Aalto University
School of Science

Political echo chambers in social media

Aristides Gionis

Department of Computer Science
Aalto University

NIPS workshop on prioritising online content
Dec 9, 2017

social media

- people use social media to
  - share information, express opinion, comment, interact, discuss, get personalized news feed
- 62% of adults in US get their news from social media

PEW RESEARCH CENTER

social media : good and bad sides

advantages

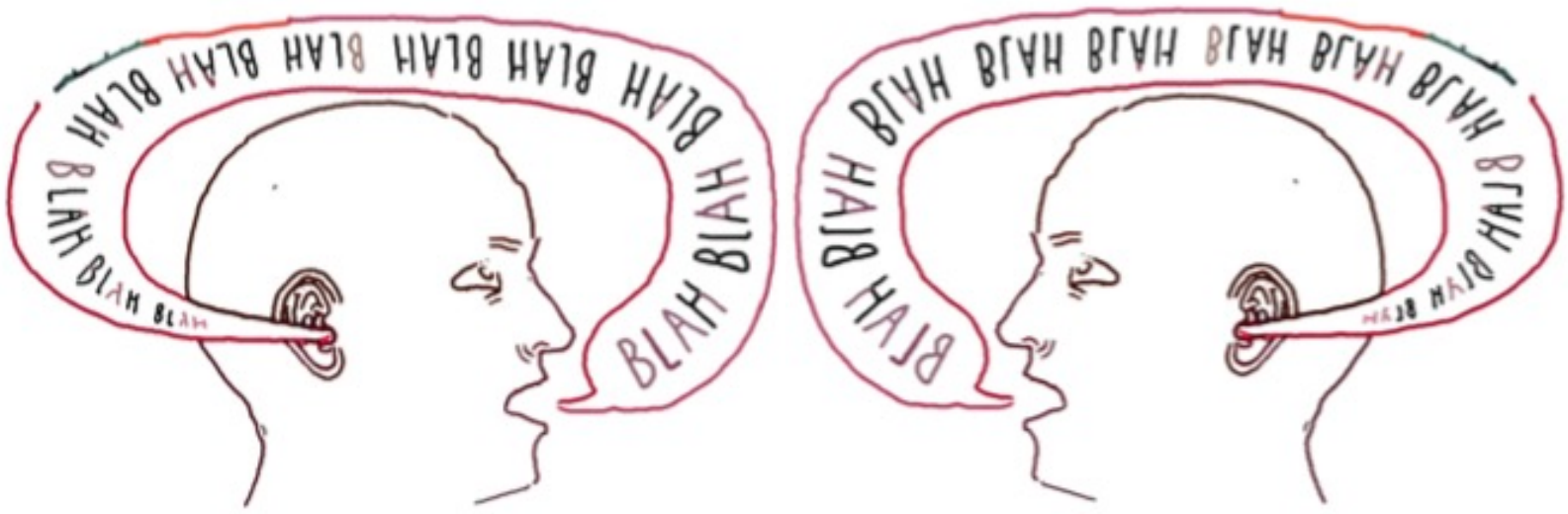
- no information barriers
- citizen journalism
- social connectivity
- democratization
- …

