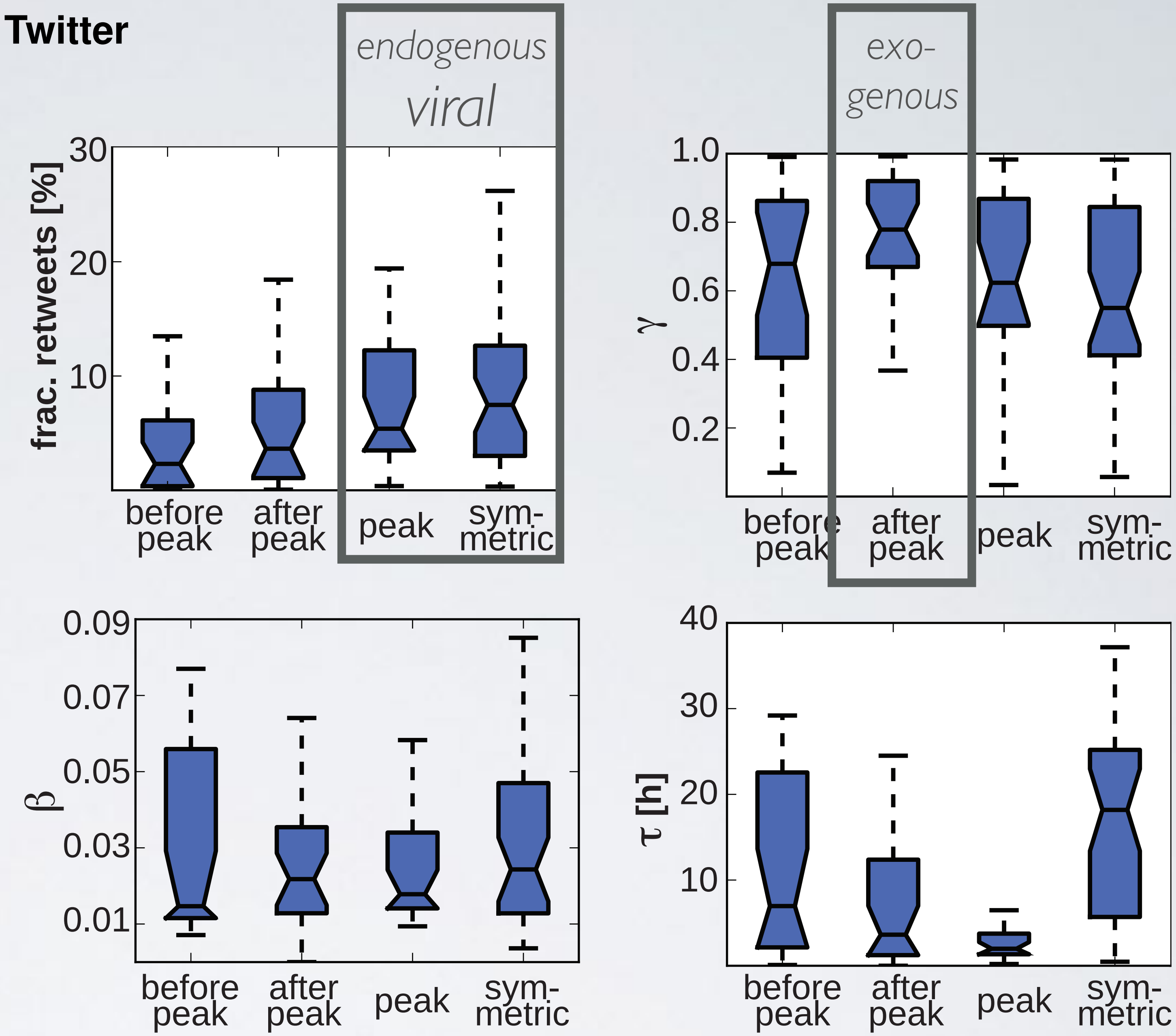


Figure 5: Parameters controlling the spreading of hashtags, broken down by hashtag class. Top left: fraction of retweets to regular tweets. Top right: fraction of seeders γ. Bottom left: fraction β of followers that adopt the hashtag after seeing it. Bottom right: average time τ between the first tweet with the hashtag and the last one.

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

parameter β measures contagiosity: it indicates whether a meme propagates from a user to their *followers*

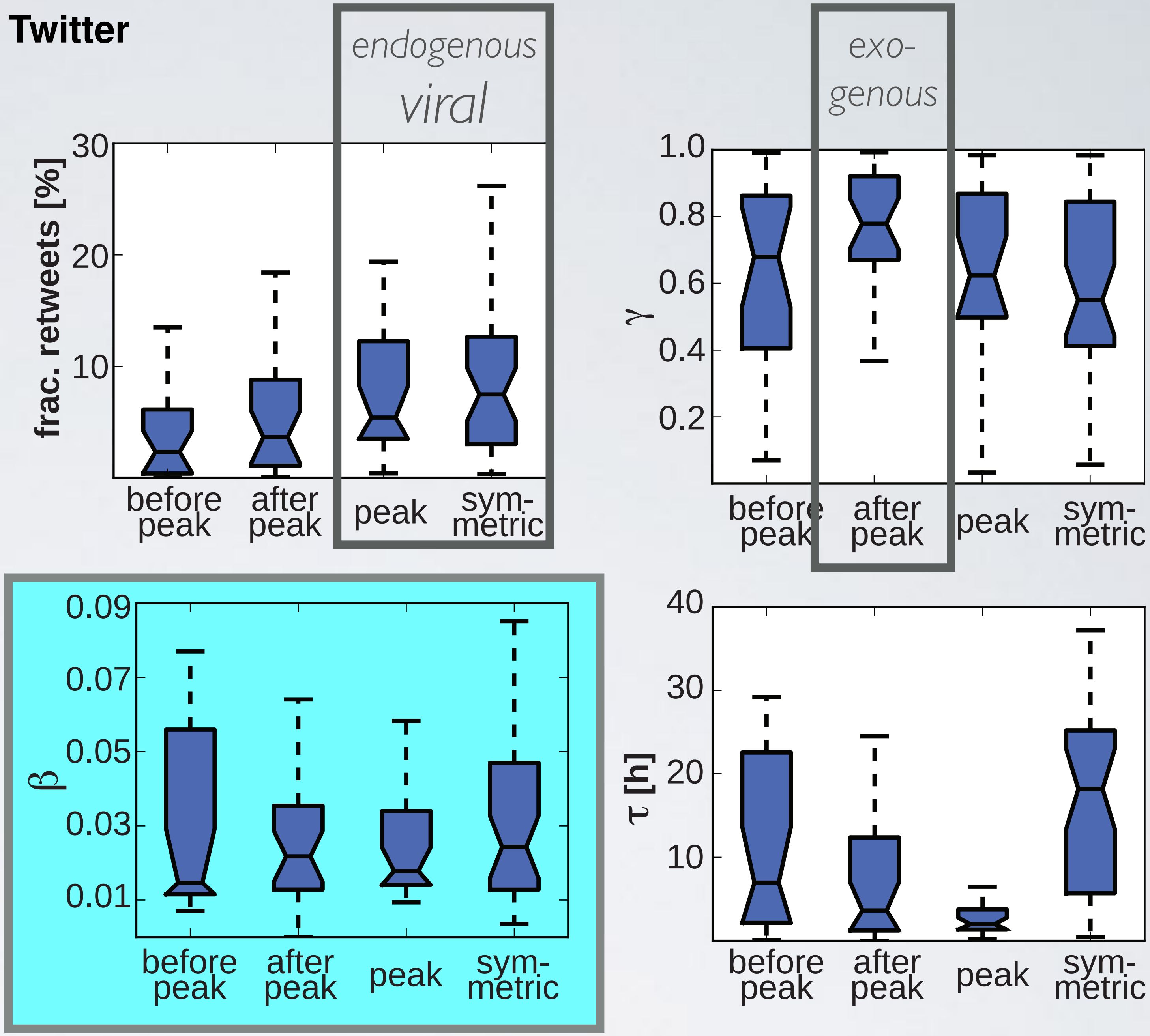


Figure 5: Parameters controlling the spreading of hashtags, broken down by hashtag class. Top left: fraction of retweets to regular tweets. Top right: fraction of seeders γ. Bottom left: fraction β of followers that adopt the hashtag after seeing it. Bottom right: average time τ between the first tweet with the hashtag and the last one.

Dynamical Classes of Collective Attention in Twitter

Janette Lehmann
Web Research Group,
Universitat Pompeu Fabra
Barcelona, Spain
janette.lehmann@gmx.de

Bruno Gonçalves
College of Computer and
Information Sciences
Northeastern University
b.goncalves@neu.edu

José J. Ramasco
IFISC (CSIC-UIB)
Palma de Mallorca, Spain
jramasco@ifisc.uib.es

Ciro Cattuto
ISI Foundation
Torino, Italy
ciro.cattuto@isi.it

(2012)

parameter β measures contagiosity: it indicates whether a meme propagates from a user to their *followers*

τ denotes the time during which users are susceptible to propagate hashtags to their followers.

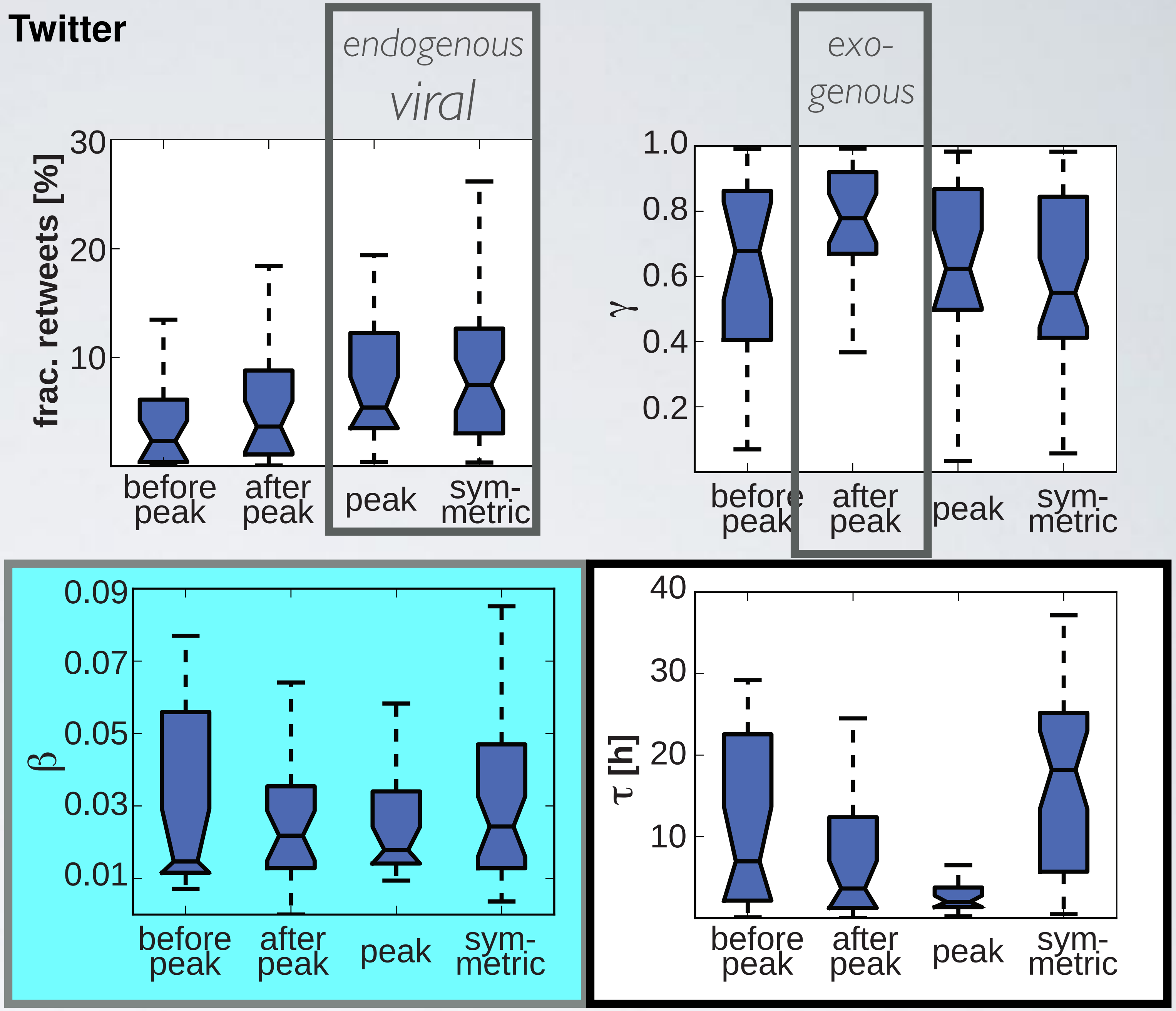
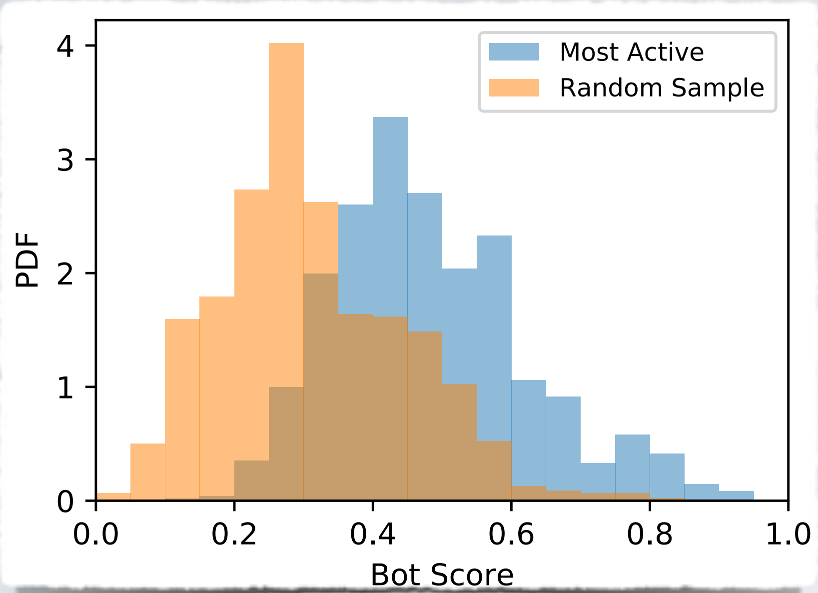


Figure 5: Parameters controlling the spreading of hashtags, broken down by hashtag class. Top left: fraction of retweets to regular tweets. Top right: fraction of seeders γ. Bottom left: fraction β of followers that adopt the hashtag after seeing it. Bottom right: average time τ between the first tweet with the hashtag and the last one.

Shao, Ciampaglia, Varol, Flammini, Menczer, 2017

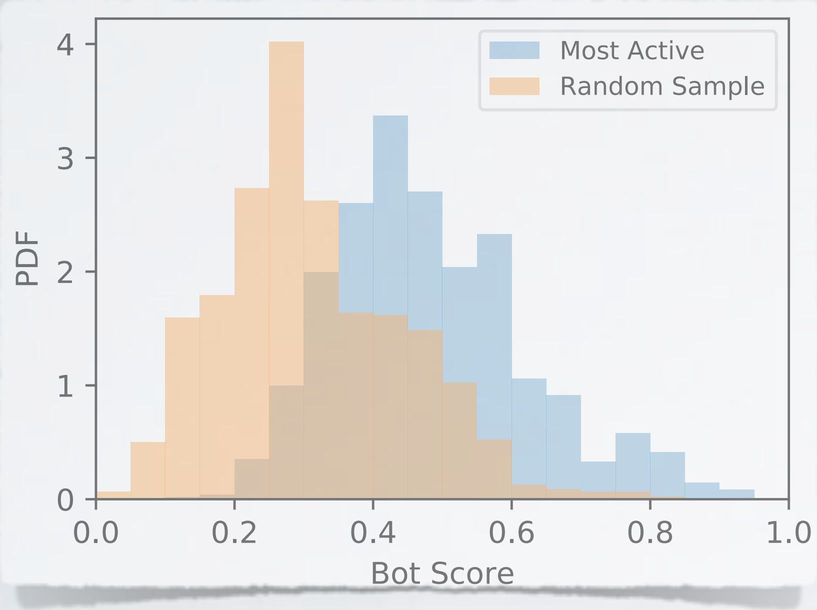
"The spread of misinformation by social bots"



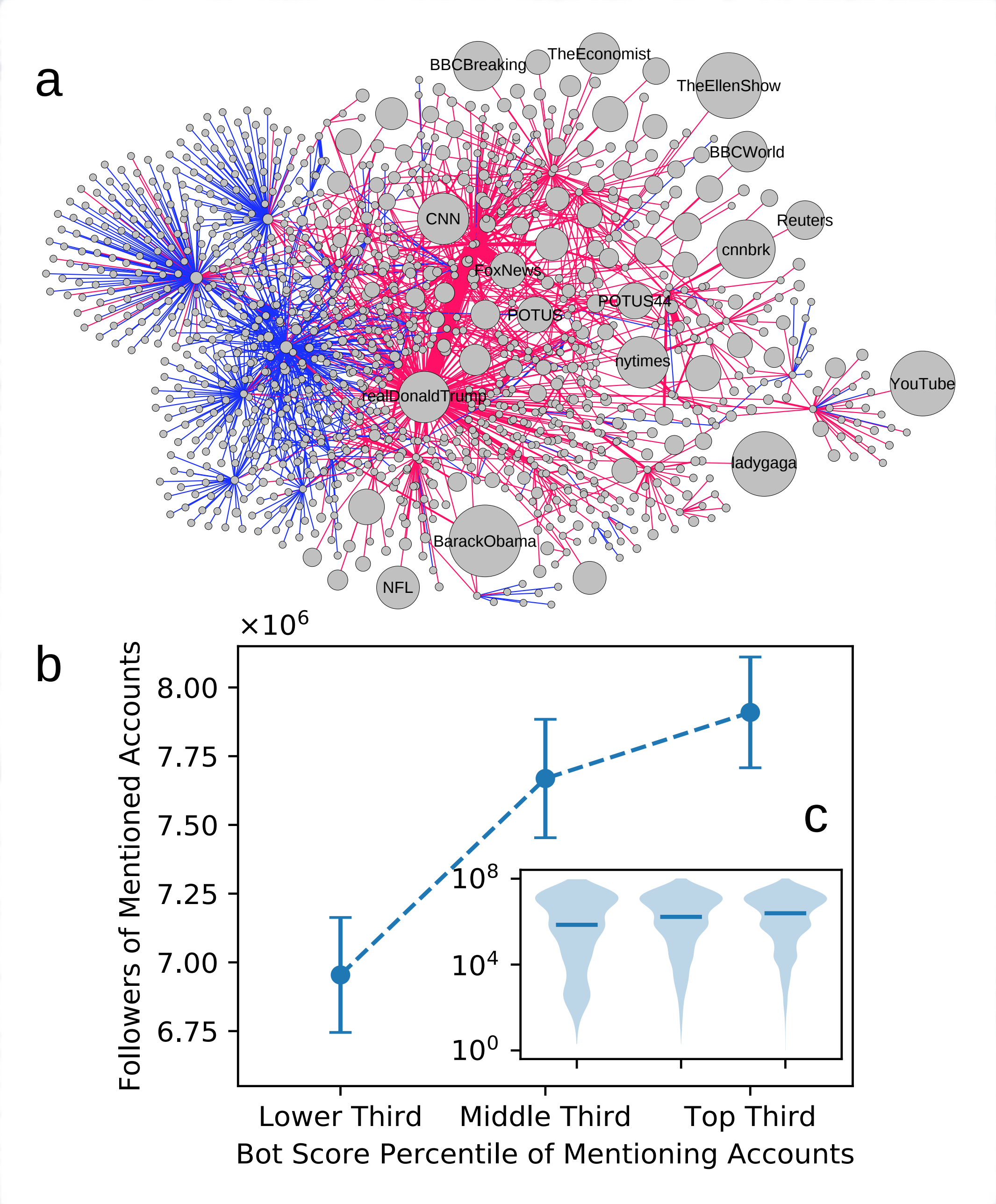
1. most active users are more likely to be bots

Shao, Ciampaglia, Varol, Flammini, Menczer, 2017

"The spread of misinformation by social bots"



1. most active users are more likely to be bots



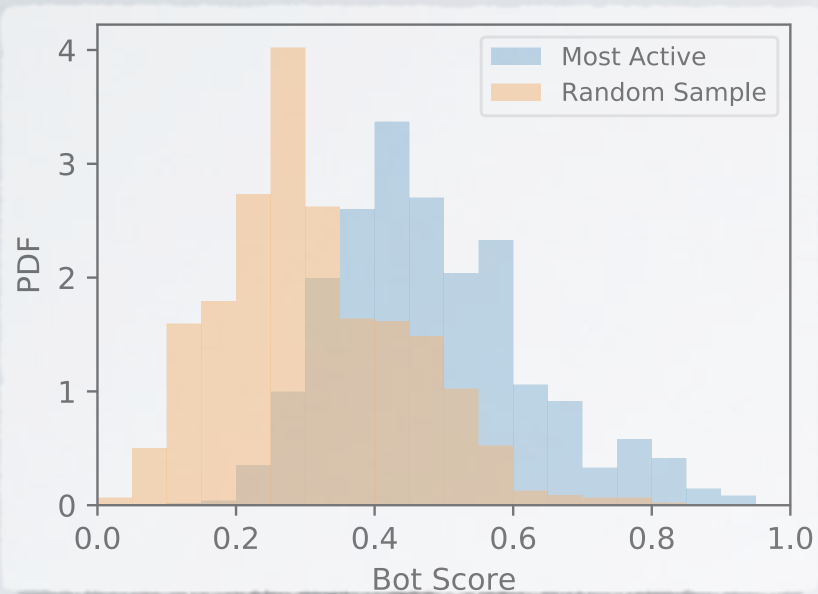
2. bots tend to *target* users with the highest number of followers