disadvantages

- harassment
- fake news
- echo chambers
- polarization
- …

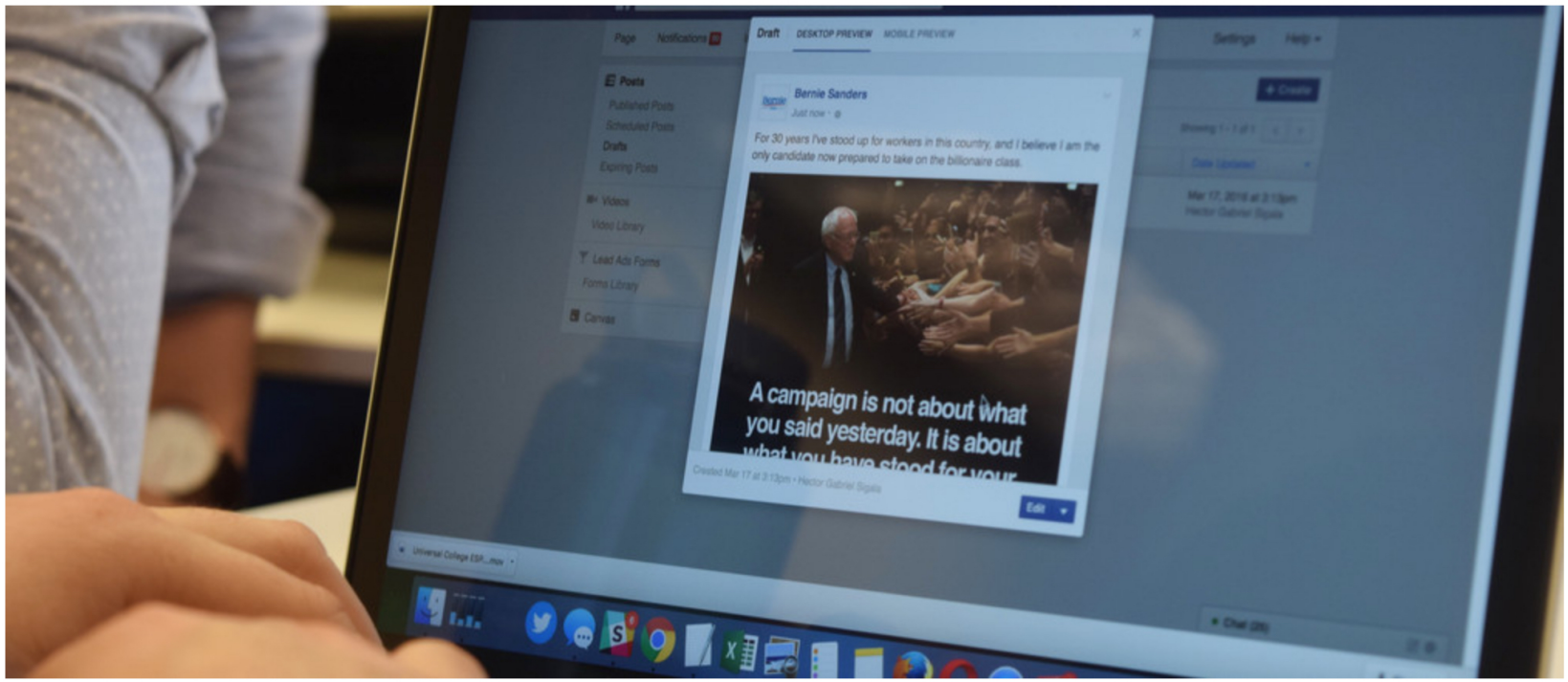
echo chambers

- a situation in which information, ideas, or beliefs are amplified or reinforced by communication and repetition inside a defined system



Global Agenda | Future of Government

The biggest threat to democracy? Your social media feed



The internet was meant to spread democracy. Could it be having the opposite effect?

Image: REUTERS/Melissa Fares

MOSTAFA M. EL-BERMAWY BUSINESS 11.18.16 05:45 AM

YOUR FILTER BUBBLE IS DESTROYING DEMOCRACY

2016 Presidential Election – Digital Analysis by the Numbers		
Hillary Clinton	Criteria	Donald Trump
238.5M	Total Social Media Shares	256.5M
16,633	Average Shares per Post	17,894
8.3M	Facebook Page Likes (Official Page)	12.2M
10.3M	Twitter Following (Official Page)	13.1M
32.9K	Number of Referring Domains	21.4K
1.51M	Number of Backlinks to Website	960K
550	Alexa Rank in The US	681

Social media

The truth about Brexit didn't stand a chance in the online bubble

Emily Bell

A political system which abandons facts and a media ecosystem which does not filter for truth asks too much of people

Obama Foundation taps social media to fight online echo chambers

The foundation's chief digital officer offers feedback forms, hashtags, and thought-provoking questions.

Rob LeFebvre, @roblef
07.05.17 in Politics

51
Comments

372
Shares

what may cause echo chambers?

individual biases

- homophily
  - tendancy to associate with similar-minded
- confirmation bias and biased assimilation
  - tendancy to interpret information so as to confirm one's beliefs
- closure
  - desire for firm answers; aversion for ambiguity

individual biases (continued)

- cognitive dissonance
  - positive feeling when presented with information that confirms one's belief
- selective exposure
  - tendancy to keep away from communication of opposite hue
- information overload
  - can act as a catalyst

group biases

- social identity
  - individuals associate themselves with social identities
    race, religion, gender, class, ...
- group polarization
  - a group tends to make decisions that are more
    extreme than the initial inclination of its members
- in-group favoritism
  - favoring in-group over out-group members

system biases

- algorithmic filtering
  - algorithmic personalization
- media bias
  - e.g., Fox news vs. MSNBC

echo chambers on twitter

a case study

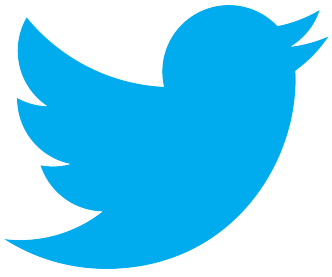
studying echo chambers

- working definition

  the political leaning of the content that users receive from the network agrees with that of the content they share