blue: *retweets*

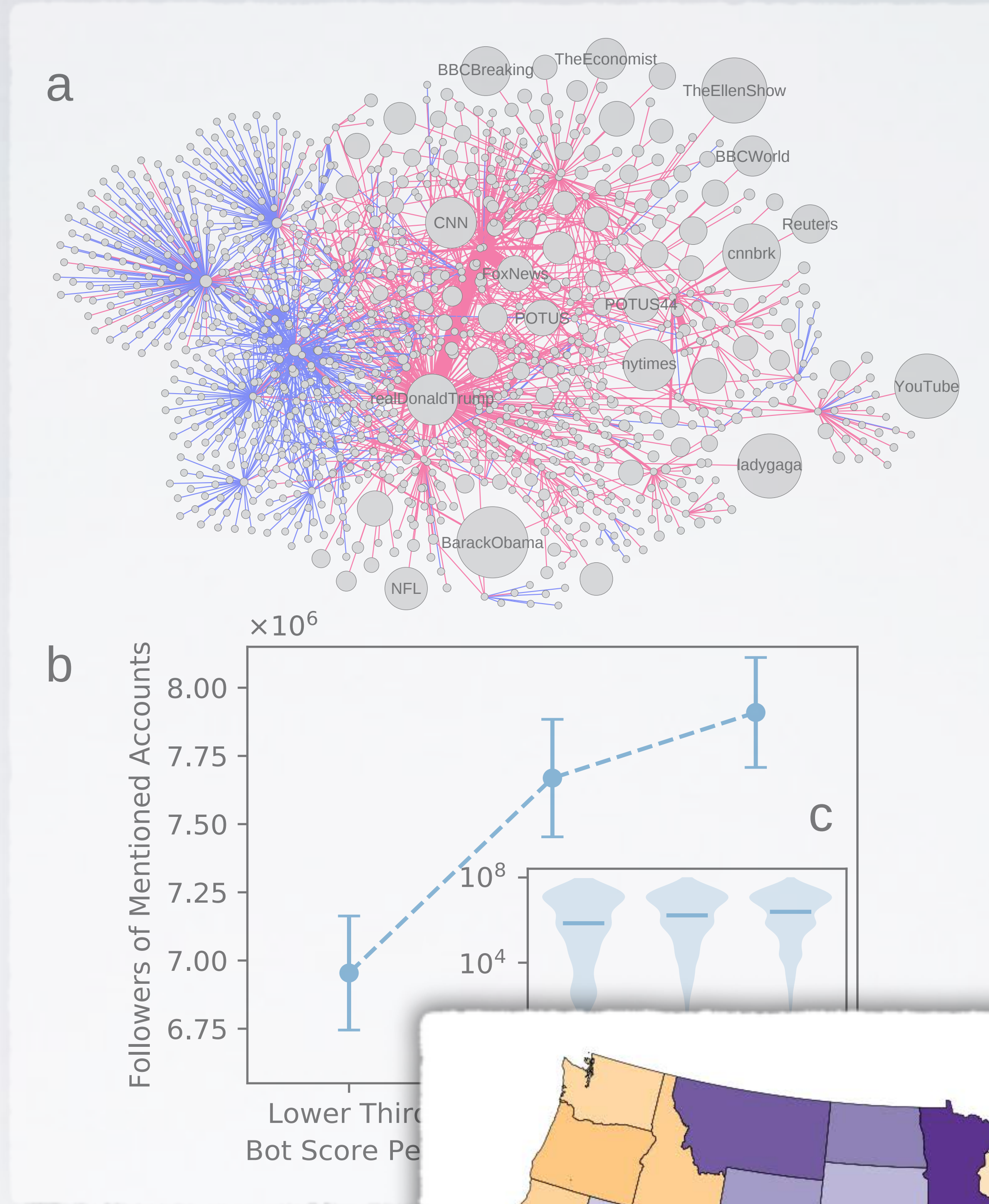
red: *mentions*

Shao, Ciampaglia, Varol, Flammini, Menczer, 2017

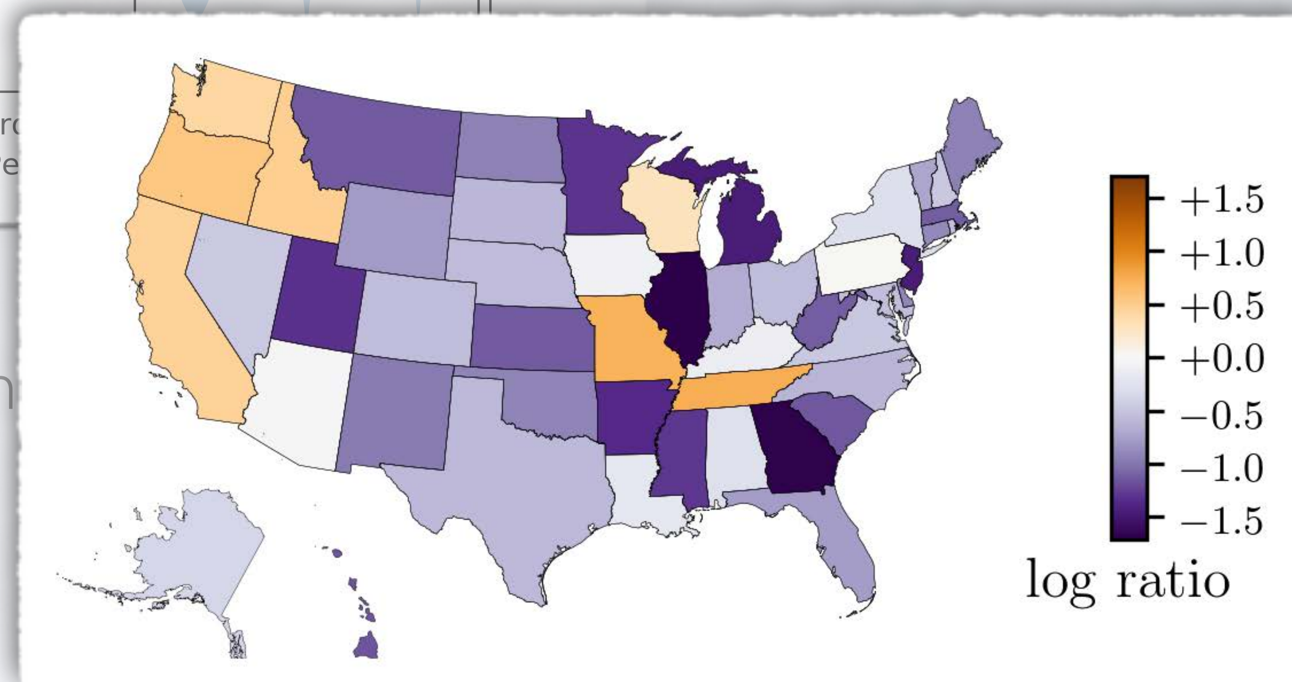
"The spread of misinformation by social bots"



1. most active users are more likely to be bots



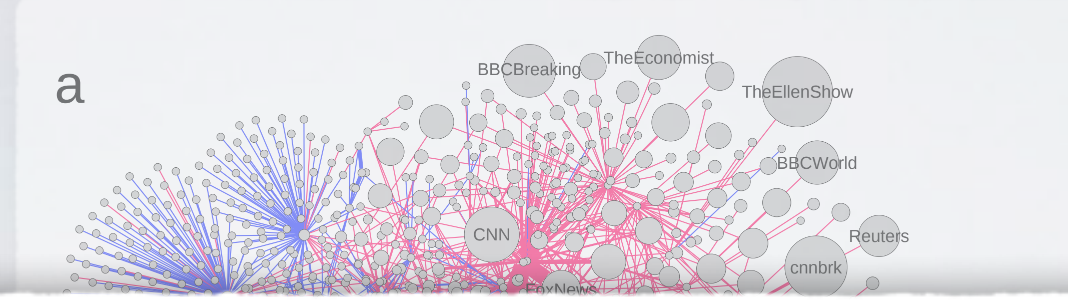
2. bots tend
the highest n

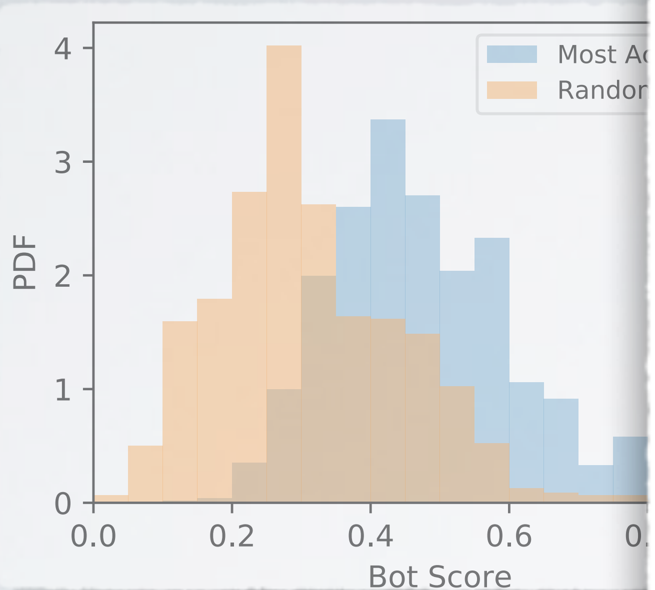


3. misinformation bots are "*overrepresented*" in some states

Shao, Ciampaglia, Varol, Flammini, Menczer, 2017

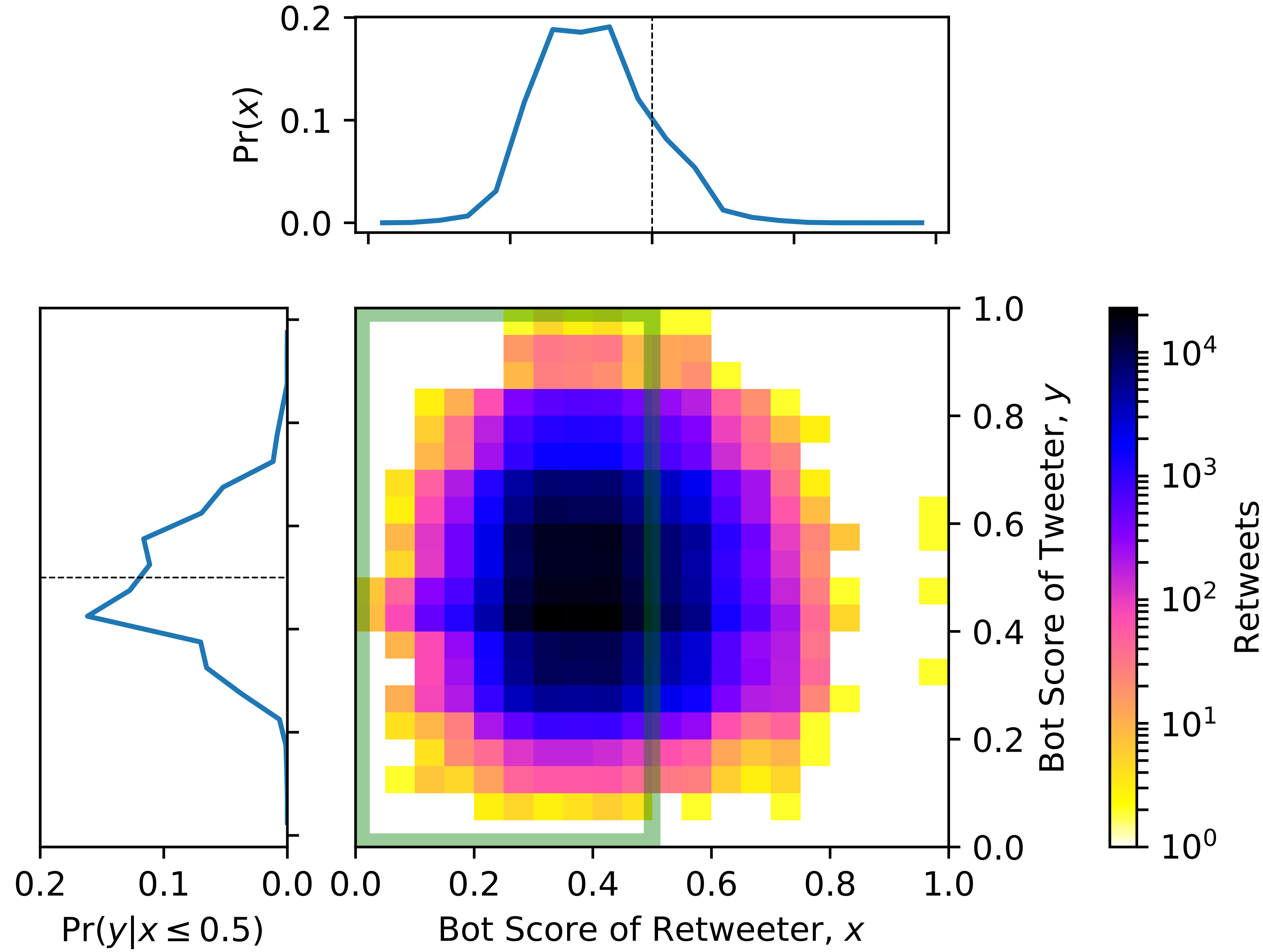
"The spread of misinformation by social bots"





1. most active users

more likely to be bo



4. humans do most of the retweeting *(top)* and retweet as much from humans than bots *(left)*

Mapping the Arabic Blogosphere: Politics, Culture, and Dissent

By Bruce Etling, John Kelly, Robert Faris, and John Palfrey

Berkman
The Berkman Center for Internet & Society at Harvard University

JUNE 2009
Berkman Center Research Publication No. 2009-06

INTERNET & DEMOCRACY project

STRUCTURE OF THE NETWORK AND METHODS OVERVIEW

- **links:**
  citations (interactional)
- **colors:**
  co-citation communities ("topical")
- the opposite could be done: similar topics define visual closeness, interaction groups define colors
- regardless, strong geographic coherence
- while national clusters aren't on national topics (e.g. youth, women's rights, bloggers' rights, poetry)
- international clusters related to international media and international political topics (including islam)

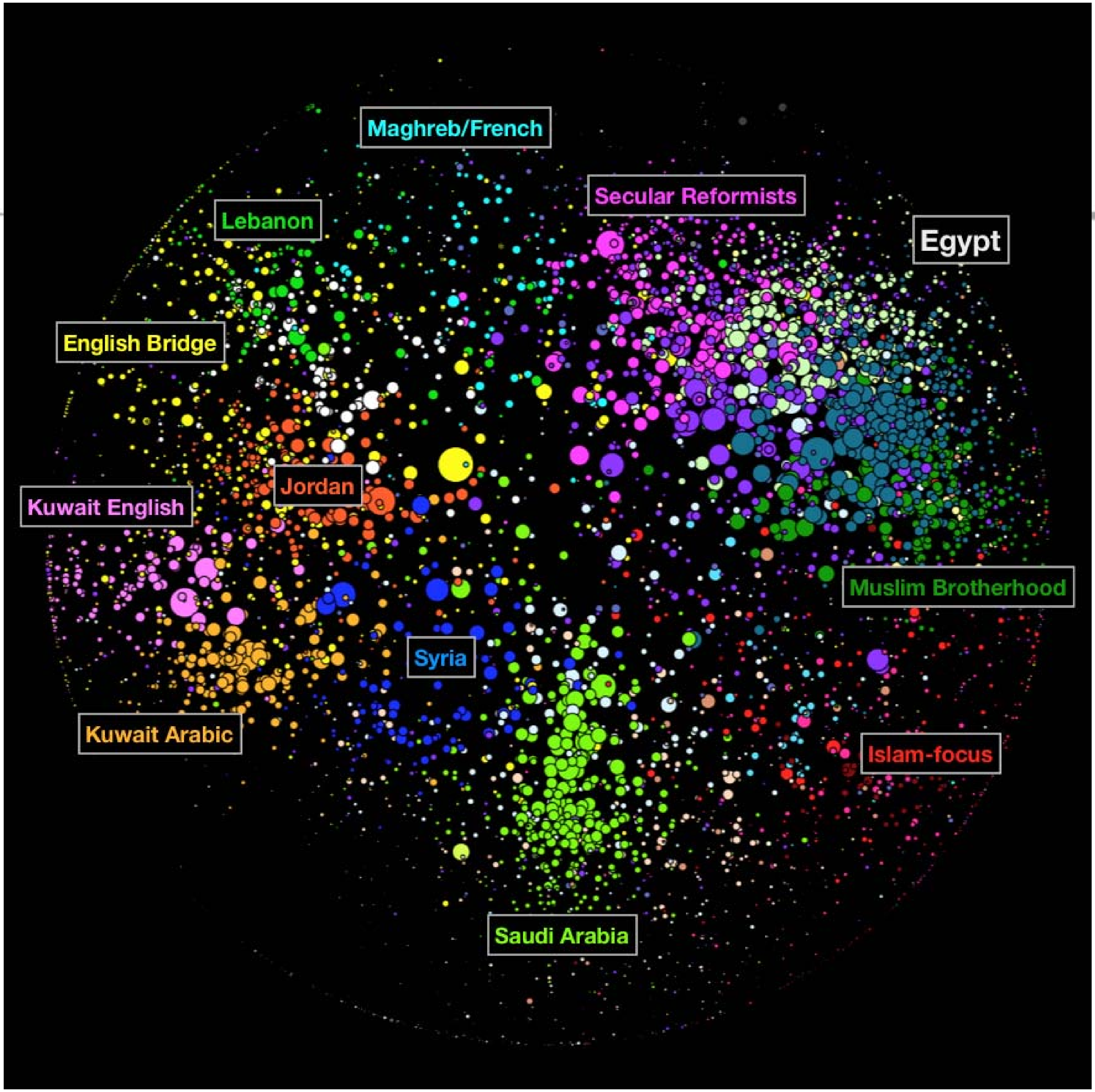
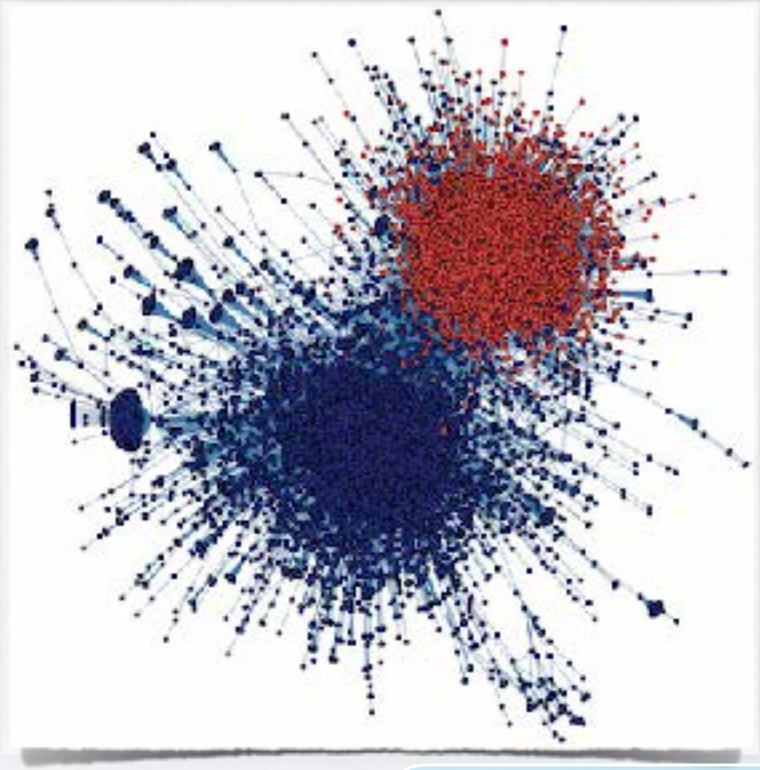


Fig. 1: Map of the Arabic Blogosphere

LEAD EXAMPLE: ONLINE FRAGMENTATION
–
BOTH INFORMATIONAL AND INTERACTIONAL

- Van Alstyne & Brynjolfsson, 1996 — *"balkanization"*
- Sunstein 2001, 09 — *"echo chambers"*
- Pariser, 2011 — *"filter bubbles"*
- Barbera et al. 2015 — *"polarization"*
- Bakshy et al. 2015 — *"selective exposure"*



blue/red network of citations on Twitter, US users

Conover et al. 2011

- *more broadly:*
  peer selection and influence, and its coevolution

e.g. Lewis et al. 2012, in a web context

LEAD EXAMPLE: ONLINE FRAGMENTATION
–
BOTH INFORMATIONAL AND INTERACTIONAL

Van Alstyne & Brynjolfsson, 1996

- "*balkanization*"

Sunstein 2001, 09

- "*echo chambers*"

Pariser, 2011

- "*filter bubbles*"

Barbera et al. 2015

- "*polarization*"

Bakshy et al. 2015

- "*selective exposure*"

- *more broadly:*
  peer selection and influence,
  and its coevolution

e.g. Lewis et al. 2012, in a web context

FRAGMENTATION : MICRO-LEVEL VIEWS

FlipFeed

FlipFeed is a Google Chrome Extension that enables Twitter users to replace their own feed with that of another real Twitter user. Powered by deep learning and social network analysis, feeds are selected based on inferred political ideology ("left" or "right") and served to users of the extension. For example, a right-leaning user who uses FlipFeed may load and navigate a left-leaning user's feed, observing the news stories, commentary, and other content they consume. The user can then decide to flip back to their own feed or repeat the process with another feed.

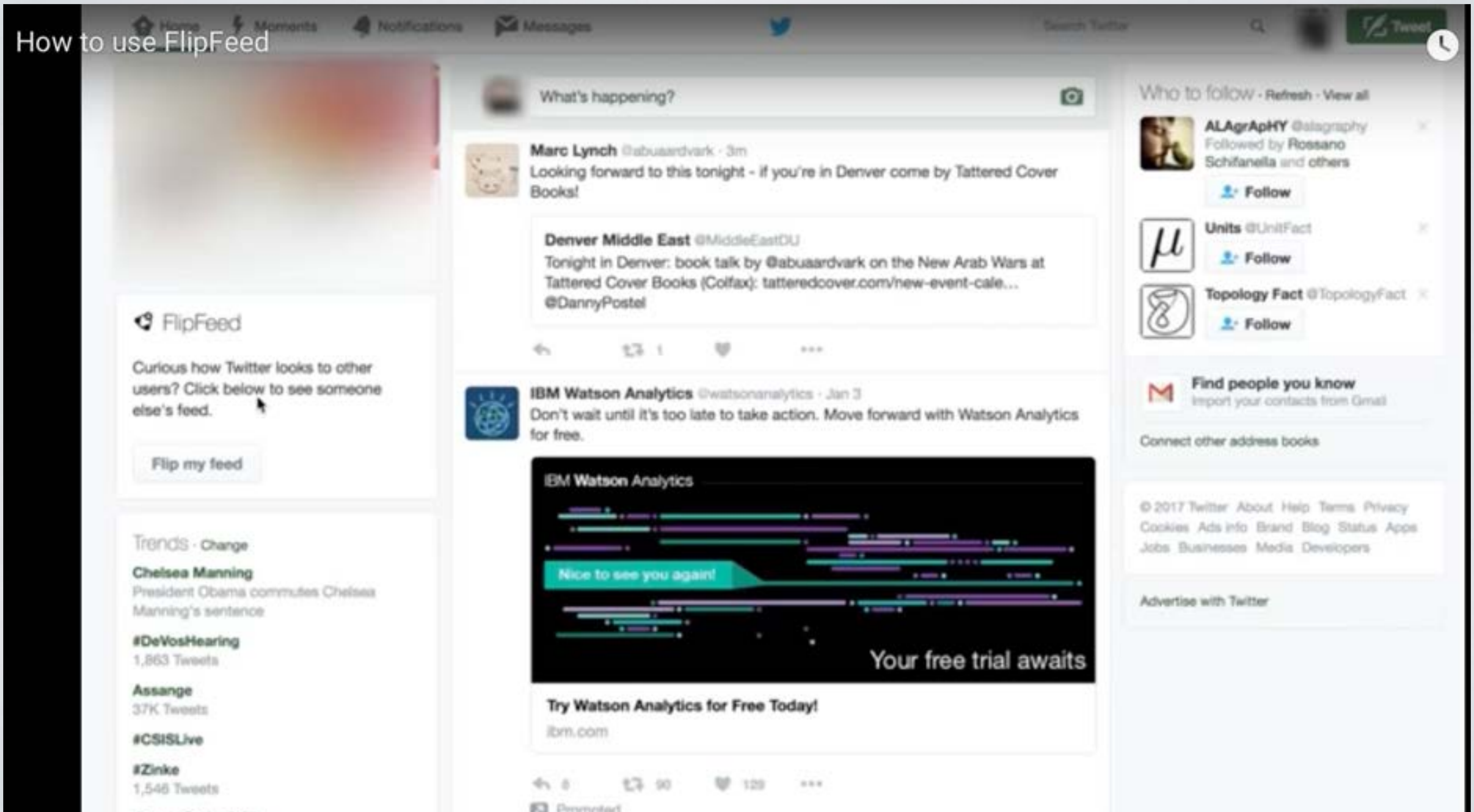
FlipFeed was built by researchers in the Laboratory for Social Machines at the MIT Media Lab to explore how social media platforms can be used to mitigate, rather than exacerbate, ideological polarization by helping people explore and empathize with different perspectives.