- consider the two components of the phenomenon
  - echo : the opinion shared (content)
  - chamber : the place it is shared (network)

datasets



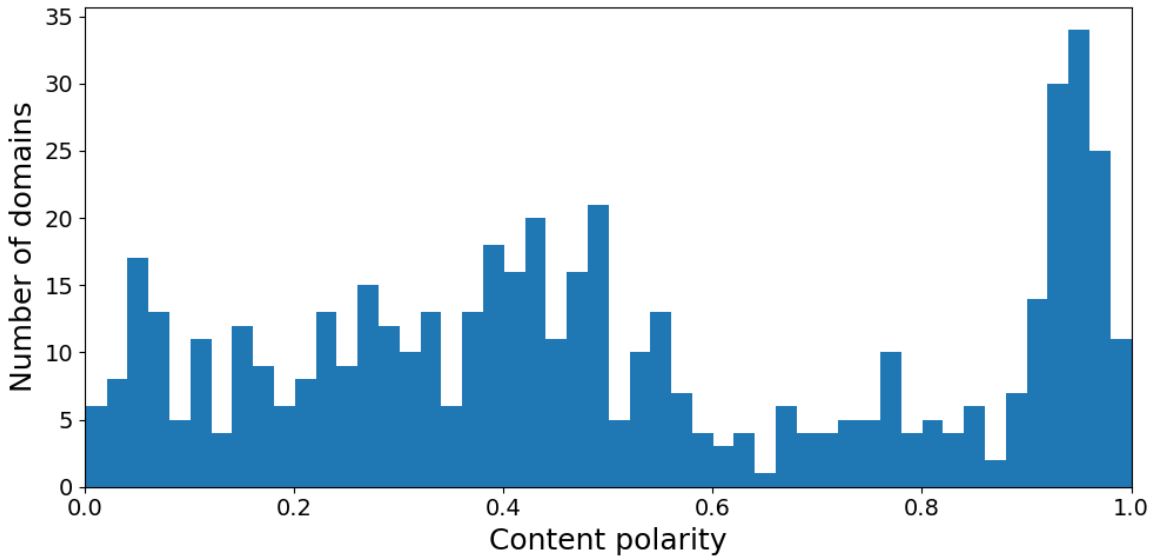
Topic	#Tweets	#Users	Event
guncontrol	19M	7506	Democrat filibuster for gun-control reforms (June 12–18, 2016)[6]
obamacare	39M	8773	Obamacare subsidies preserved in us supreme court ruling (June 22–29, 2015)[7]
abortion	34M	3995	Supreme court strikes down Texas abortion restrictions (June 27–July 3, 2016)[8]
combined	19M	6391	2016 US election result night (Nov 6–12, 2016)
large	2.6B	676 996	Tweets from users retweeting a U.S. presidential/vice presidential candidate (from [4], 2009–2016)
#ff	4M	3204	
#gameofthrones	5M	2159	
#love	3M	2940	filtering for these hashtags
#tbt	28M	12 778	
#foodporn	8M	3904	

content

- focus on news sources e.g., nyt, bbc, cnn, etc.
- assign content polarity score at each source