social machines

mit media lab

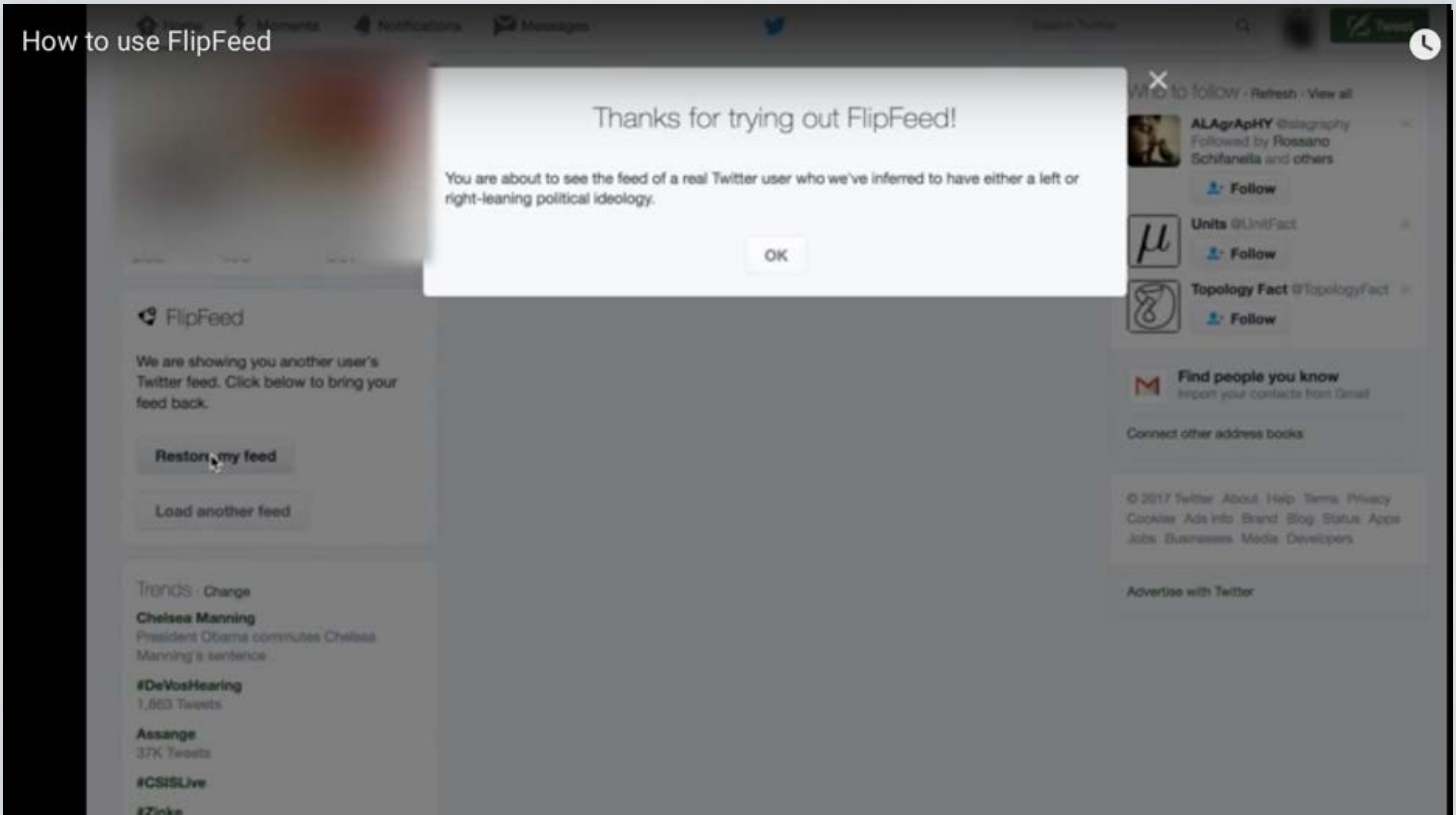
FRAGMENTATION : MICRO-LEVEL VIEWS

FlipFeed



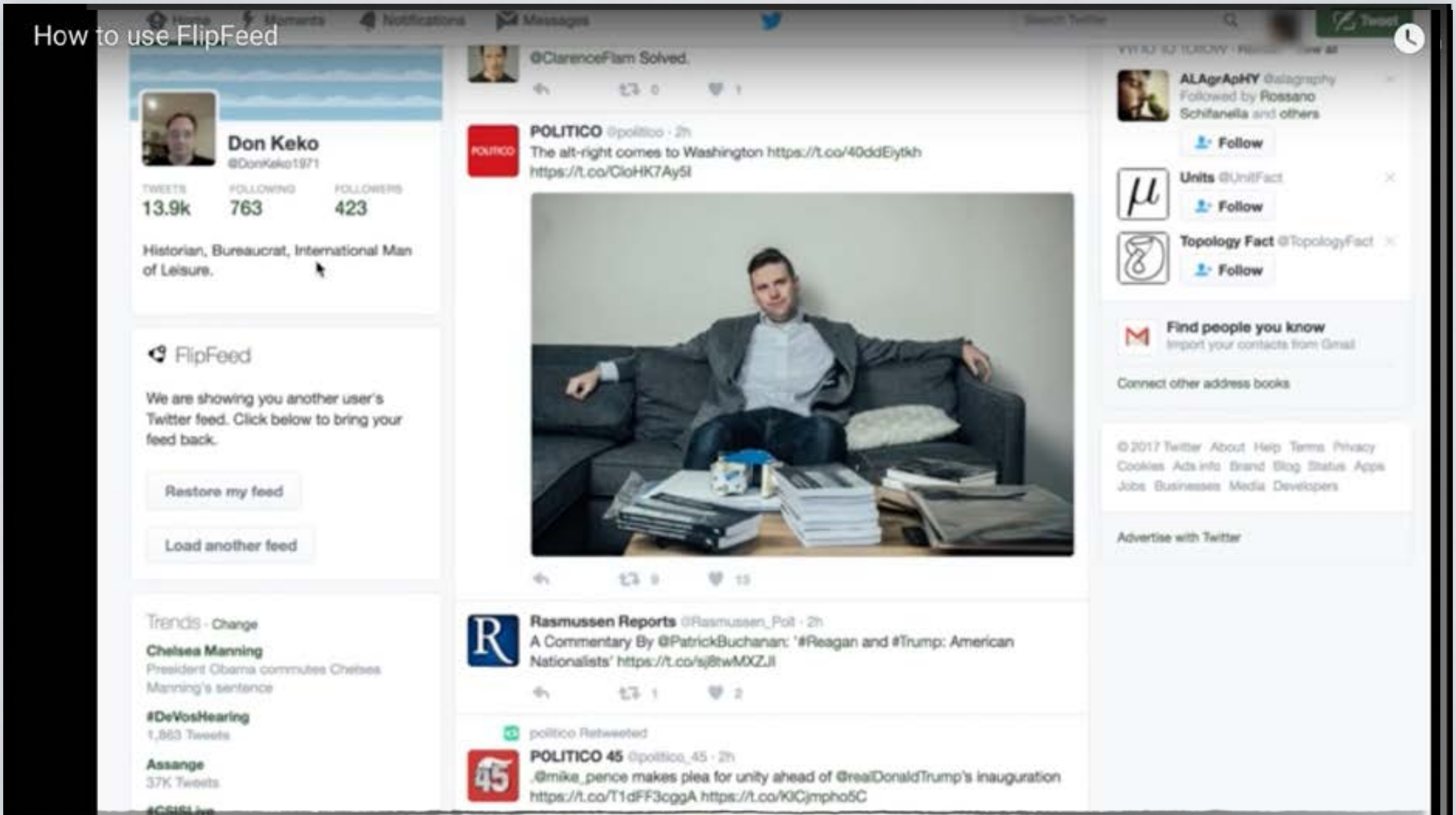
FRAGMENTATION : MICRO-LEVEL VIEWS

FlipFeed



FRAGMENTATION : MICRO-LEVEL VIEWS

FlipFeed



My political bubble

Made from my friends list using PolitEcho.org

What is PolitEcho?

PolitEcho shows you the political biases of your Facebook friends and news feed. The app assigns each of your friends a score based on our prediction of their political leanings then displays a graph of your friend list. Then it calculates the political bias in the content of your news feed and compares it with the bias of your friends list to highlight possible differences between the two.

How do I use it?

PolitEcho is a Google Chrome extension. Click on the button above to find PolitEcho's Google Chrome store page and click the "Add to Chrome" button to install PolitEcho on your computer. Once it is installed, simply log into Facebook and click on the PolitEcho icon in your navigation bar to get started.

FRAGMENTATION : MACRO-LEVEL VIEWS

Adamic & Glance, 2004

"Divided They Blog"

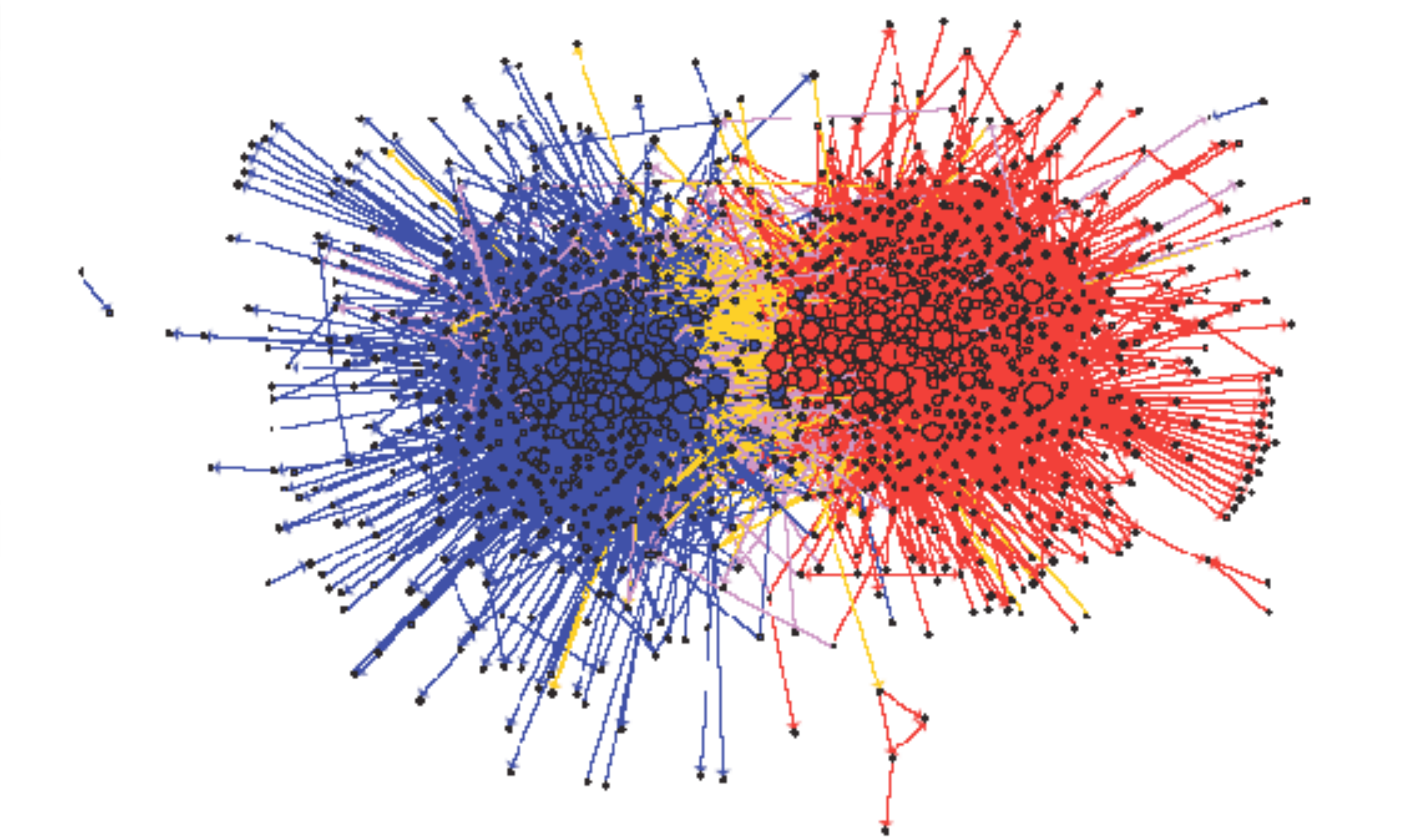


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

~1.5k blogs, 50/50 blue/red

FRAGMENTATION : MACRO-LEVEL VIEWS

Adamic & Glance, 2004

"Divided They Blog"

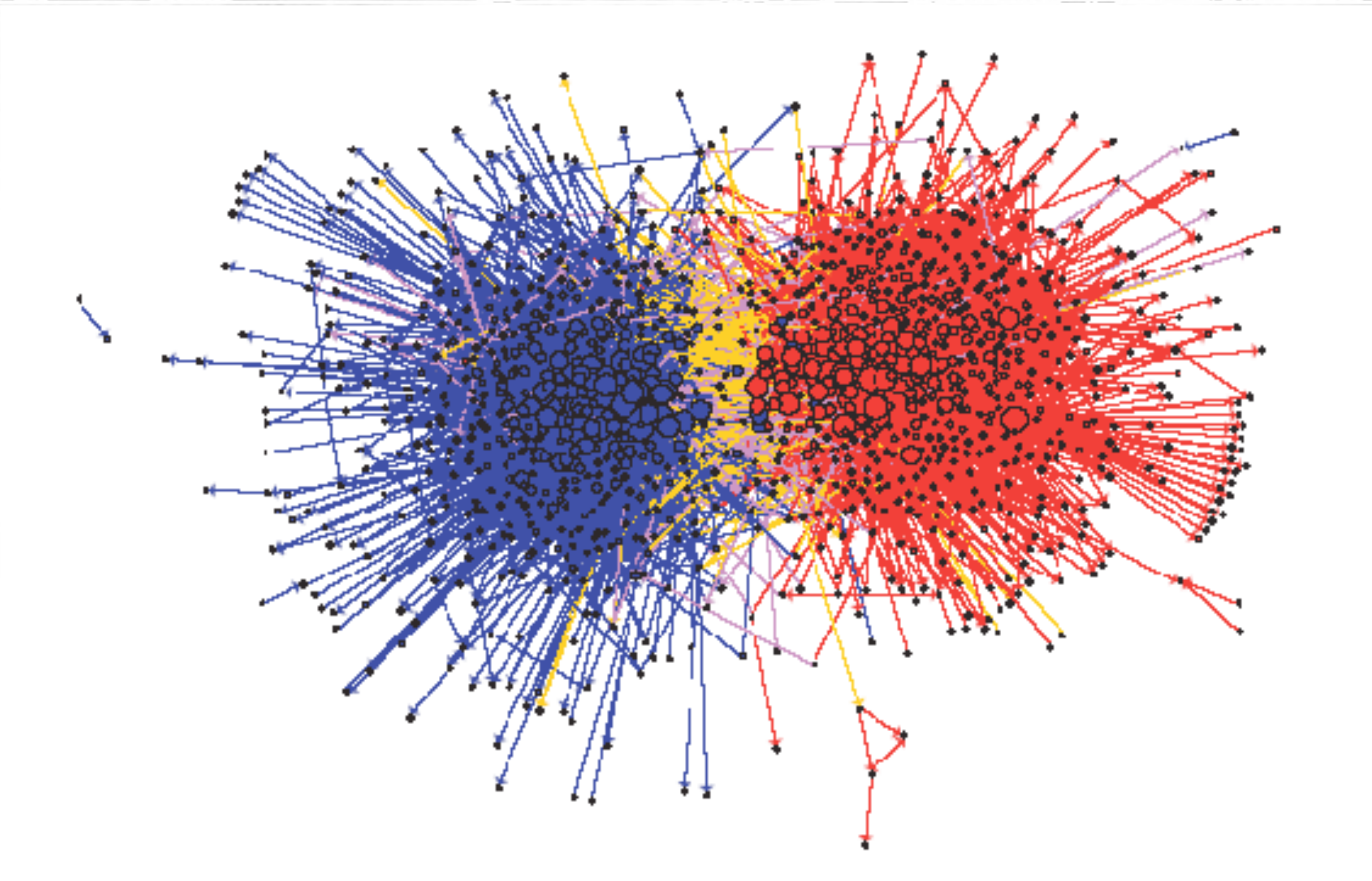
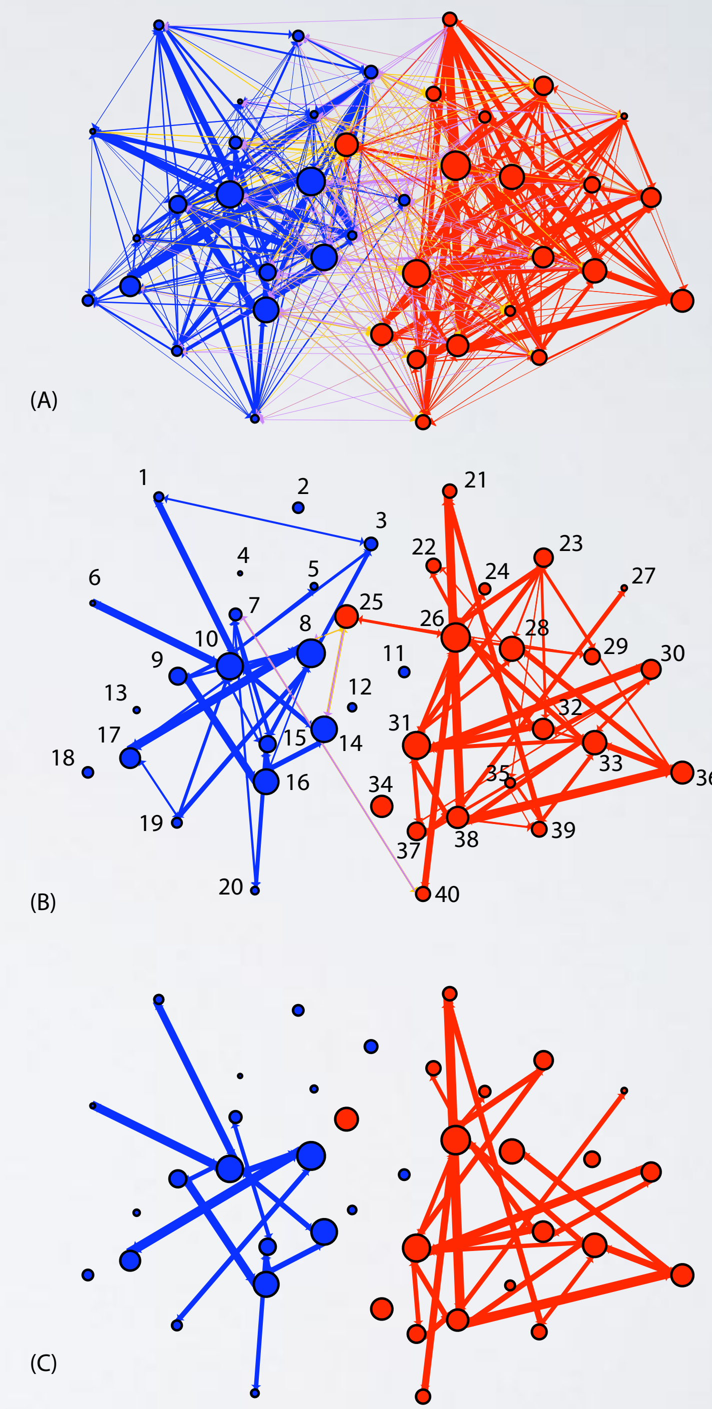


Figure 1: Community structure of political blogs (expanded set), shown using utilizing a GEM layout [11] in the GUESS[3] visualization and analysis tool. The colors reflect political orientation, red for conservative, and blue for liberal. Orange links go from liberal to conservative, and purple ones from conservative to liberal. The size of each blog reflects the number of other blogs that link to it.

~1.5k blogs, 50/50 blue/red

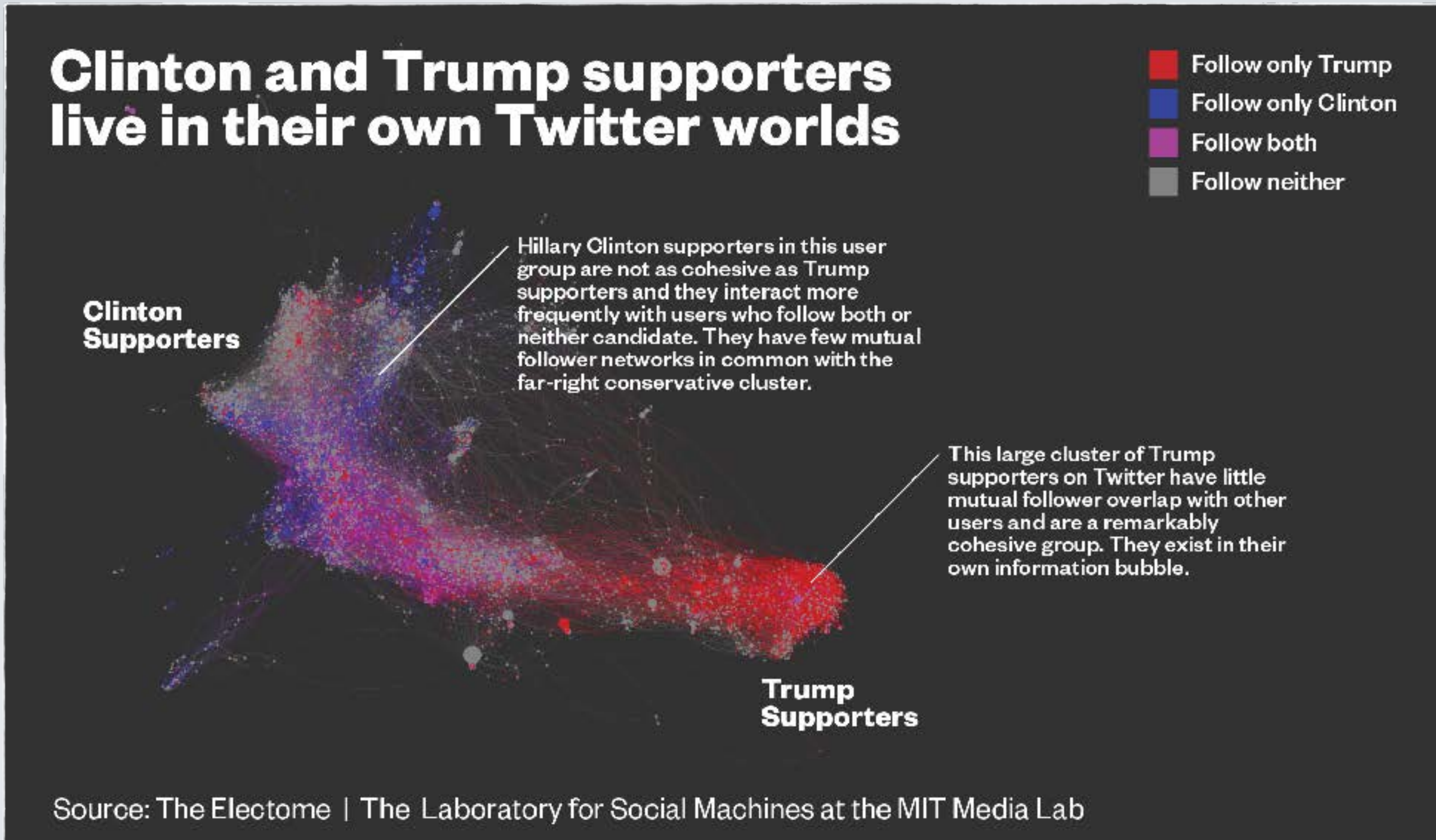


1 Digbys Blog
2 James Walcott
3 Pandagon
4 blog.johnkerry.com
5 Oliver Willis
6 America Blog
7 Crooked Timber
8 Daily Kos
9 American Prospect
10 Eschaton
11 Wonkette
12 Talk Left
13 Political Wire
14 Talking Points Memo
15 Matthew Yglesias
16 Washington Monthly
17 MyDD
18 Juan Cole
19 Left Coaster
20 Bradford DeLong

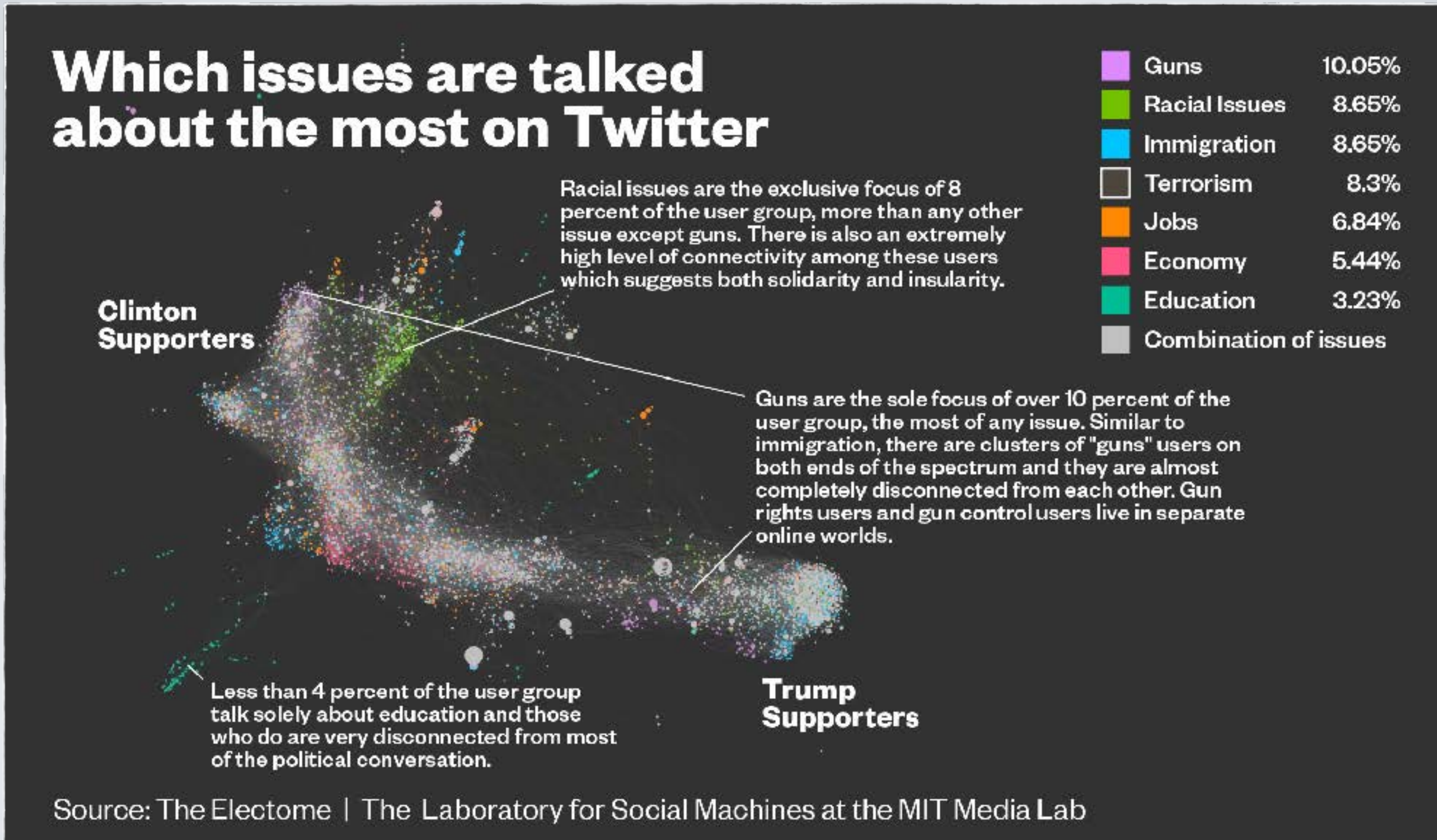
21 JawaReport
22 Voka Pundit
23 Roger L Simon
24 Tim Blair
25 Andrew Sullivan
26 Instapundit
27 Blogs for Bush
28 Little Green Footballs
29 Belmont Club
30 Captain's Quarters
31 Powerline
32 Hugh Hewitt
33 INDC Journal
34 Real Clear Politics
35 Winds of Change
36 Allahpundit
37 Michelle Malkin
38 WizBang
39 Dean's World
40 Volokh

Figure 3: Aggregate citation behavior prior to the 2004 election. Blogs are colored according to political orientation, and the size of the circle reflects how many citations from the top 40 the blog has received. The thickness of the line reflects the number of citations between two blogs. (A) All directed edges are shown. (B) Edges having fewer than 5 citations in either or both directions are removed. (C) Edges having fewer than 25 combined citations are removed.

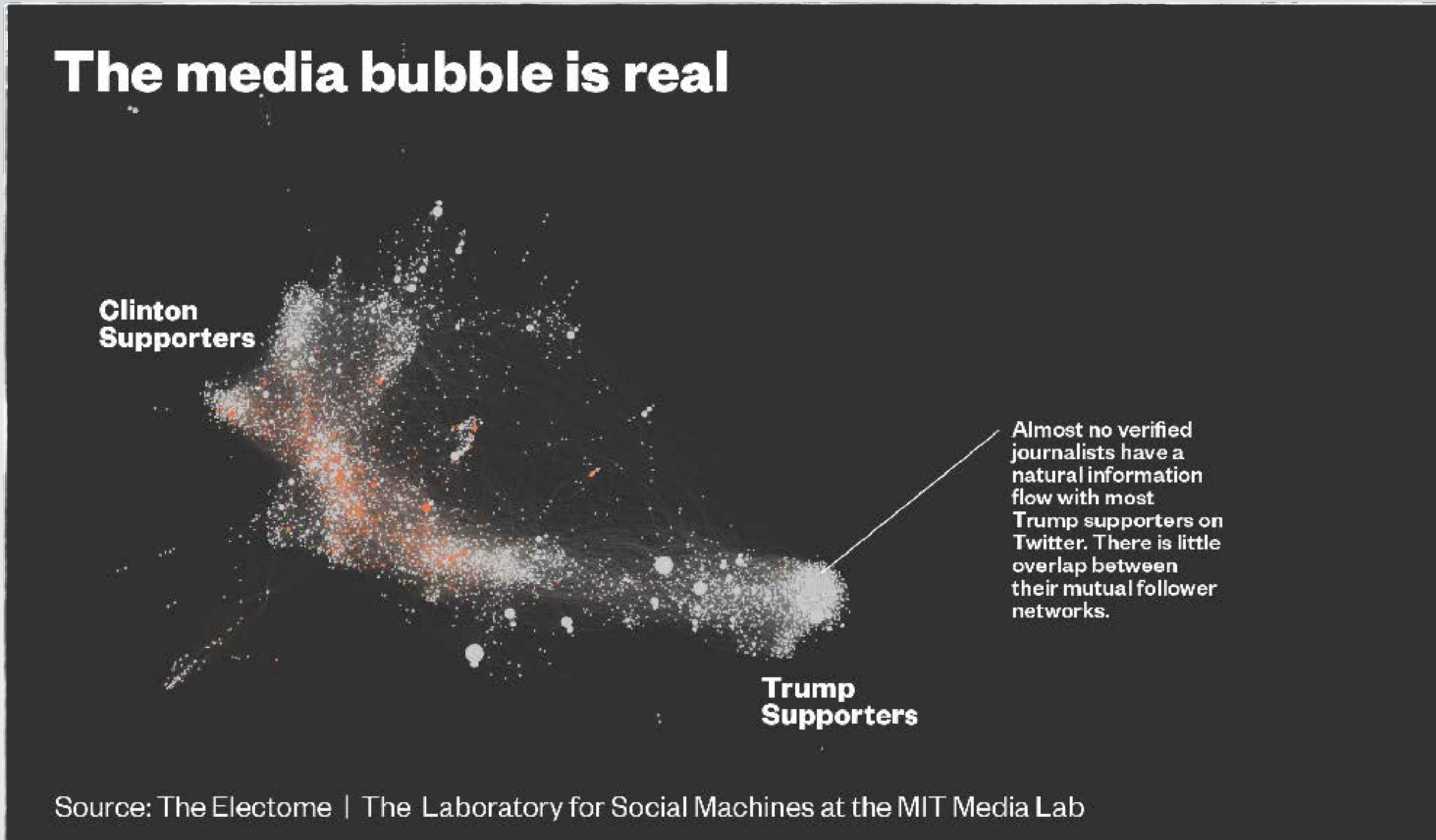
Thompson, 2016



Thompson, 2016

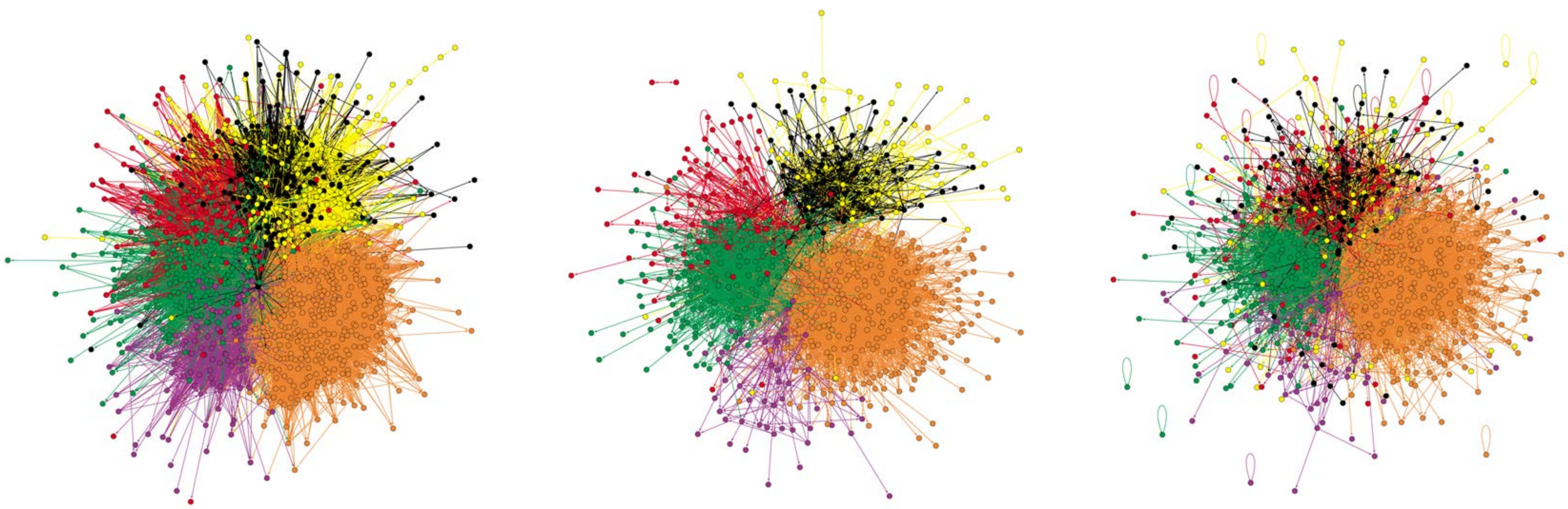


Thompson, 2016



Lietz, Wagner, Bleier, Strohmaier, 2014

When Politicians Talk: Assessing Online Conversational Practices of Political Parties on Twitter



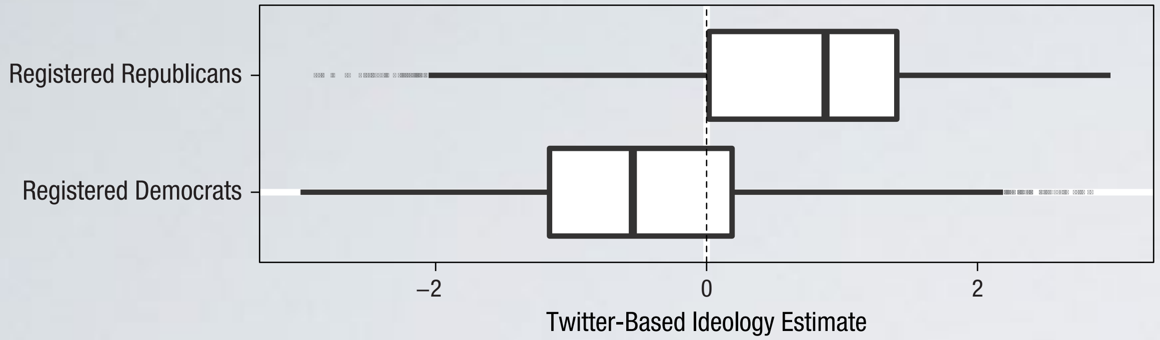
(a) Following (H=0.83) (b) Retweeting (H=0.90) (c) Mentioning (H=0.79)

Figure 1: **Examples of online conversational practices on Twitter**: Structures of the aggregate following, retweeting, and mentioning networks of German politicians from 9 weeks before to 4 weeks after the federal election 2013. The vertices in the networks correspond to user handles and are color-coded by party affiliation (colors given in Table 1). Arcs correspond to following/retweeting/mentioning relationships and are colored by sender. Structural differences between different practices can be observed: For example, homophily H effects are lower in the mentioning network (0.79) than in the following (0.83) and retweeting (0.90) networks. CDU/CSU and FDP, which formed the last government coalition in Germany, are tightly knit in the follow and retweet networks. The Pirates are largely decoupled from a relatively pluralistic mentioning space where all other parties transact. The networks were laid out using the Kamada-Kawai algorithm.

Party	Election Result	Politicians
CDU/CSU	41.5%	158
SPD	25.7%	143
FDP	4.8%	143
Greens	8.4%	178
Left	8.6%	97
Pirates	2.2%	312
Total	**91.2%**	**1,031**

homophily is stronger in retweet and follower networks than in mentions

- *references are more fragmented than conversations*



Barberá, Jost, Nagler, Tucker, Bonneau, 2015

"Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber"

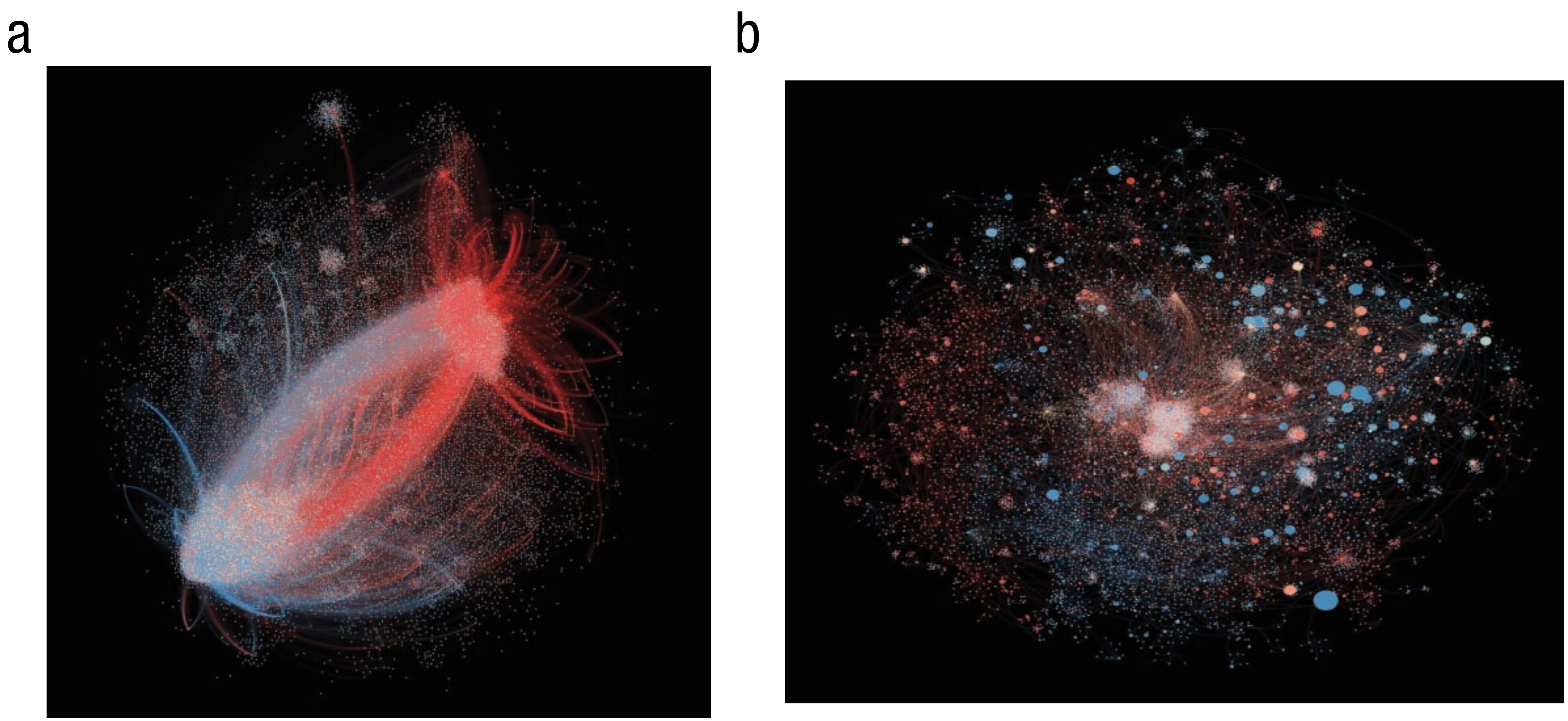
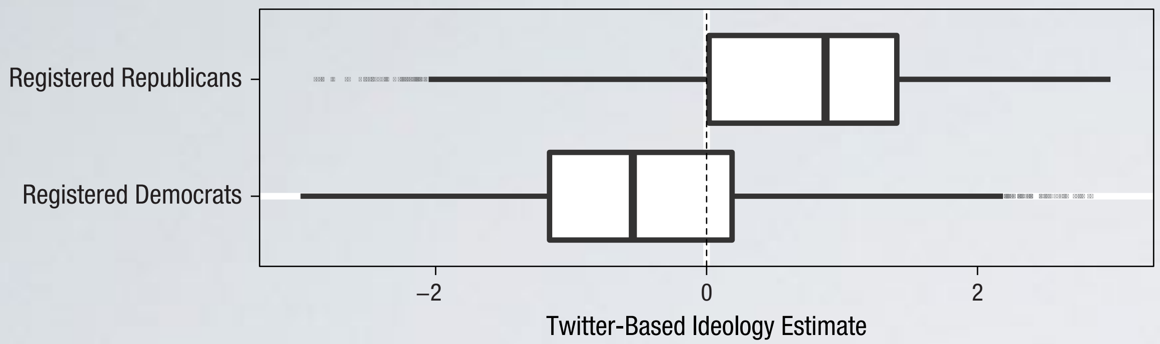
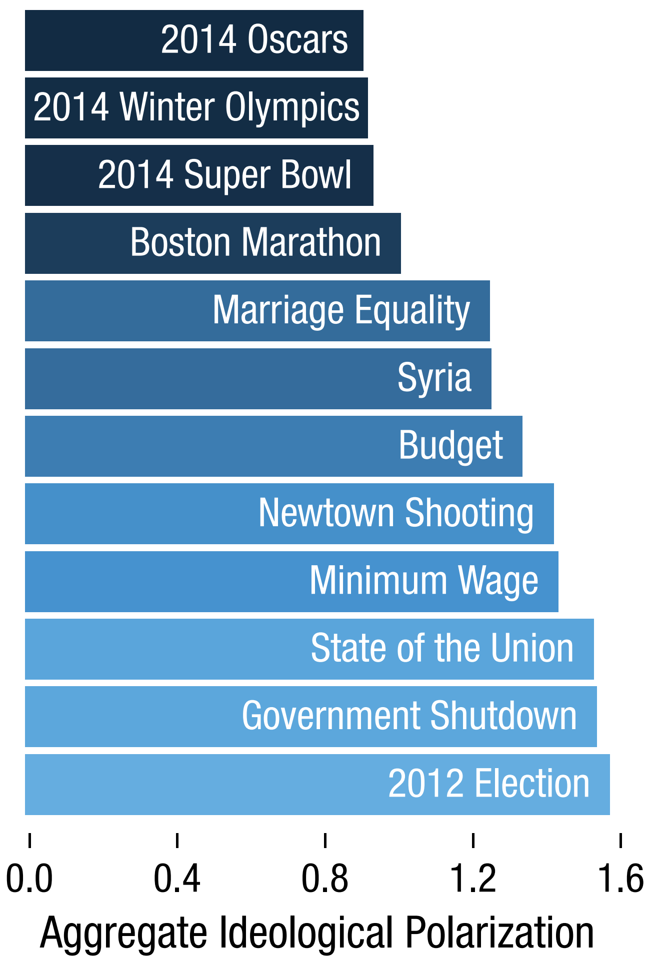


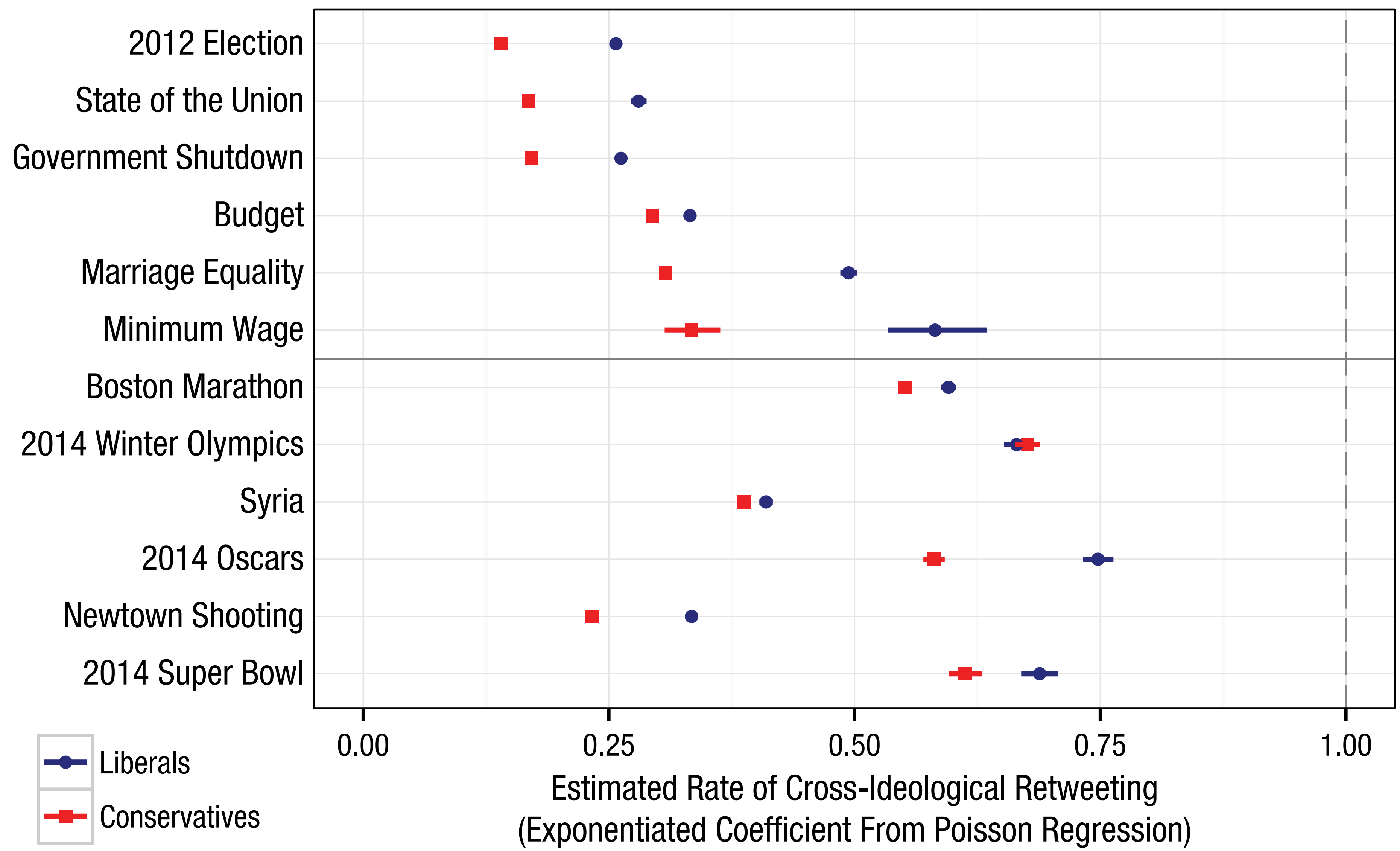
Fig. 3. Additional results on polarization in retweeting behavior. The graphics in (a) and (b), which were created using a force-directed layout algorithm, depict the retweet networks for the tweet collections on the 2012 election and the 2014 Super Bowl. Each node (dot) represents one user (from a random sample, weighted by activity), and each edge (line) represents a retweet. Nodes are colored according to the ideology estimate of the corresponding user, from very conservative (dark red) to very liberal (dark blue). Edges are colored according to the ideology estimate of the user whose tweet was retweeted. White color denotes areas with a large number of nodes whose placement in the figure overlap.