  0 : liberal — 1 : conservative
- obtain ground-truth scores for top-500 sources

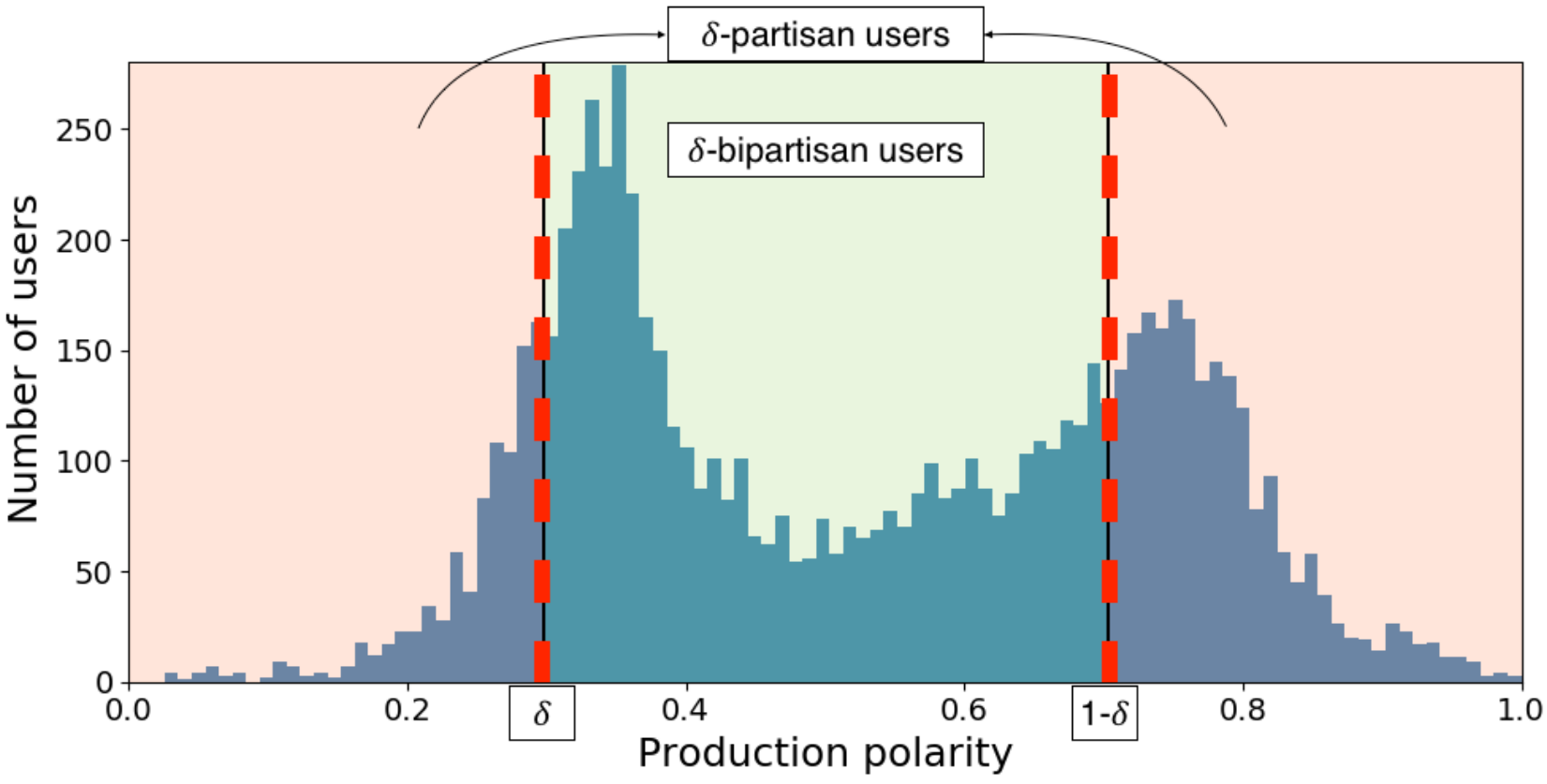
[Bakshy et al., Science, 2015]



users

- production polarity : avg polarity of shared content
- partisan : user with “extreme” production polarity
- bi-partisan : user with “medium” production polarity
- consumption polarity : avg polarity of followees’ content
- consumer : user with extreme consumption polarity
- gatekeeper : partisan but not consumer

users — production-polarity distribution



network features

- user polarity (democrat vs. republican)

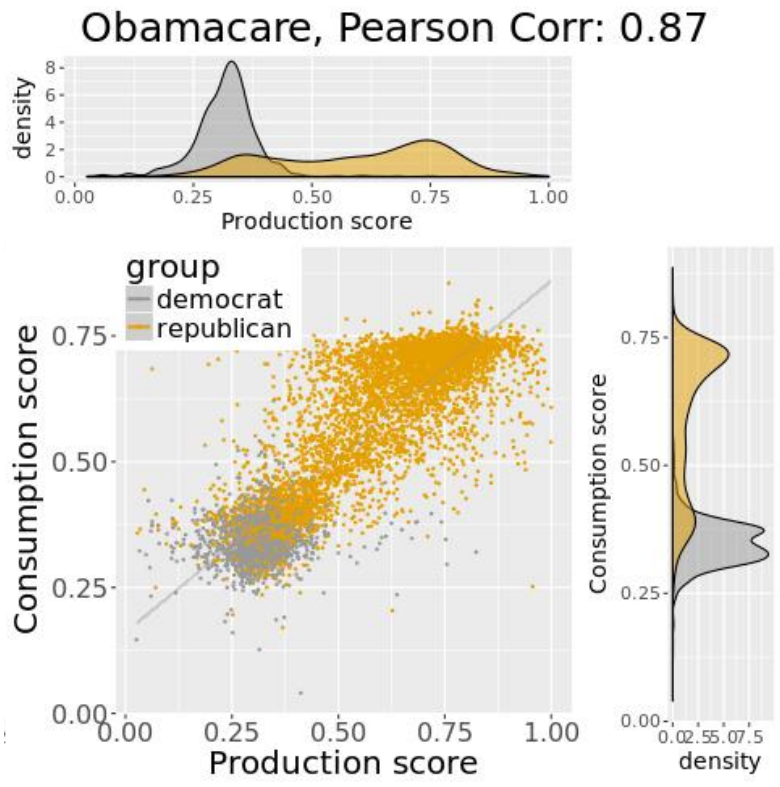
  [Barberá et al., Psychological Science, 2015]

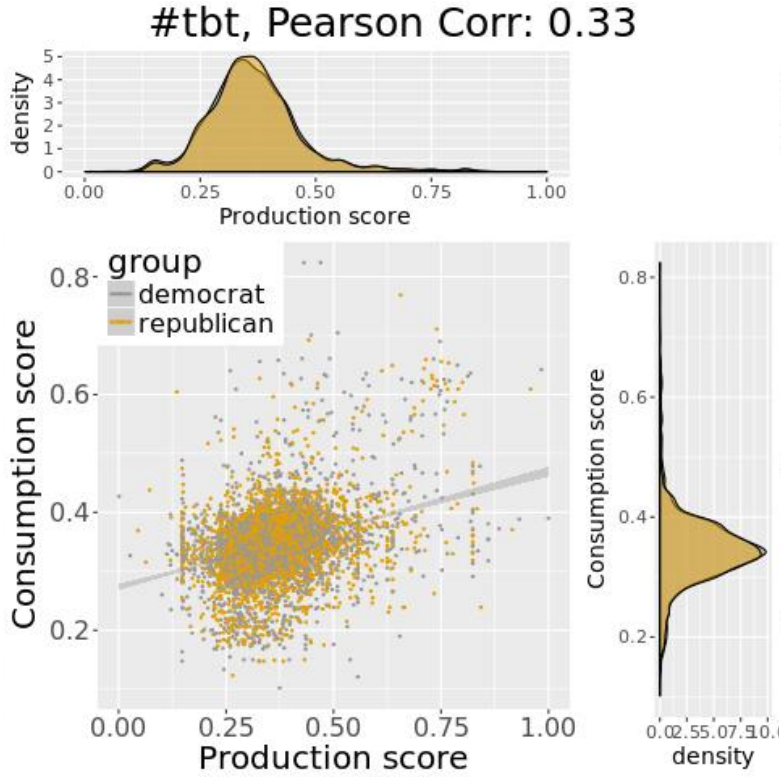
- network centrality : PageRank, in-degree
- clustering coefficient
- retweet ratio
- retweet volume

questions

- are there echo chambers?
- is there an advantage in being partisan?
- who are the users who act as gatekeepers?
- can we predict if a user is partisan or gatekeeper?

echo chambers
content production and consumption





partisans vs. gatekeepers

Features	Partisans	Gatekeepers
PageRank	✓	✓
clustering coefficient	✓	✓ (-)
user polarity	✓	✓ (-)
degree	✓	✓
retweet rate	✓	✗
retweet volume	✓	✗
favorite rate	✓	✗
favorite volume	✓	✗
# followers	✗	✗
# friends	✗	✗
# tweets	✗	✗
age on Twitter	✗	✗

also, predicting partisans can be achieved with higher accuracy than predicting gatekeepers

summary of findings

- echo chambers observed in politically contentious topics
- echo chambers not observed in non-contentious topics
- bi-partisan users pay a price in terms of network centrality, community connection, and endorsements
- gatekeepers : who are they and what is their role?

  e.g., ordinary open-minded citizens?

how to mitigate echo chambers?

mitigation action 1

improve awareness

improve awareness

- develop tools for users to perceive their “news diet”
- visualize/navigate in the underlying ideology space,
  their position, the accounts they follow, the news they read
- offer functionalities such as

  “find a high-quality article on the same topic
  from the opposing viewpoint”

learning of ideological leanings

- infer ideological stances of users and content
  e.g., liberal–conservative space
- common latent space for users and content
- e.g., substitute ground-truth polarities in previous study with learned polarities
- joint non-negative matrix-factorization task

[Lahoti et al., WSDM, 2018]

intuition

- map users and content in a joint latent ideology space, s.t.
- similar users are more likely to follow each other
- similar users are more likely to share similar content
- similar content is more likely to be shared by similar users