Barberá, Jost, Nagler, Tucker, Bonneau, 2015

"Tweeting From Left to Right: Is Online Political Communication More Than an Echo Chamber"





FRAGMENTATION : MACRO/MICRO

Conover, Ratkiewicz, Francisco, Gonçalves, Flammini, Menczer, 2011

"Political Polarization on Twitter"

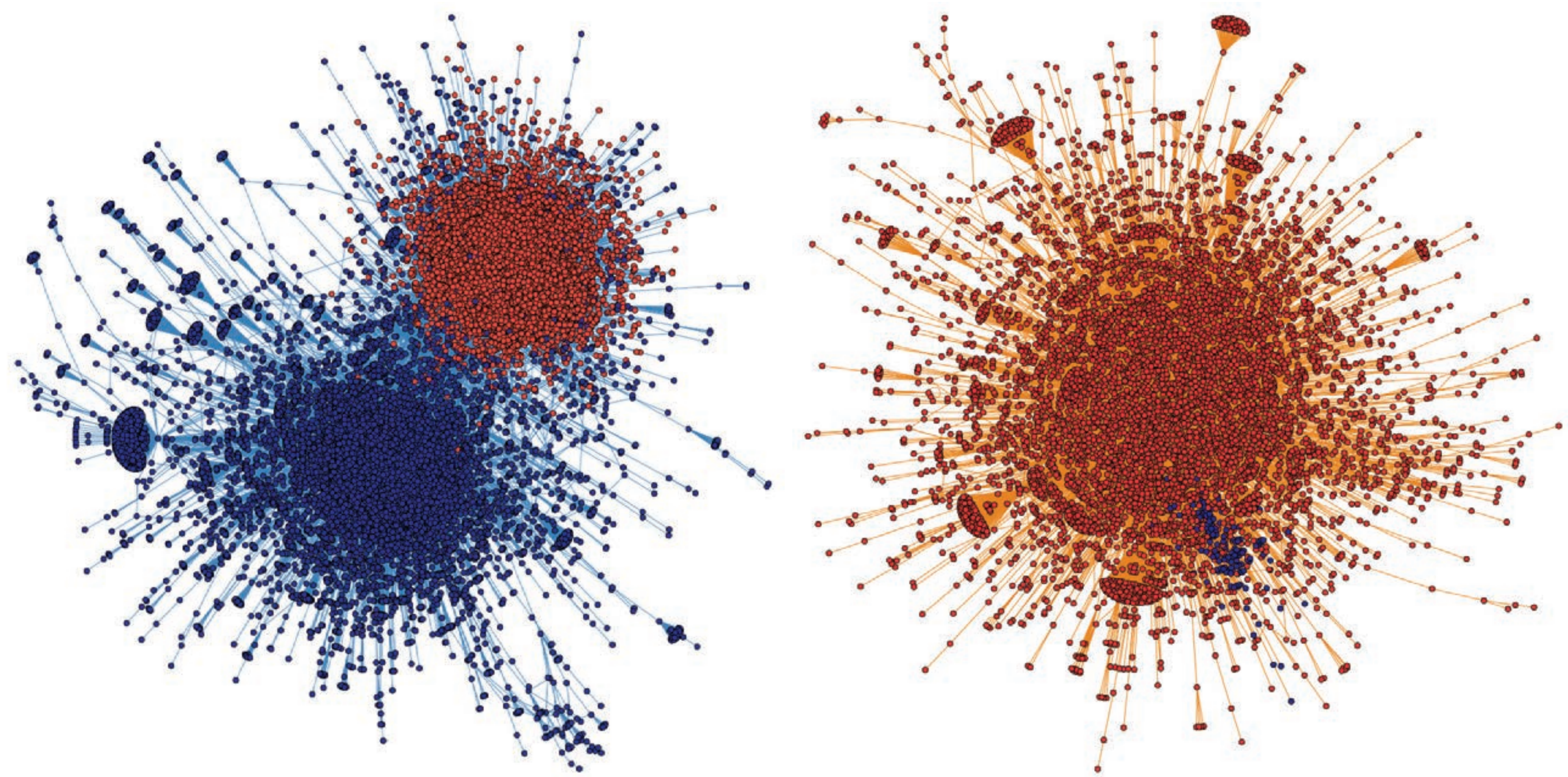


Figure 1: The political retweet (left) and mention (right) networks, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see § 3.1). Community structure is evident in the retweet network, but less so in the mention network. We show in § 3.3 that in the retweet network, the red cluster A is made of 93% right-leaning users, while the blue cluster B is made of 80% left-leaning users.

manual coding of a thousand of users, randomly selected from #p2 ("Progressives 2.0") or #tcot ("Top Conservatives on Twitter")

Table 5: Ratios between observed and expected number of links between users of different political alignments in the mention and retweet networks.

	Mention		Retweet	
	→ Left	→ Right	→ Left	→ Right
Left	1.23	0.68	1.70	0.05
Right	0.77	1.31	0.03	2.32

FRAGMENTATION : MACRO/MICRO

Conover, Ratkiewicz, Francisco, Gonçalves, Flammini, Menczer, 2011

"Political Polarization on Twitter"

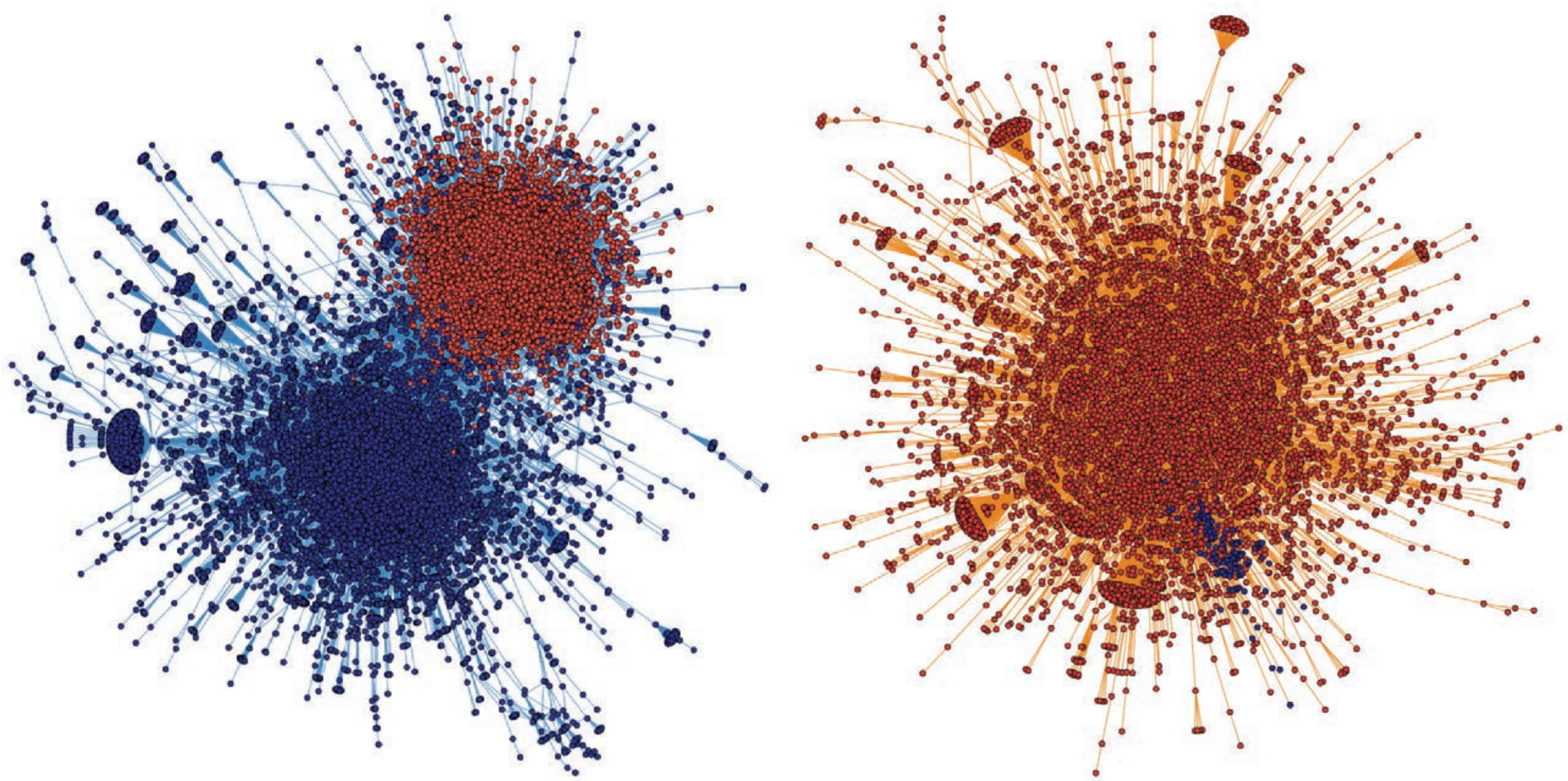


Figure 1: The political retweet (left) and mention (right) networks, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see § 3.1). Community structure is evident in the retweet network, but less so in the mention network. We show in § 3.3 that in the retweet network, the red cluster A is made of 93% right-leaning users, while the blue cluster B is made of 80% left-leaning users.

manual coding of a thousand of users, randomly selected from #p2 ("Progressives 2.0") or #tcot ("Top Conservatives on Twitter")

Table 5: Ratios between observed and expected number of links between users of different political alignments in the mention and retweet networks.

	Mention		Retweet	
	→ Left	→ Right	→ Left	→ Right
Left	1.23	0.68	1.70	0.05
Right	0.77	1.31	0.03	2.32

Table 7: Hashtags in tweets by users across the political spectrum, grouped by valence quintiles.

Far Left	Moderate Left	Center	Moderate Right	Far Right
#healthcare	#aarp #women	#democrats #social	#rangel #waste	#912project #twisters
#judaism #hollywood	#citizensunited	#seniors #dnc	#saveamerica	#gop2112 #israel
#2010elections	#democratic	#budget #political	#american #gold	#foxnews #mediabias
#capitalism #recession	#banksters #energy	#goproud #christian	#repeal #mexico	#constitution
#security #dreamact	#sarahpalin	#media #nobel	#terrorism #gopleader	#patriots #rednov
#publicoption	#progressives		#palin12	#abortion
#topprogs	#stopbeck #iraq			

FRAGMENTATION : MACRO/MICRO

Conover, Ratkiewicz, Francisco, Gonçalves, Flammini, Menczer, 2011

"Political Polarization on Twitter"

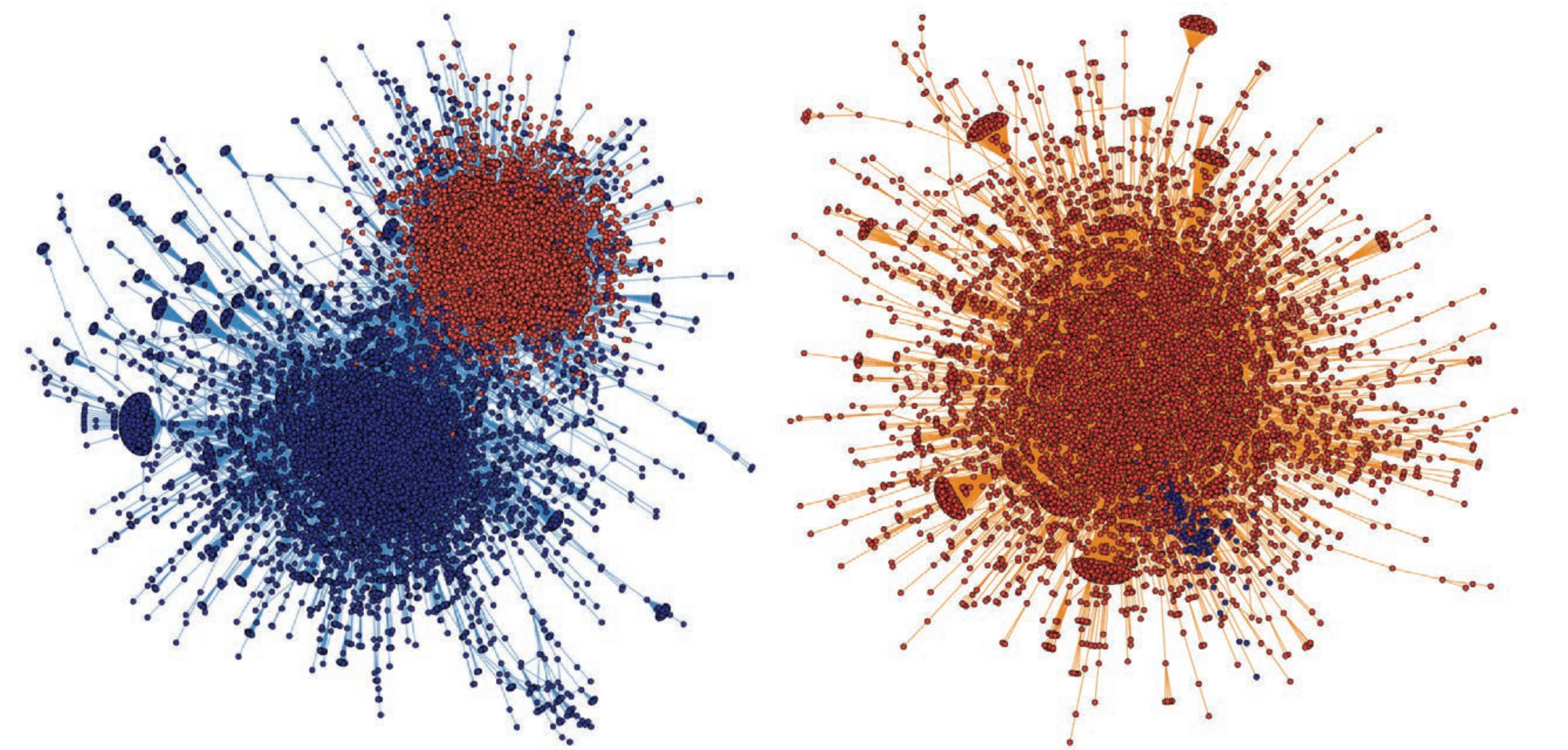


Figure 1: The political retweet (left) and mention (right) networks, laid out using a force-directed algorithm. Node colors reflect cluster assignments (see § 3.1). Community structure is evident in the retweet network, but less so in the mention network. We show in § 3.3 that in the retweet network, the red cluster A is made of 93% right-leaning users, while the blue cluster B is made of 80% left-leaning users.

manual coding of a thousand of users, randomly selected from #p2 ("Progressives 2.0") or #tcot ("Top Conservatives on Twitter")

Table 5: Ratios between observed and expected number of links between users of different political alignments in the mention and retweet networks.

	Mention		Retweet	
	→ Left	→ Right	→ Left	→ Right
Left	1.23	0.68	1.70	0.05
Right	0.77	1.31	0.03	2.32

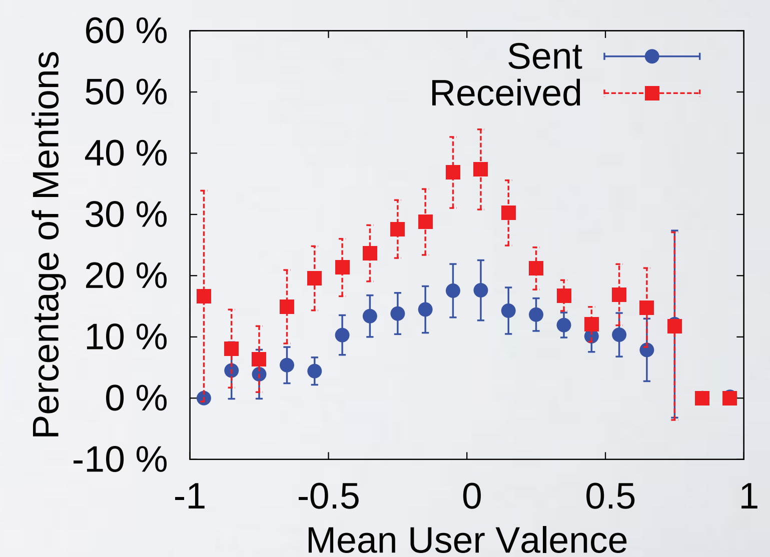


Figure 3: Proportion of mentions a user sends and receives to and from ideologically-opposed users relative to her valence. Points represent binned averages. Error bars denote 95% confidence intervals.

Table 7: Hashtags in tweets by users across the political spectrum, grouped by valence quintiles.

Far Left	Moderate Left	Center	Moderate Right	Far Right
#healthcare	#aarp #women	#democrats #social	#rangel #waste	#912project #twisters
#judaism #hollywood	#citizensunited	#seniors #dnc	#saveamerica	#gop2112 #israel
#2010elections	#democratic	#budget #political	#american #gold	#foxnews #mediabias
#capitalism #recession	#banksters #energy	#goproud #christian	#repeal #mexico	#constitution
#security #dreamact	#sarahpalin	#media #nobel	#terrorism #gopleader	#patriots #rednov
#publicoption	#progressives		#palin12	#abortion
#topprogs	#stopbeck #iraq			

Hashtag	# Tweets	Retweet graph		Follow graph		Description and collection period (2015)
		$\|V\|$	$\|E\|$	$\|V\|$	$\|E\|$	
#beefban	422 908	21 590	30 180	9525	204 332	Government of India bans beef, Mar 2–5
#nemtsov	371 732	43 114	77 330	17 717	155 904	Death of Boris Nemtsov, Feb 28–Mar 2
#netanyahuspeech	1 196 215	122 884	280 375	49 081	2 009 277	Netanyahu's speech at U.S. Congress, Mar 3–5
#russia_march	317 885	10 883	17 662	4844	42 553	Protests after death of Boris Nemtsov ("march"), Mar 1–2
#indiasdaughter	776 109	68 608	144 935	38 302	131 566	Controversial Indian documentary, Mar 1–5
#baltimoreriots	1 989 360	289 483	432 621	214 552	690 944	Riots in Baltimore after police kills a black man, Apr 28–30
#indiana	972 585	43 252	74 214	21 909	880 814	Indiana pizzeria refuses to cater gay wedding, Apr 2–5
#ukraine	514 074	50 191	91 764	31 225	286 603	Ukraine conflict, Feb 27–Mar 2
#gunsense	1 022 541	30 096	58 514	17 335	841 466	Gun violence in U.S., Jun 1–30
#leadersdebate	2 099 478	54 102	136 290	22 498	1 211 956	Debate during the U.K. national elections, May 3
#sxsw	343 652	9304	11 003	4558	91 356	SXSW conference, Mar 13–22
#1dfamheretostay	501 960	15 292	26 819	3151	20 275	Last OneDirection concert, Mar 27–29
#germanwings	907 510	29 763	39 075	2111	7329	Germanwings flight crash, Mar 24–26
#mothersday	1 798 018	155 599	176 915	2225	14 160	Mother's day, May 8
#nepal	1 297 995	40 579	57 544	4242	42 833	Nepal earthquake, Apr 26–29
#ultralive	364 236	9261	15 544	2113	16 070	Ultra Music Festival, Mar 18–20
#FF	408 326	5401	7646	3899	63 672	Follow Friday, Jun 19
#jurassicworld	724 782	26 407	32 515	4395	31 802	Jurassic World movie, Jun 12-15
#wcw	156 243	10 674	11 809	3264	23 414	Women crush Wednesdays, Jun 17
#nationalkissingday	165 172	4638	4816	790	5927	National kissing day, Jun 19

Fig. 2. Sets of related hashtags for the topics (a) #baltimoreriots and (b) #netanyahuspeech.

Garimella, Mathioudakis, De Francisci Morales, Gionis, 2016

Hashtag	# Tweets	Retweet graph		Follow graph		Description and collection period (2015)
		$\|V\|$	$\|E\|$	$\|V\|$	$\|E\|$	
#beefban	422 908	21 590	30 180	9525	204 332	Government of India bans beef, Mar 2–5
#nemtsov	371 732	43 114	77 330	17 717	155 904	Death of Boris Nemtsov, Feb 28–Mar 2
#netanyahuspeech	1 196 215	122 884	280 375	49 081	2 009 277	Netanyahu's speech at U.S. Congress, Mar 3–5
#russia_march	317 885	10 883	17 662	4844	42 553	Protests after death of Boris Nemtsov ("march"), Mar 1–2
#indiasdaughter	776 109	68 608	144 935	38 302	131 566	Controversial Indian documentary, Mar 1–5
#baltimoreriots	1 989 360	289 483	432 621	214 552	690 944	Riots in Baltimore after police kills a black man, Apr 28–30
#indiana	972 585	43 252	74 214	21 909	880 814	Indiana pizzeria refuses to cater gay wedding, Apr 2–5
#ukraine	514 074	50 191	91 764	31 225	286 603	Ukraine conflict, Feb 27–Mar 2
#gunsense	1 022 541	30 096	58 514	17 335	841 466	Gun violence in U.S., Jun 1–30
#leadersdebate	2 099 478	54 102	136 290	22 498	1 211 956	Debate during the U.K. national elections, May 3
#sxsw	343 652	9304	11 003	4558	91 356	SXSW conference, Mar 13–22
#1dfamheretostay	501 960	15 292	26 819	3151	20 275	Last OneDirection concert, Mar 27–29
#germanwings	907 510	29 763	39 075	2111	7329	Germanwings flight crash, Mar 24–26
#mothersday	1 798 018	155 599	176 915	2225	14 160	Mother's day, May 8
#nepal	1 297 995	40 579	57 544	4242	42 833	Nepal earthquake, Apr 26–29
#ultralive	364 236	9261	15 544	2113	16 070	Ultra Music Festival, Mar 18–20
#FF	408 326	5401	7646	3899	63 672	Follow Friday, Jun 19
#jurassicworld	724 782	26 407	32 515	4395	31 802	Jurassic World movie, Jun 12-15
#wcw	156 243	10 674	11 809	3264	23 414	Women crush Wednesdays, Jun 17
#nationalkissingday	165 172	4638	4816	790	5927	National kissing day, Jun 19

Fig. 2. Sets of related hashtags for the topics (a) #baltimoreriots and (b) #netanyahuspeech.

Random Walk Controversy (RWC)

"

Consider two random walks, one ending in partition X and one ending in partition Y.

RWC is the difference of the probabilities of two events:

(i) both random walks started from the partition they ended in and

(ii) both random walks started in a partition other than the one they ended in."

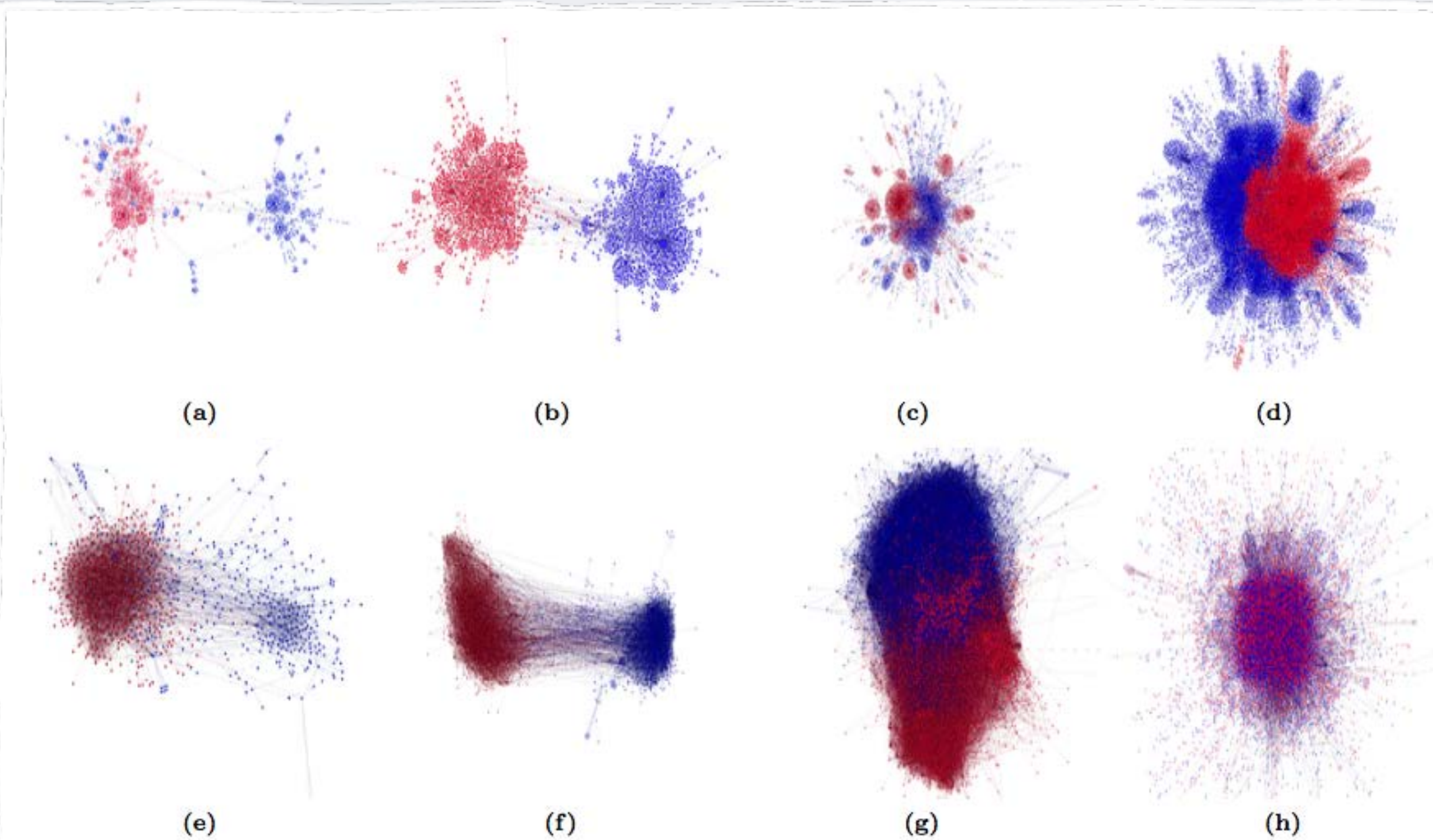


Fig. 3. Sample conversation graphs with retweet (top) and follow (bottom) aspects (visualized using the force-directed layout algorithm in Gephi). The left side is controversial, (a,e) #beefban, (b,f) #russia_march, while the right side is non-controversial, (c,g) #sxsw, (d,h) #germanwings. Only the largest connected component is shown.