*similar means close in the latent ideology space

the problem setting

- social network $G = (V, E)$
  - adjacency matrix $\mathbf{A} \in \mathbb{R}^{n \times n}$
- user–content matrix $\mathbf{C} \in \mathbb{R}^{m \times n}$
- latent matrix representing user ideology $\mathbf{U} \in \mathbb{R}^{n \times k}$
- latent matrix representing content ideology $\mathbf{V} \in \mathbb{R}^{m \times k}$
- decompose

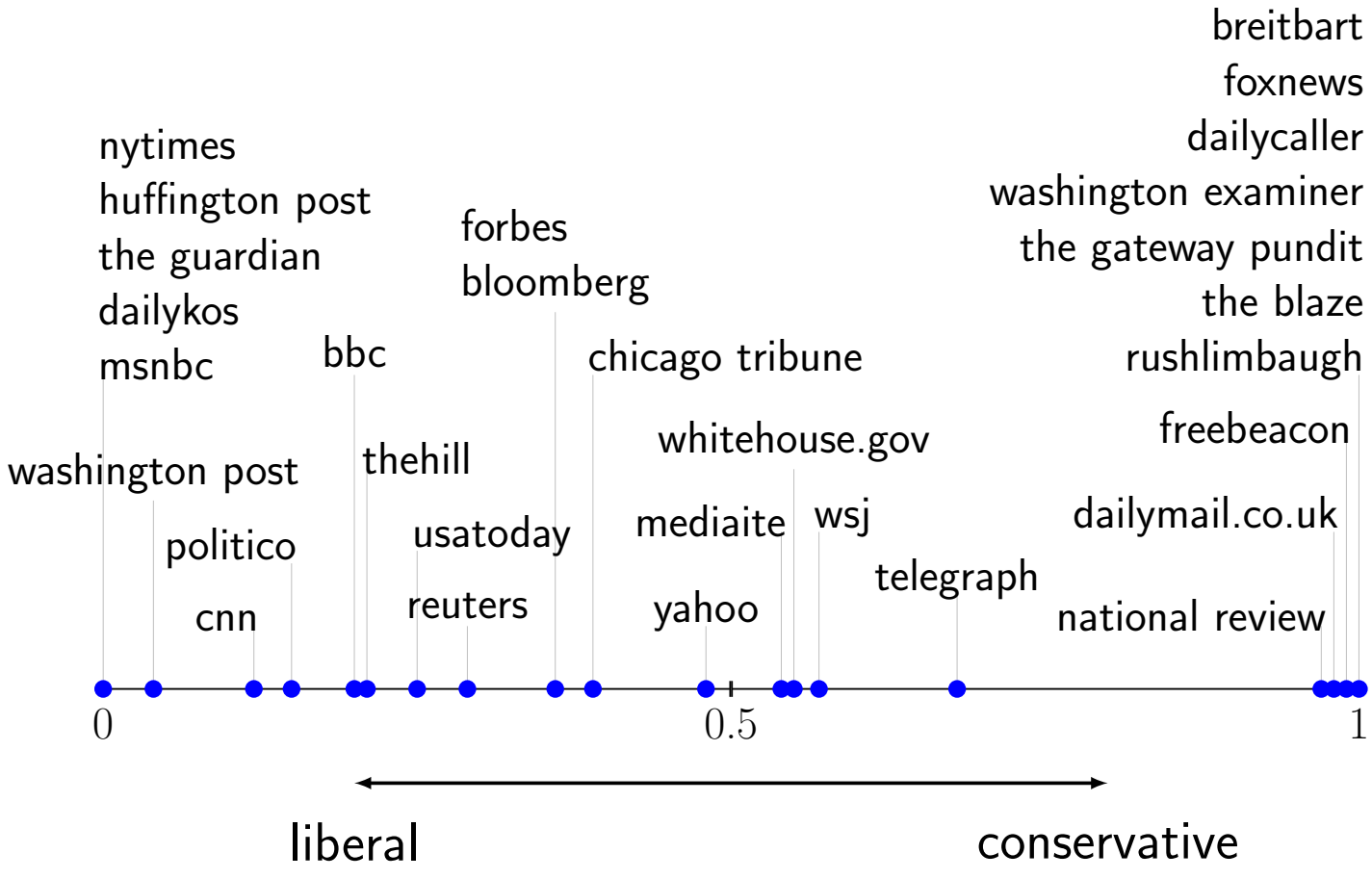
$$\mathbf{A} \approx \mathbf{U}\mathbf{H}_u\mathbf{U}^T \quad \text{and} \quad \mathbf{C} \approx \mathbf{U}\mathbf{H}_v\mathbf{V}^T$$

subject to orthonormal $\mathbf{U}$ and $\mathbf{V}$ and graph-regularization

in practice

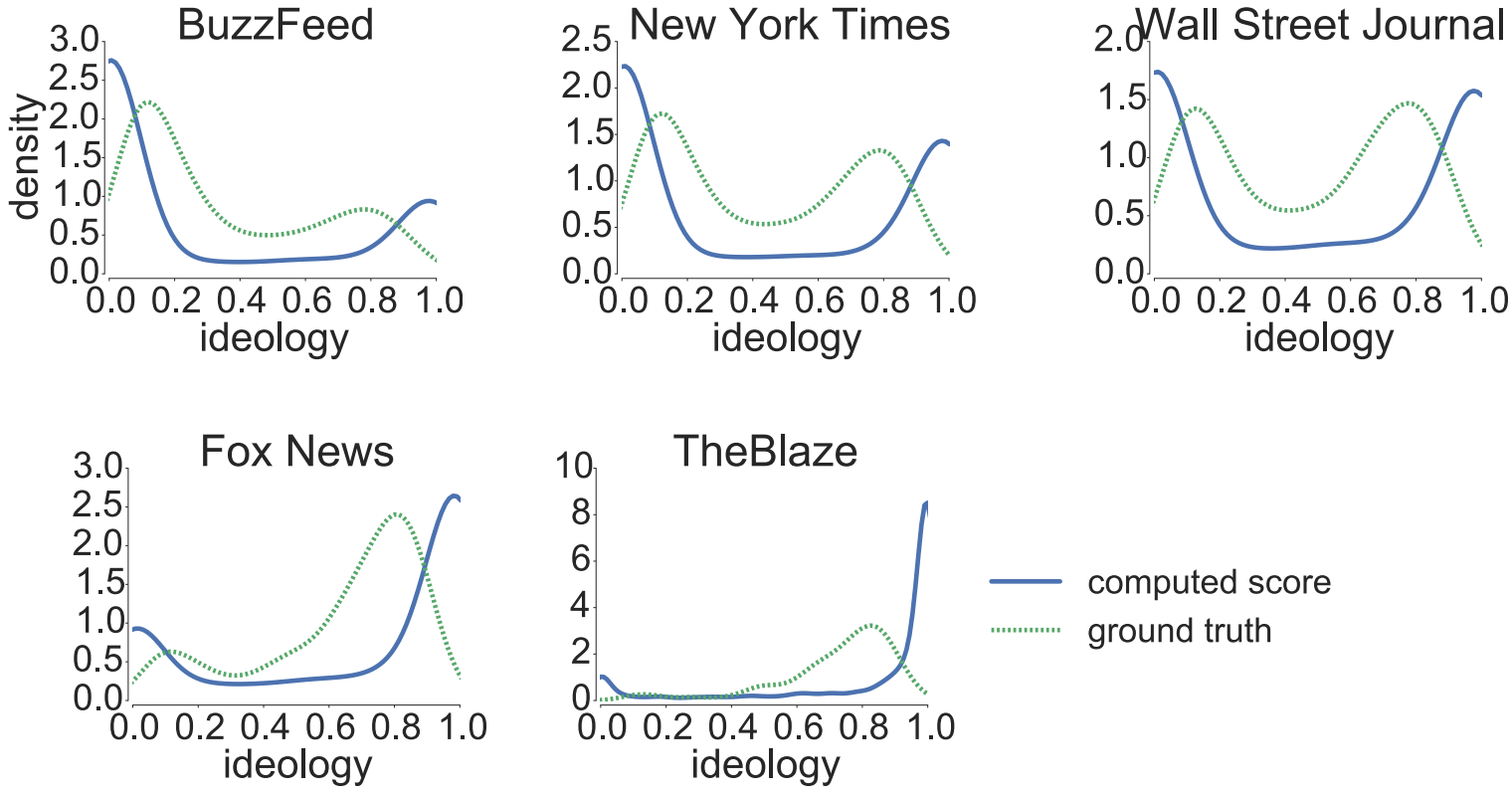
- twitter data from 2011 to 2016, focusing on controversial topics (gun control, abortion, obamacare)
- 6 391 users and 19 million tweets
- user matrix **A** represents follow graph
- content items represent url hostnames
- gather ground-truth polarity scores
  - content polarity [Bakshy et al., 2015]
  - user polarity [Barberá et al., 2015]