Garimella, Mathioudakis, De Francisci Morales, Gionis, 2016

The measure is quantified as:

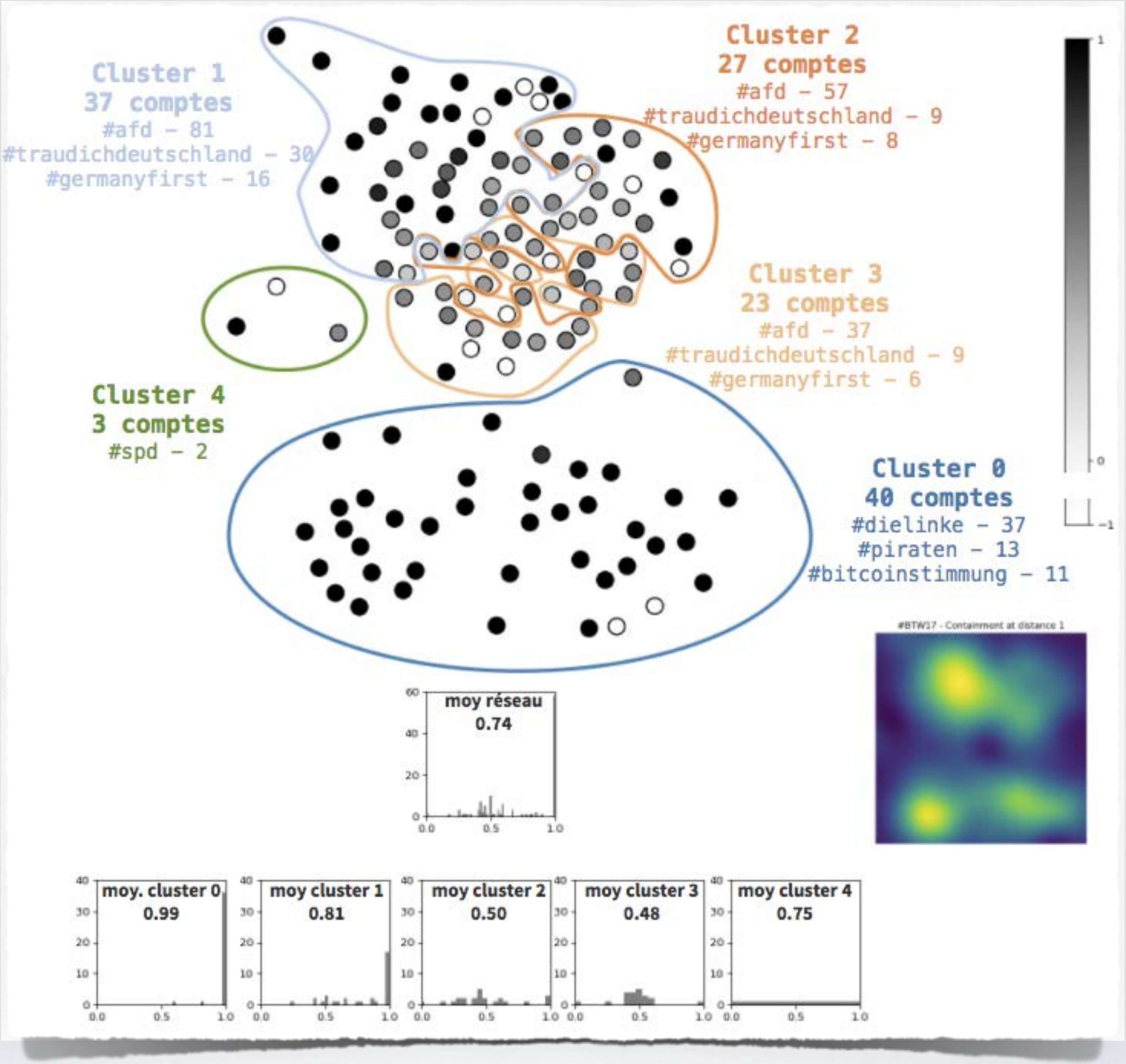
$$RWC = P_{XX}P_{YY} - P_{YX}P_{XY}$$

$P_{AB} = Pr[\text{start in partition } A \mid \text{end in partition } B]$

"

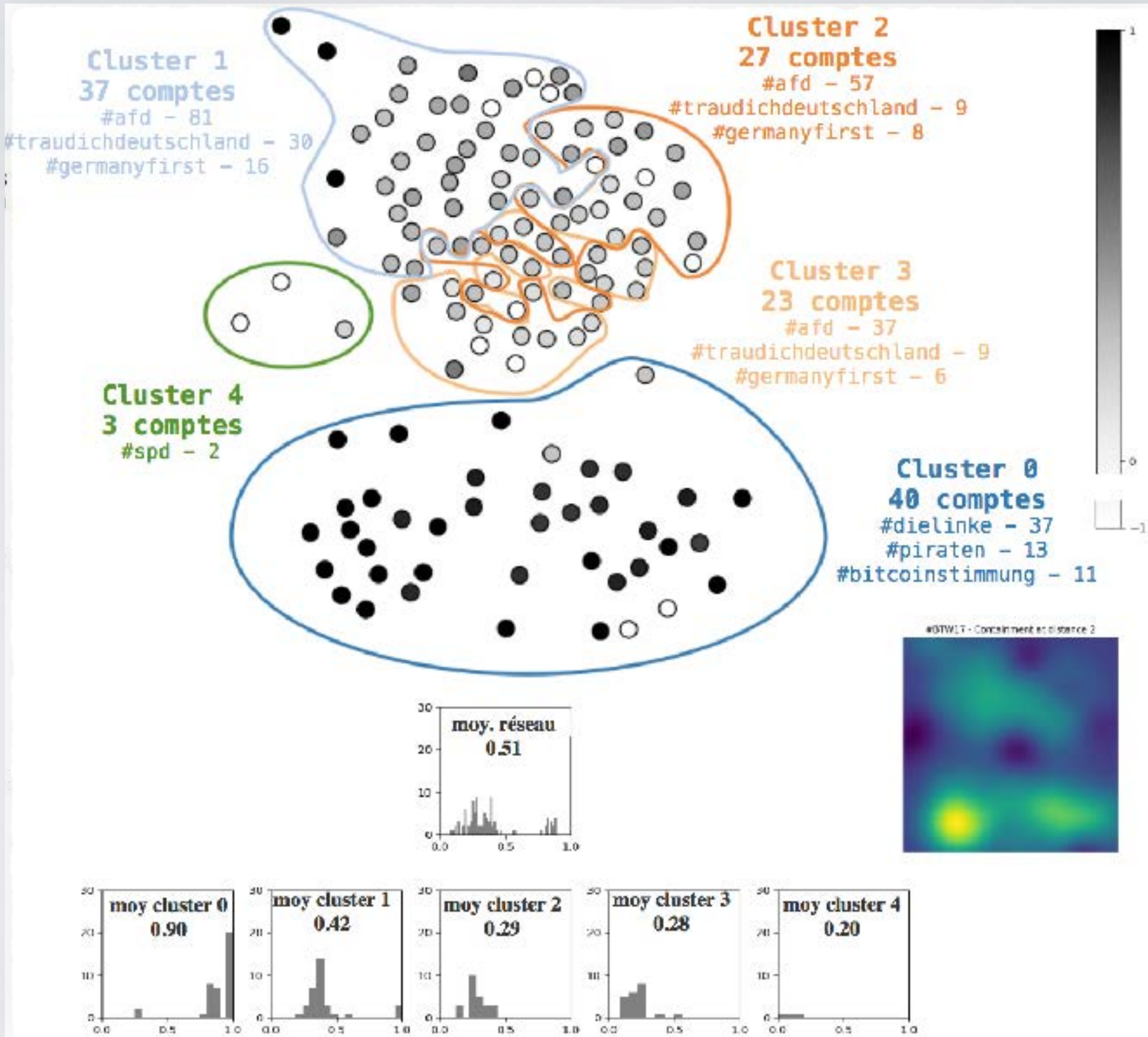
MIXING VIEWS

distance 1



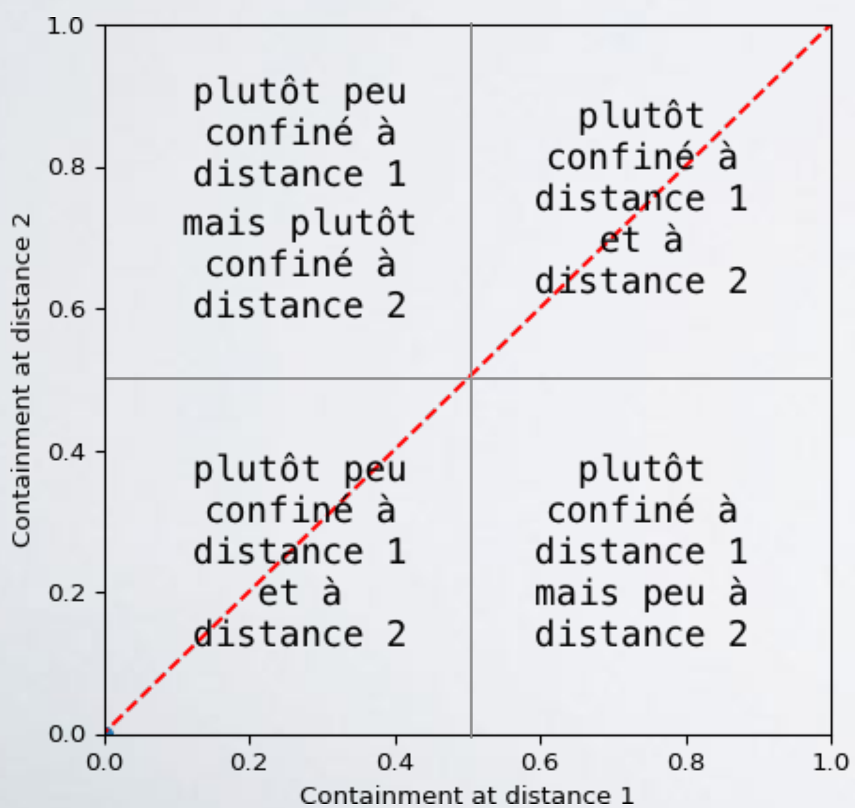
MIXING VIEWS

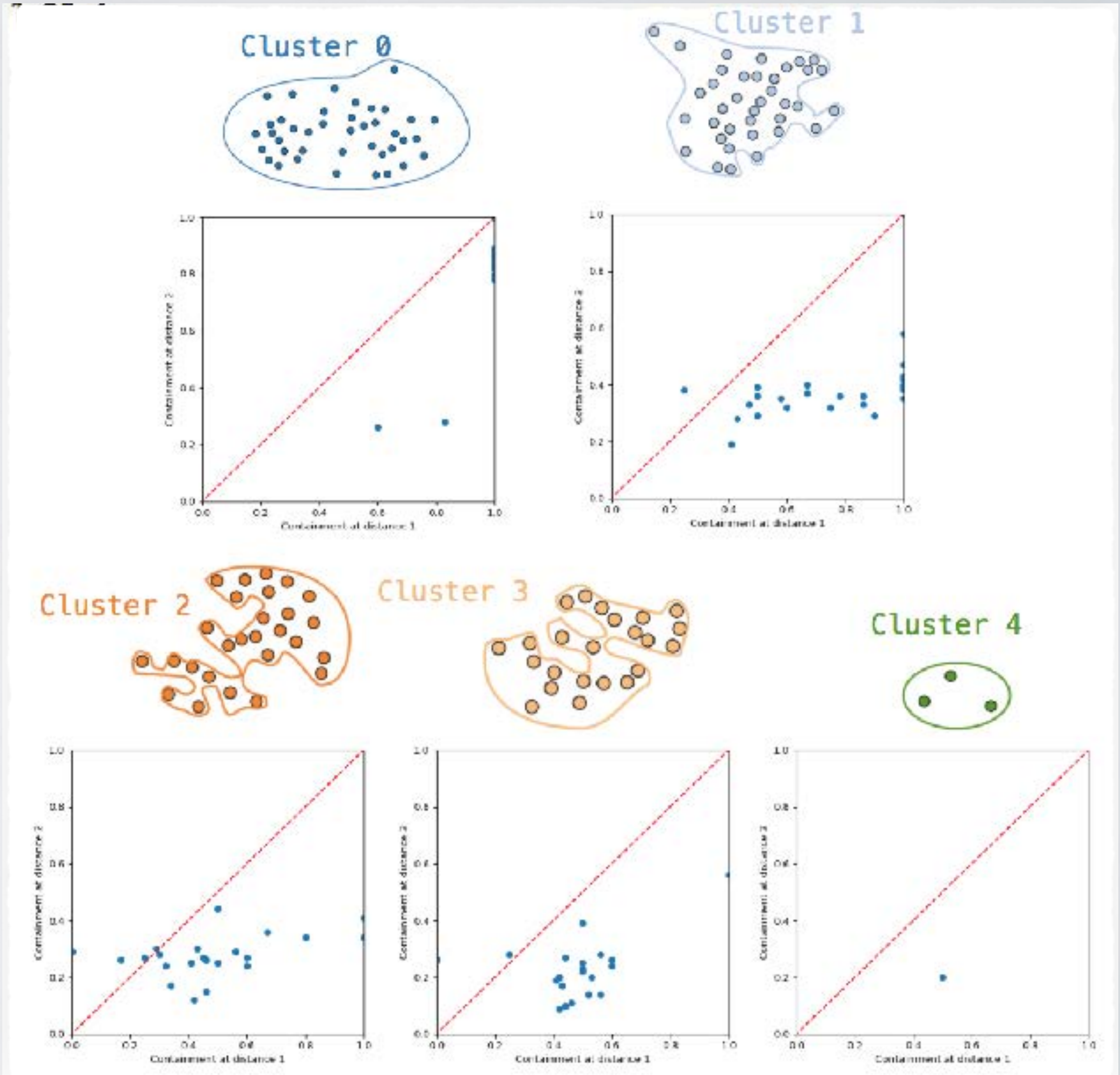
distance 2



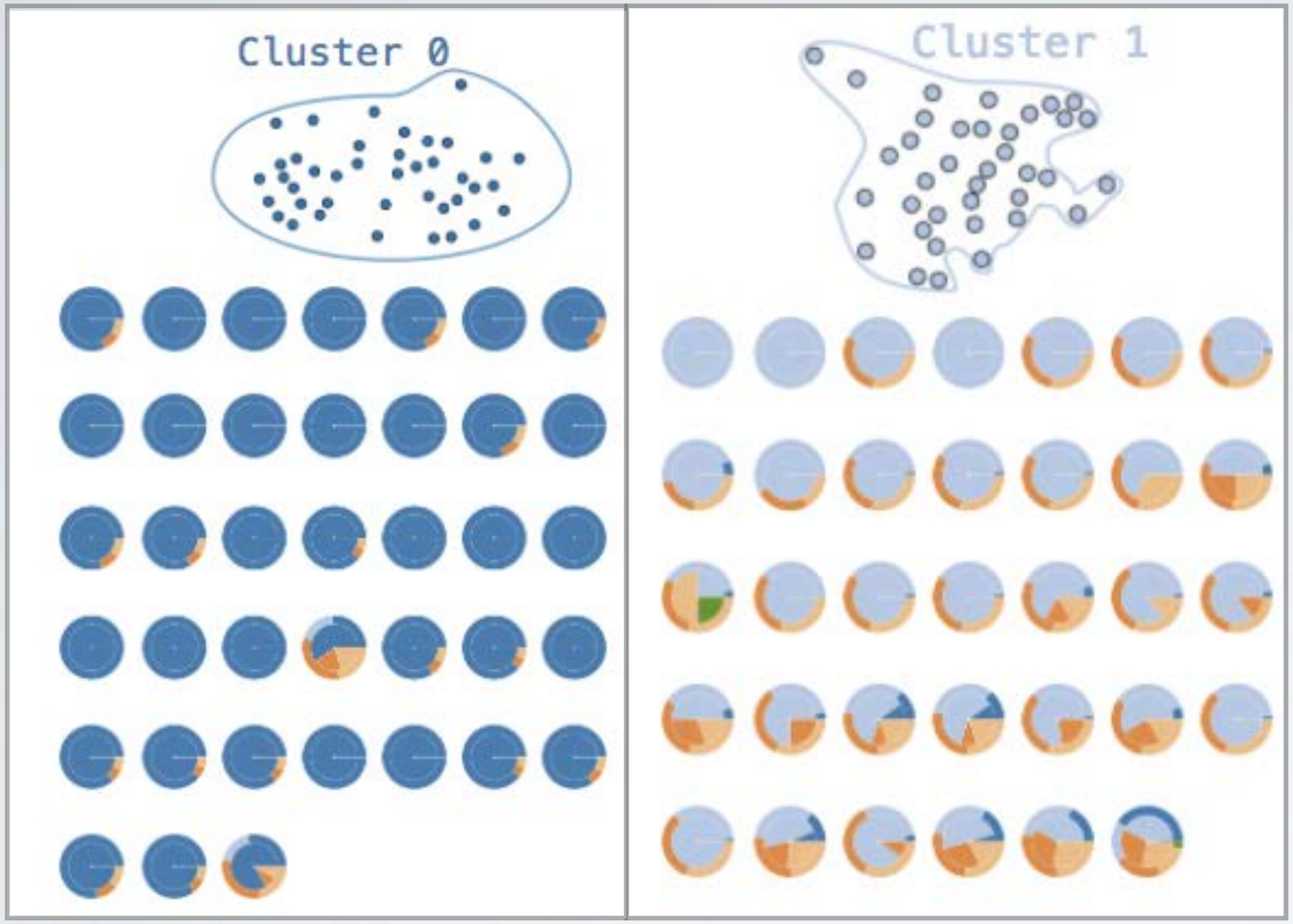
MIXING VIEWS

distance 1 vs. 2

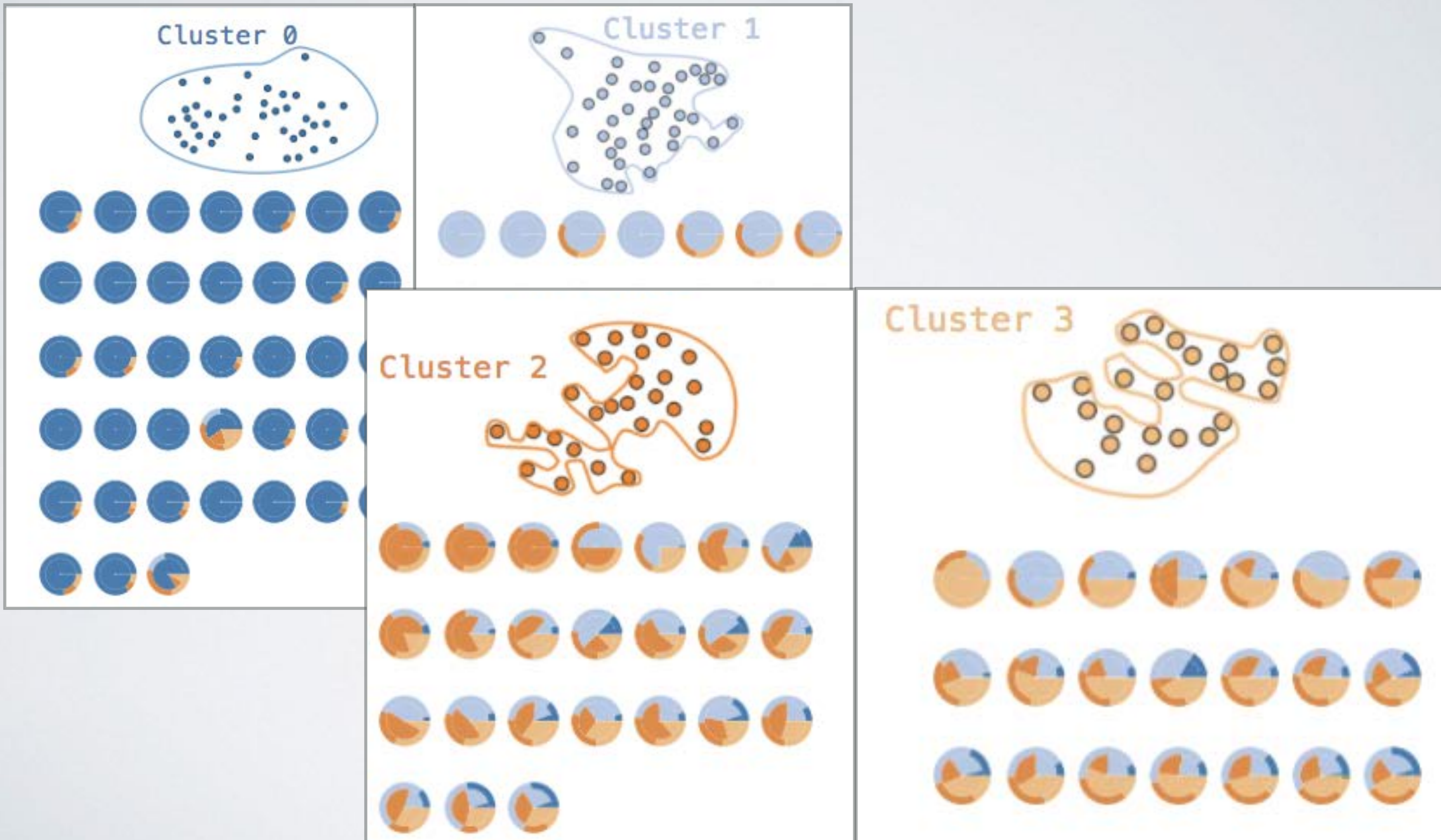




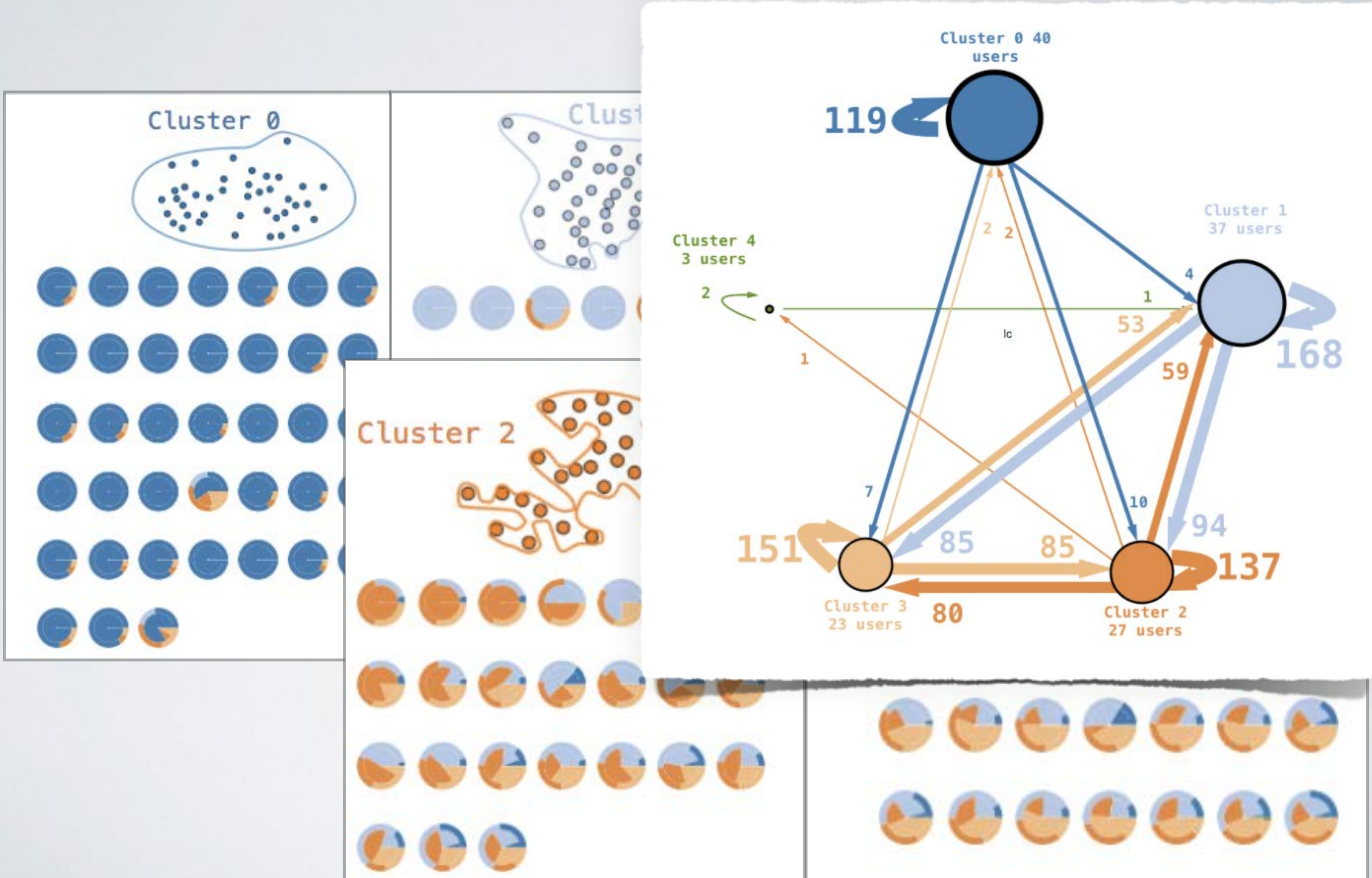
MIXING VIEWS



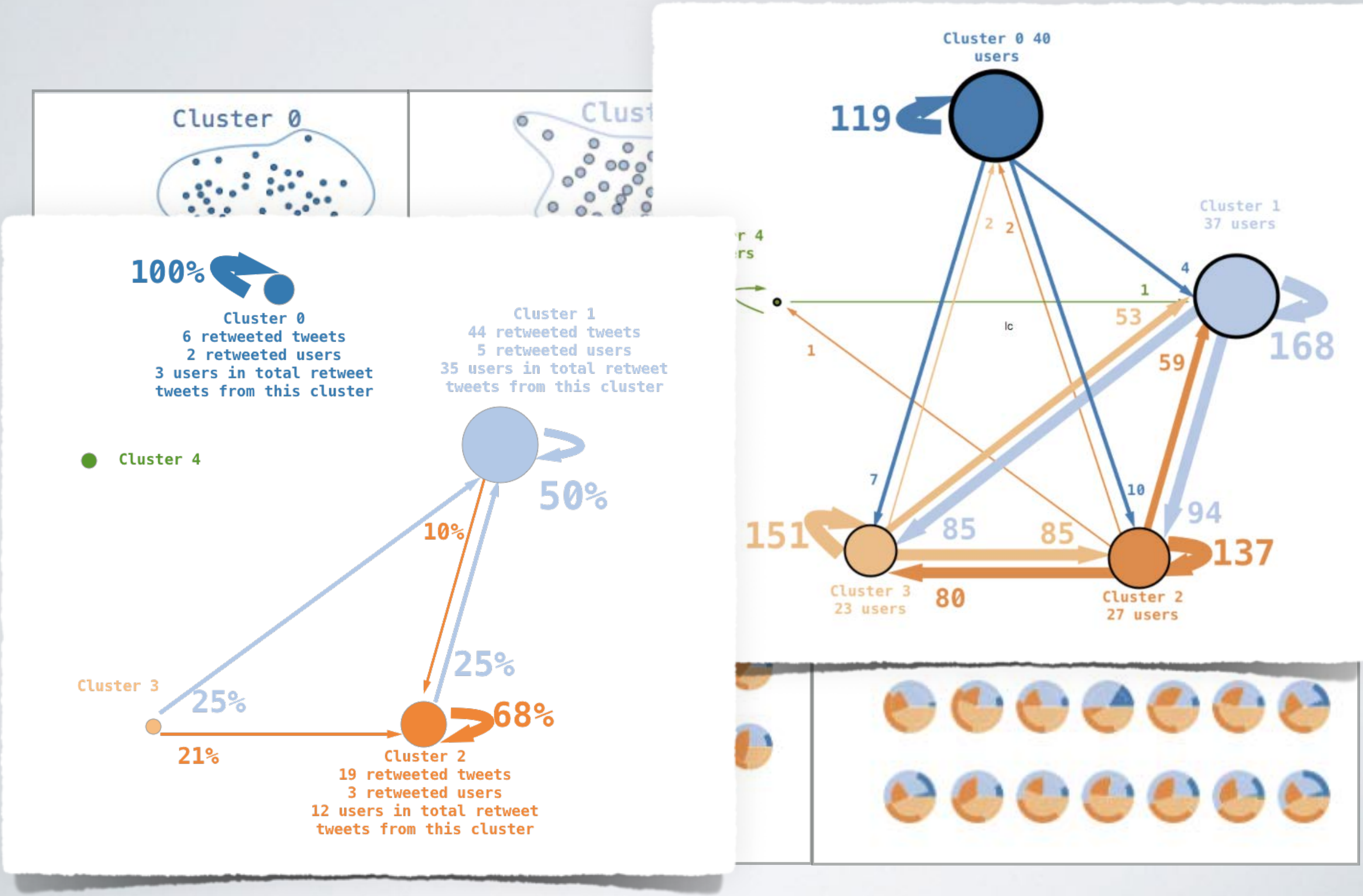
MIXING VIEWS



MIXING VIEWS



MIXING VIEWS



CONTEXT

- selection of the portion of users who are the most active on the topic

  ("#IPCC", strongly unambiguous keyword, around the publication of the latest WG2 & WG3 reports, April 2014)

a notion of
core/periphery
based on
engagement

CONTEXT

- selection of the portion of users who are the most active on the topic

  ("#IPCC", strongly unambiguous keyword, around the publication of the latest WG2 & WG3 reports, April 2014)

35k accounts –> 629 "core" nodes

a notion of
core/periphery
based on
engagement

CONTEXT

- selection of the portion of users who are the most active on the topic

  ("#IPCC", strongly unambiguous keyword, around the publication of the latest WG2 & WG3 reports, April 2014)

  35k accounts –> 629 "core" nodes

- manual coding of the possible alignment of core actors

a notion of
core/periphery
based on
engagement

CONTEXT

- selection of the portion of users who are the most active on the topic

("#IPCC", strongly unambiguous keyword, around the publication of the latest WG2 & WG3 reports, April 2014)

a notion of **core/periphery** based on **engagement**

35k accounts –> 629 "core" nodes

- manual coding of the possible alignment of core actors

Critical tweet
(critical to the scientific basis of human-induced climate change):

IPCC Insider Rejects Global-Warming Report - National Review Online (blog) http://t.co/0bYSPze2Vh

Uncommitted tweet
(gives mainly information related to the IPCC report publication)

Climate scientists meet in Japan for IPCC's WG2 report - Top Asia-Pacific News 25.03.14 http://t.co/vG2OkFqhXE

Supportive
(supporting the scientific basis of human-induced climate change):

#IPCC: Climate change is everywhere. That means we need to take action. Now. http://t.co/eItB60qMiR

CONTEXT

- selection of the portion of users who are the most active on the topic

("#IPCC", strongly unambiguous keyword, around the publication of the latest WG2 & WG3 reports, April 2014)

a notion of **core/periphery** based on **engagement**

35k accounts –> 629 "core" nodes

- manual coding of the possible alignment of core actors

Critical (C)	Supportive (S)	Uncommitted (U)
60 *9.5%*	229 *36.4%*	340 *54.1%*

Critical tweet
(critical to the scientific basis of human-induced climate change):

IPCC Insider Rejects Global-Warming Report - National Review Online (blog) http://t.co/ObYSPze2Vh

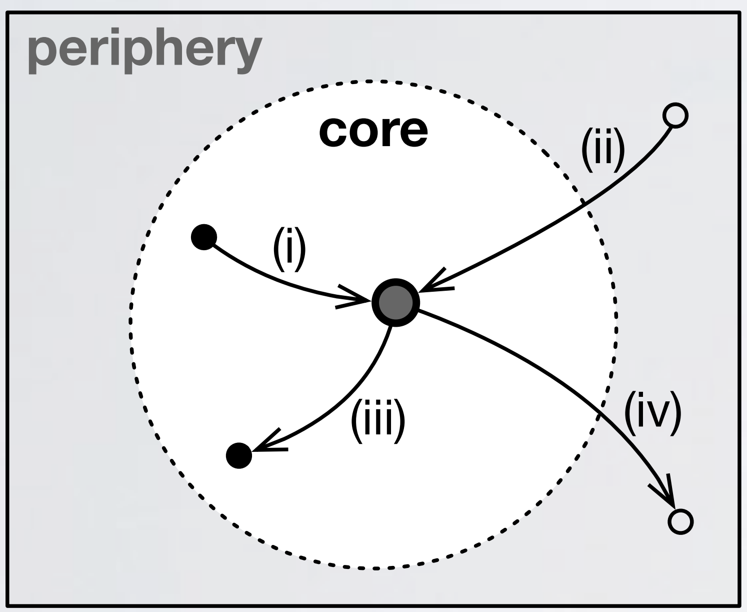
Uncommitted tweet
(gives mainly information related to the IPCC report publication)

Climate scientists meet in Japan for IPCC's WG2 report - Top Asia-Pacific News 25.03.14 http://t.co/vG2OkFqhXE

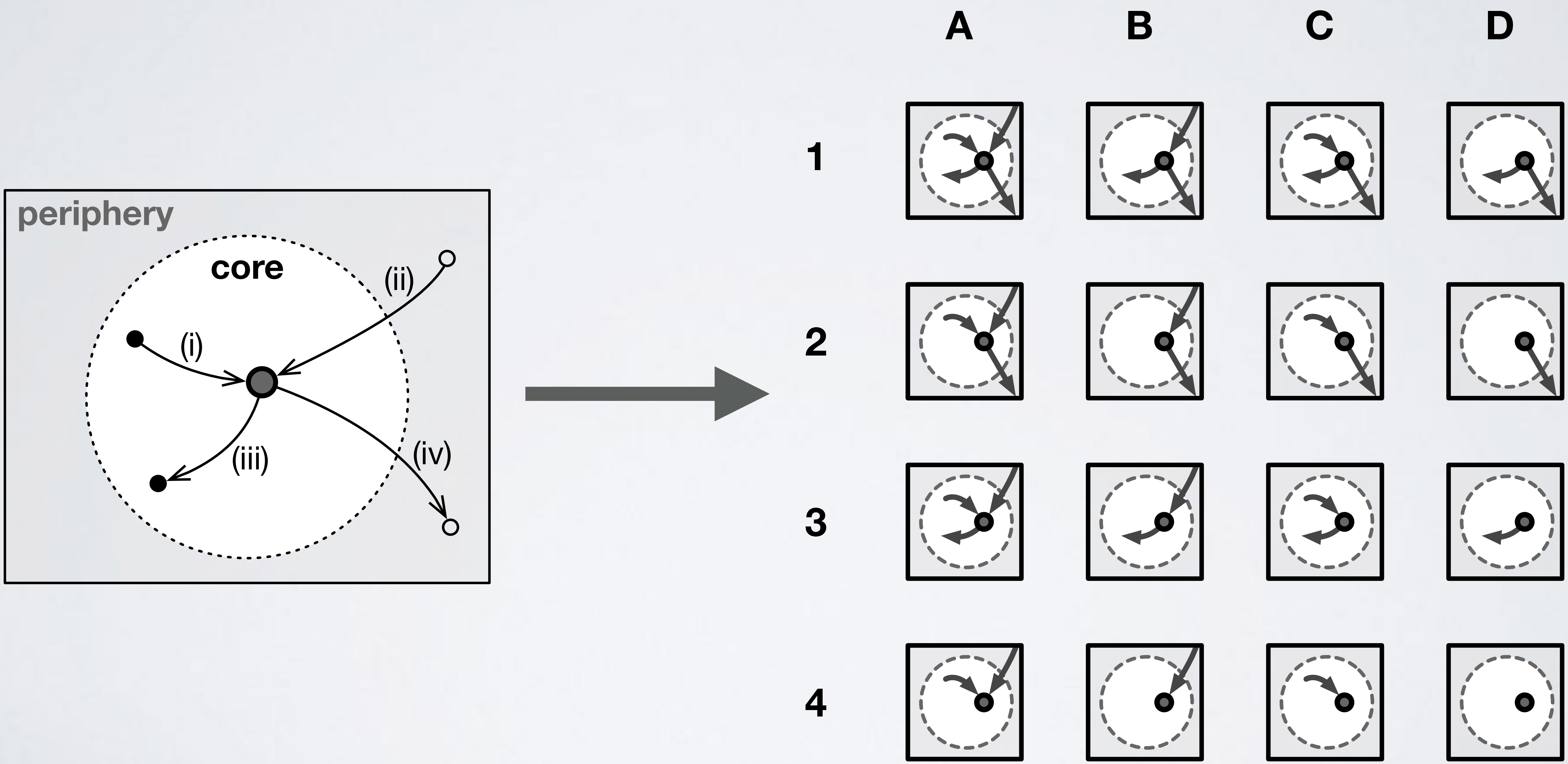
Supportive
(supporting the scientific basis of human-induced climate change):

#IPCC: Climate change is everywhere. That means we need to take action. Now. http://t.co/eItB60qMiR

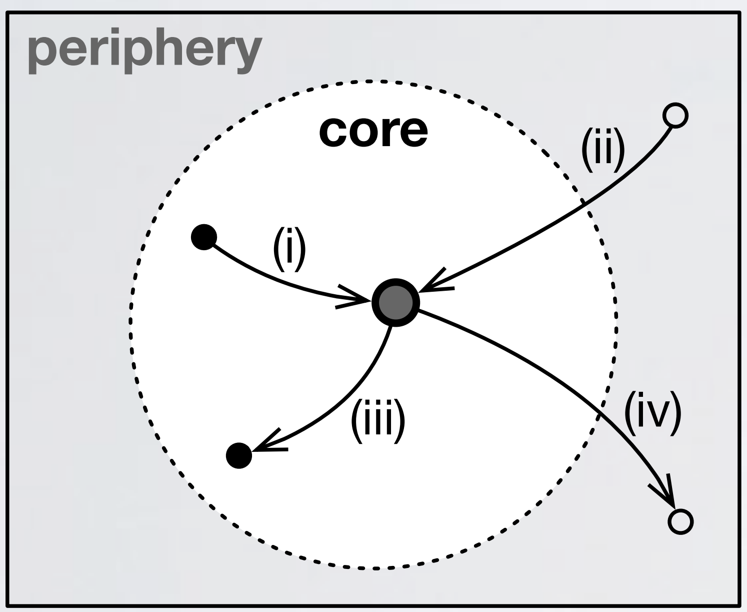
STRUCTURAL MODEL



STRUCTURAL MODEL

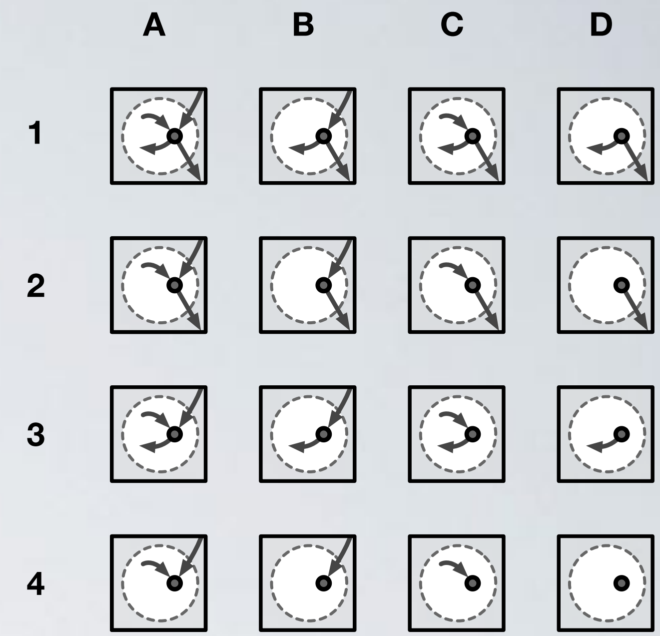


STRUCTURAL MODEL

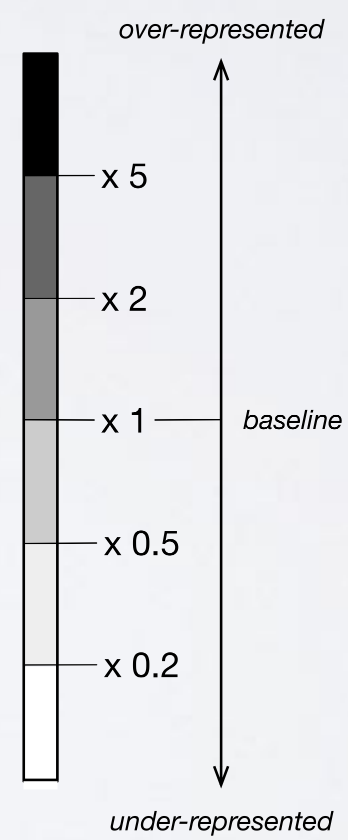


	A	B	C	D
1	p^4 0.16%	$p^3(1-p)$ 0.64%	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%
2	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%
3	$p^3(1-p)$ 0.64%	$p^2(1-p)^2$ 2.56%	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%
4	$p^2(1-p)^2$ 2.56%	$p(1-p)^3$ 10.2%	$p(1-p)^3$ 10.2%	$(1-p)^4$ 41.%

REMARKABLE POSITIONS

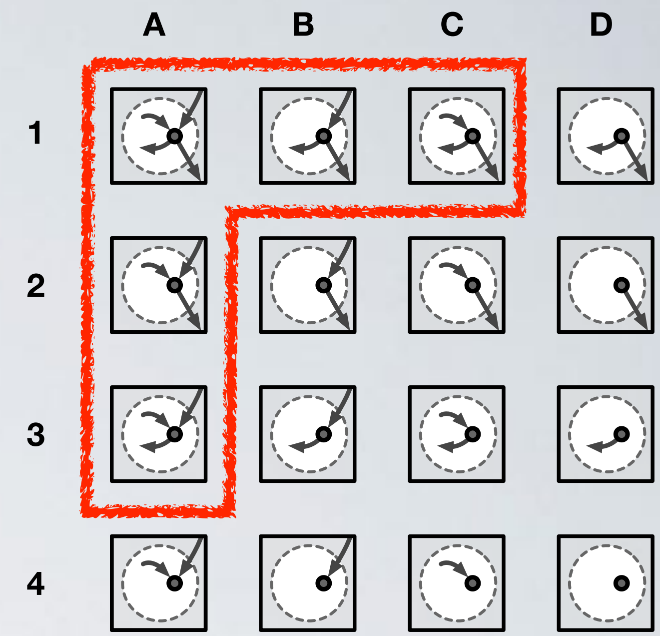


	A	B	C	D
1	**17** 0.5	**6** 2.6	**9** 2.2	**31** 11
2	**14** 3.3	**5** 15	**1** 14	**36** 69
3	**4** 2.2	**0** 11	**2** 10	**17** 47
4	**49** 14	**24** 69	**10** 60	**404** 297



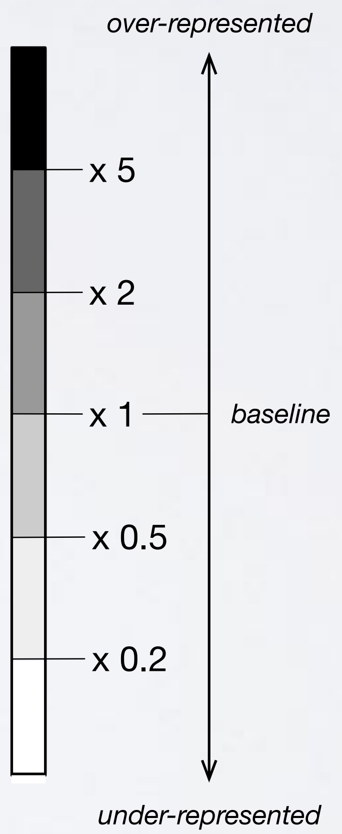
(629 core nodes)

REMARKABLE POSITIONS



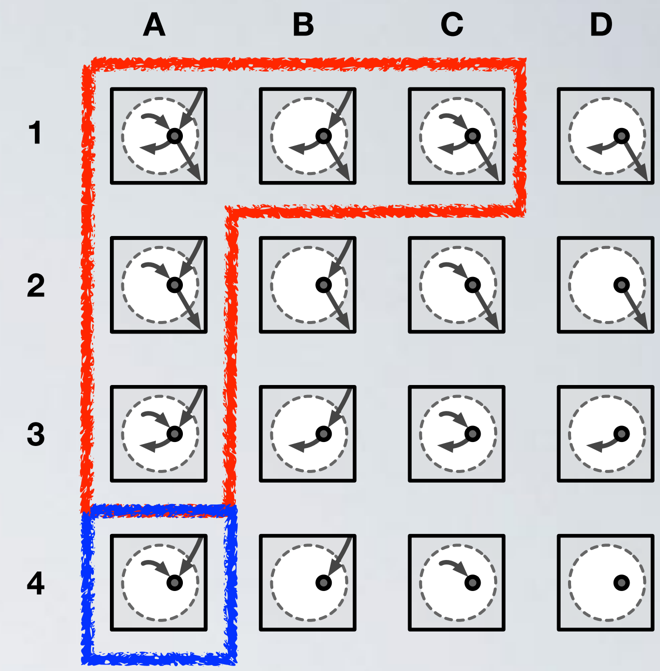
star

	A	B	C	D
1	17 0.5	6 2.6	9 2.2	31 11
2	14 3.3	5 15	1 14	36 69
3	4 2.2	0 11	2 10	17 47
4	49 14	24 69	10 60	404 297



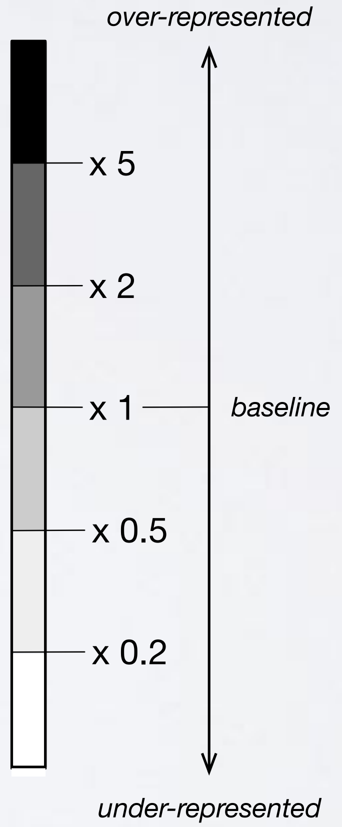
(629 core nodes)

REMARKABLE POSITIONS



star

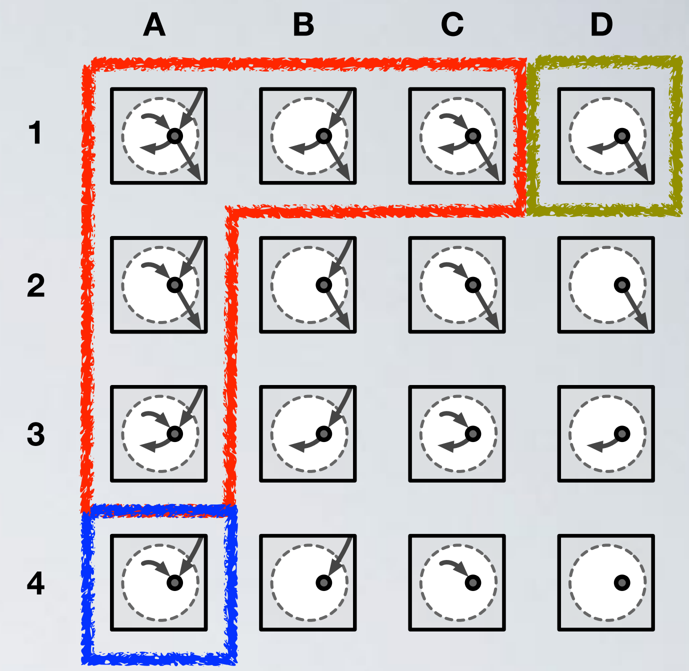
	A	B	C	D
1	17 0.5	6 2.6	9 2.2	31 11
2	14 3.3	5 15	1 14	36 69
3	4 2.2	0 11	2 10	17 47
4	49 14	24 69	10 60	404 297



famous

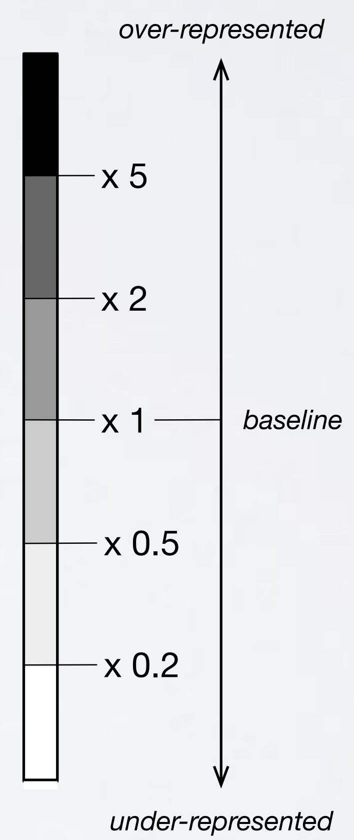
(629 core nodes)

REMARKABLE POSITIONS



star **curious**

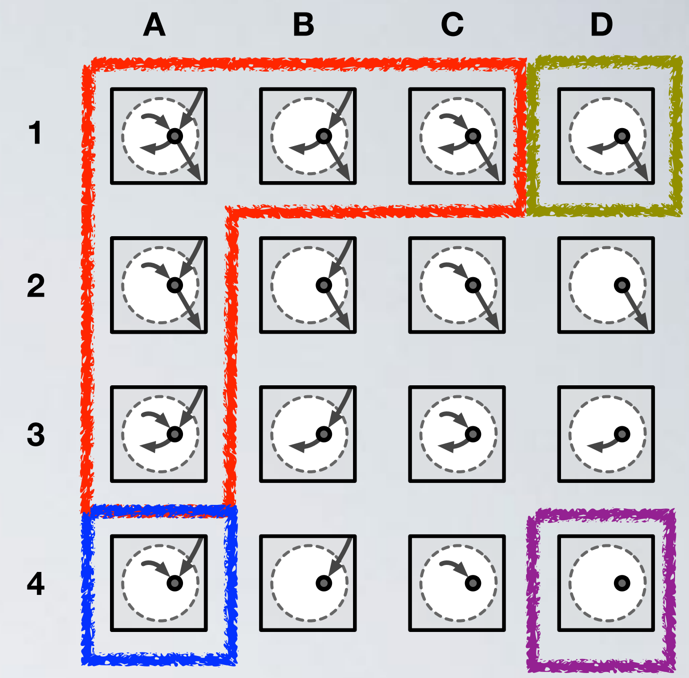
	A	B	C	D
1	17 0.5	6 2.6	9 2.2	31 11
2	14 3.3	5 15	1 14	36 69
3	4 2.2	0 11	2 10	17 47
4	49 14	24 69	10 60	404 297



famous

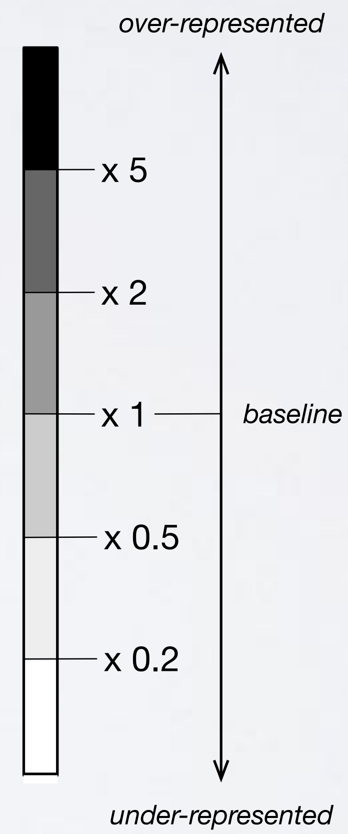
(629 core nodes)

REMARKABLE POSITIONS



star **curious**

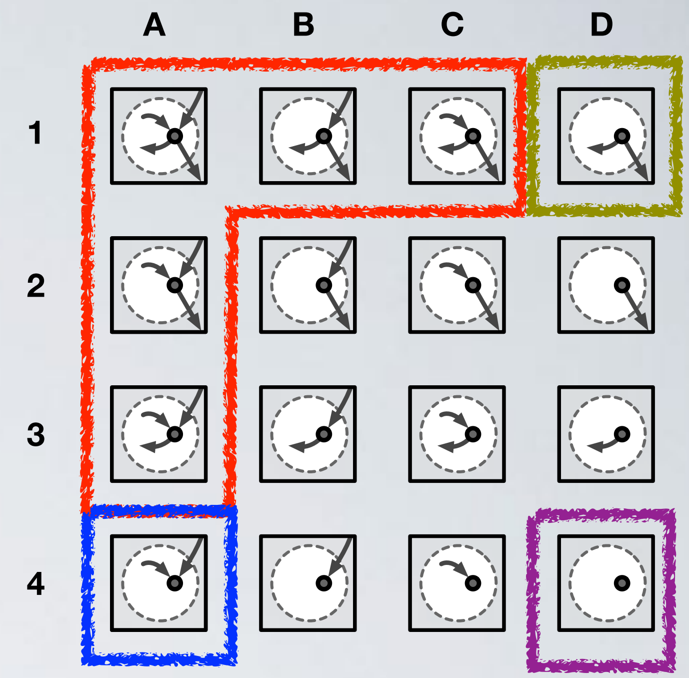
	A	B	C	D
1	**17** 0.5	**6** 2.6	**9** 2.2	**31** 11
2	**14** 3.3	**5** 15	**1** 14	**36** 69
3	**4** 2.2	**0** 11	**2** 10	**17** 47
4	**49** 14	**24** 69	**10** 60	**404** 297



famous **silent**

(629 core nodes)

REMARKABLE POSITIONS AND ALIGNMENT



Critical

	A	B	C	D
1	5 1.6	1 0.6	3 0.9	5 3
2	4 1.3	0 0.5	0 0.1	3 3.4
3	0 0.4		0 0.2	1 1.6
4	4 4.7	3 2.3	1 1	30 39

(60 nodes)

Supportive

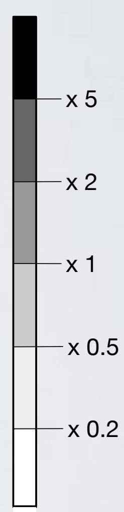
	A	B	C	D
1	3 6.2	2 2.2	3 3.3	13 11.3
2	3 5.1	0 1.8	0 0.4	11 13.1
3	1 1.5		1 0.7	9 6.2
4	15 17.8	7 8.7	5 3.6	156 147

(229 nodes)

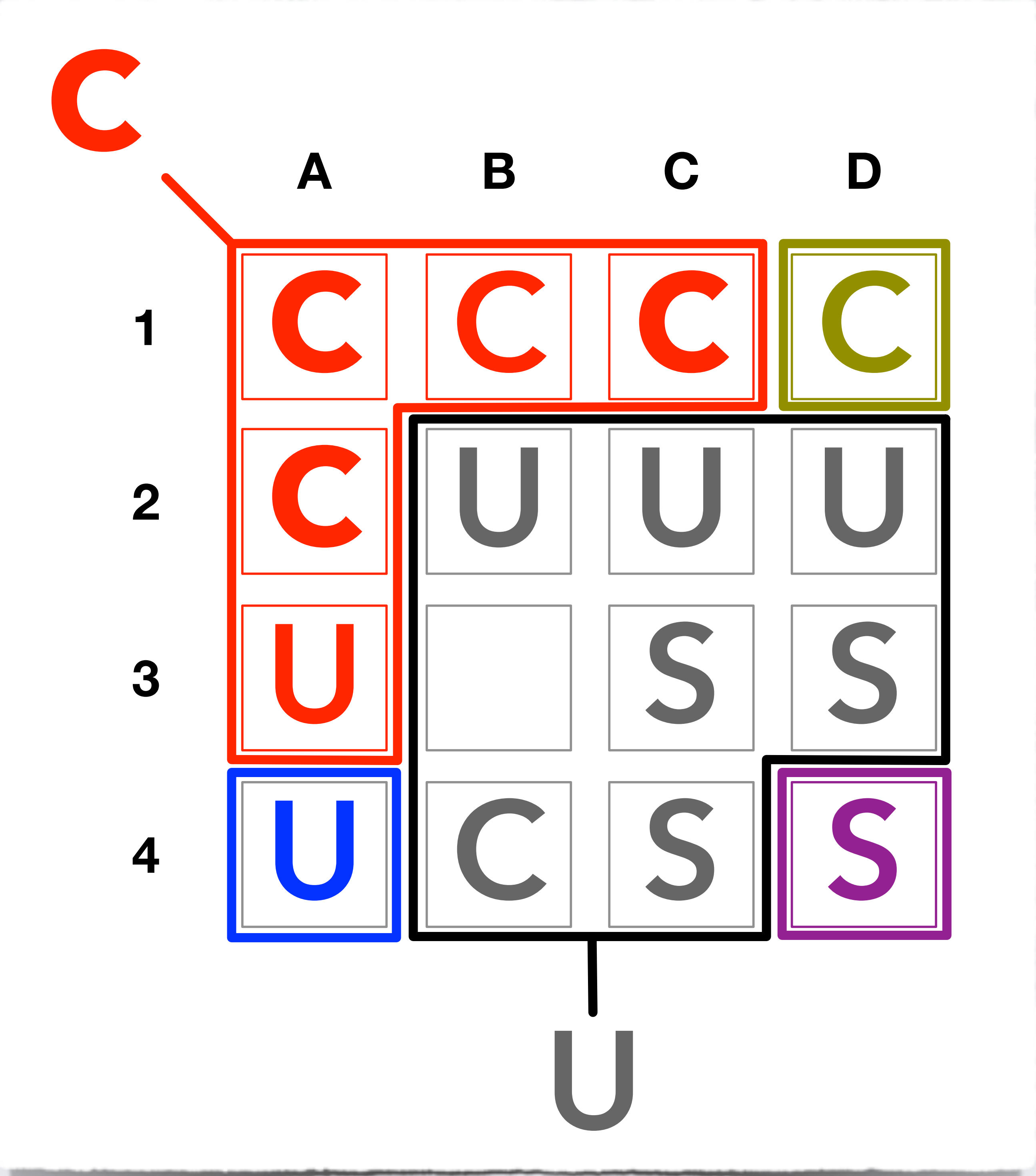
Uncommitted

	A	B	C	D
1	9 9.2	3 3.2	3 2.2	13 11
2	7 7.6	5 2.7	1 14	22 69
3	3 2.2		1 10	7 47
4	30 26.5	14 69	4 60	218 297

(340 nodes)



DOMINANT ALIGNMENTS IN REMARKABLE POSITIONS



SEMANTICS

$\mathbf{tf}_c(w)$: term frequency of word *w* in category *c*
(proportion of users using w *in category* c*)*

SEMANTICS

$\mathrm{tf}_c(w)$: term frequency of word *w* in category c

(proportion of users using w *in category* c*)*

$$\mathrm{rf}_c(w) = \frac{\mathrm{tf}_c(w)}{\langle \mathrm{tf}_{c'}(w) \rangle_{c' \in \mathcal{C}}}$$

relative term frequency of word *w* in category c

SEMANTICS

$\mathrm{tf}_c(w)$: term frequency of word *w* in category c
(proportion of users using w *in category* c*)*

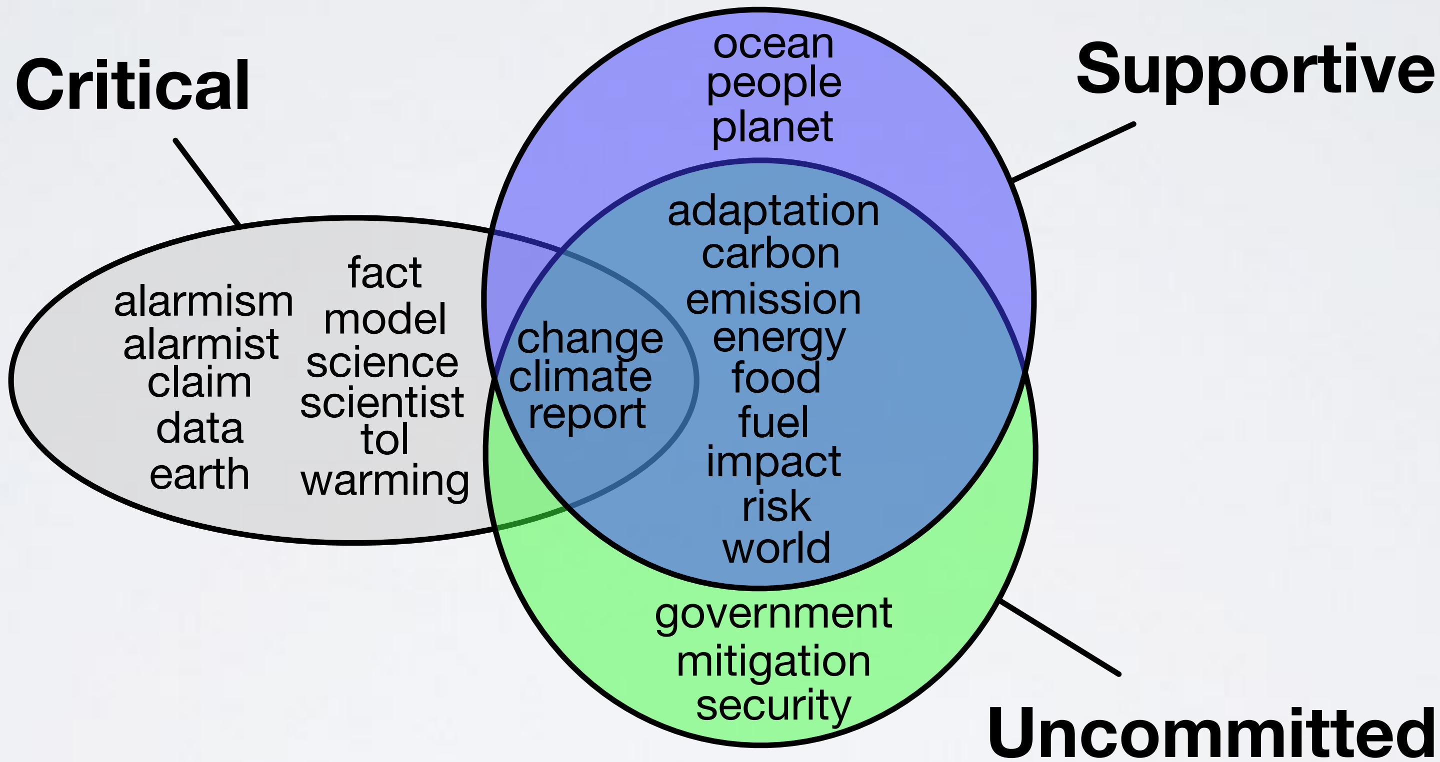
$$\mathrm{rf}_c(w) = \frac{\mathrm{tf}_c(w)}{\langle \mathrm{tf}_{c'}(w) \rangle_{c' \in \mathcal{C}}}$$

relative term frequency of word *w* in category c

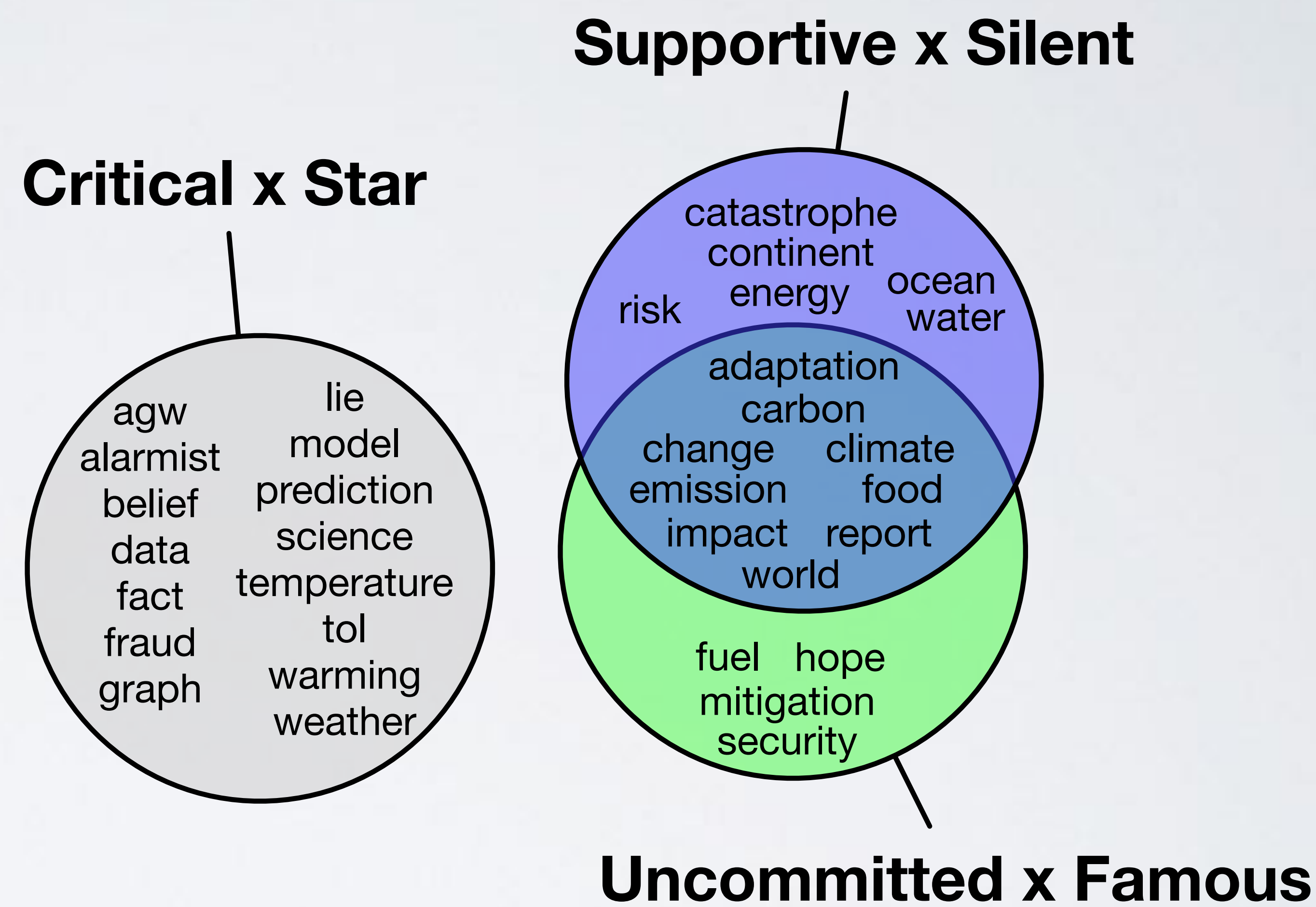
$$s_c(w) = \mathrm{tf}_c(w) \cdot \log \mathrm{rf}_c(w)$$

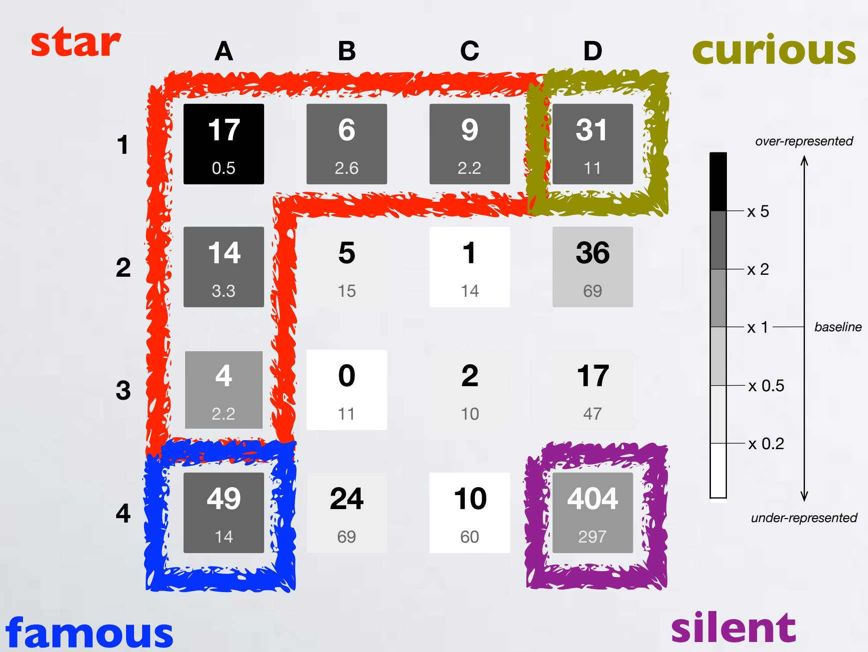
score ("typicality") of *w* in c

SEMANTICS AND ALIGNMENT



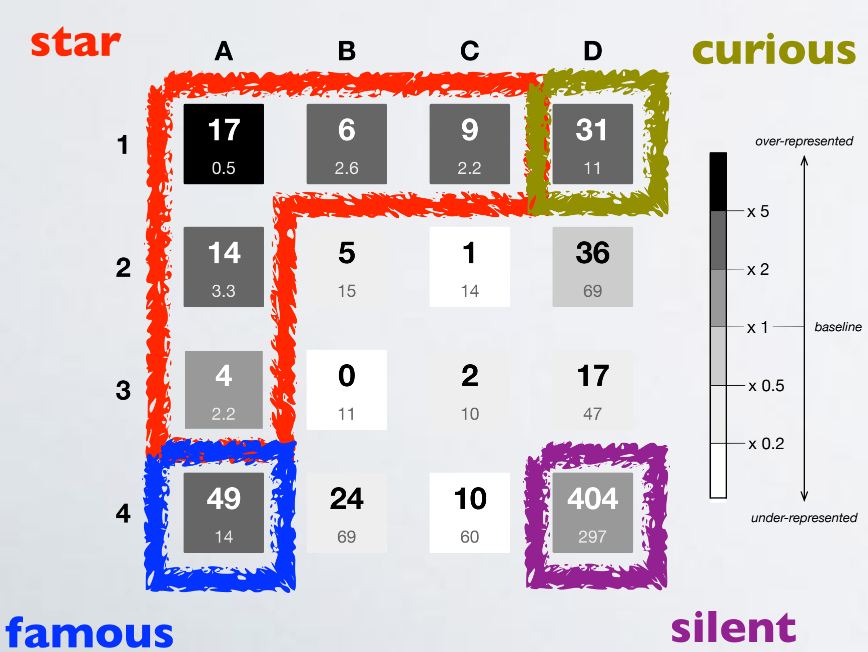
SEMANTICS AND REMARKABLE POSITIONS





SEMANTICS AND REMARKABLE POSITIONS

Slot group	Most typical terms
star	agw, alarmist, ar5, author, belief, co2, data, global, model, paper, prediction, science, tol, weather
famous	adaptation, continent, emission, food, fuel, hope, impact, security, water
curious	assessment, cost, earth, graph, ocean
silent	catastrophe



CONCLUSIONS

- Critical voices are in minority but occupy "star" position on Twitter:
  - *discussing 'beliefs', 'models' and 'data'*
    (CLIMATE CHANGE AS SCIENTIFIC ISSUE)
- Supportive majority of Twitter users in climate change debate are rather "silent":
  - *discussing 'impacts', 'emissions' and 'adaptation'*
    (CLIMATE CHANGE AS POLITICAL ISSUE)