content ideology scores



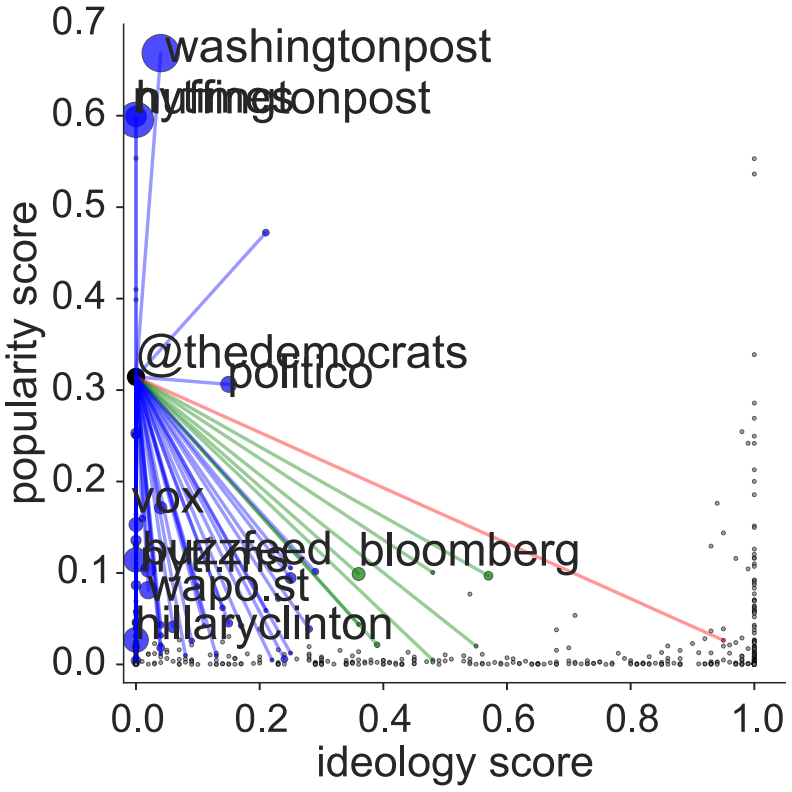
correlation with ground-truth scores 0.82

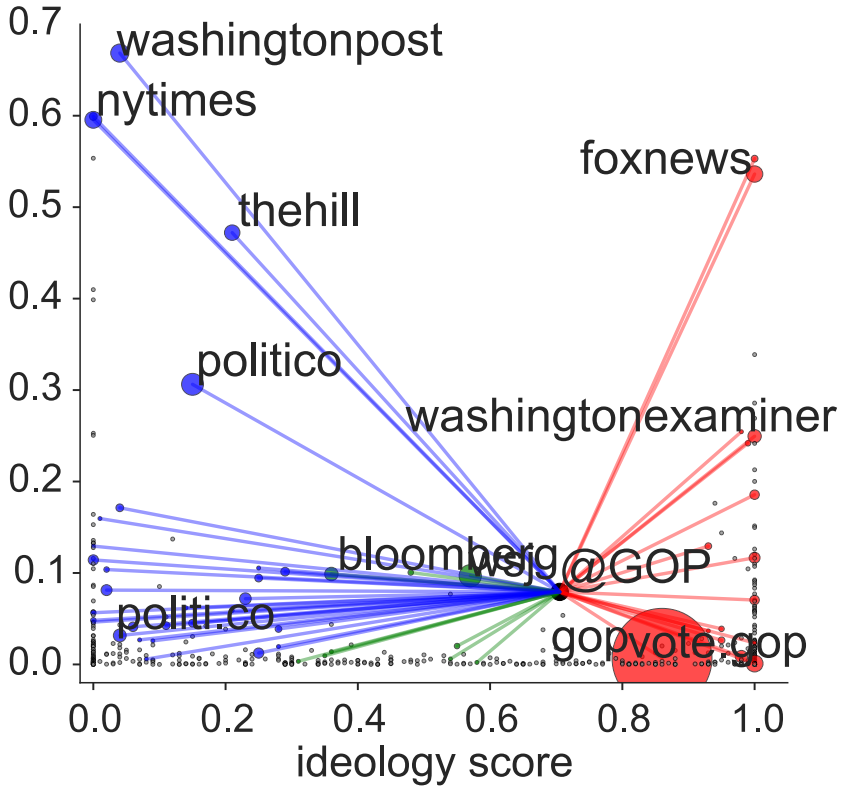
audience ideology scores



correlation of user ideology scores with ground-truth 0.90

visualizing the information bubble



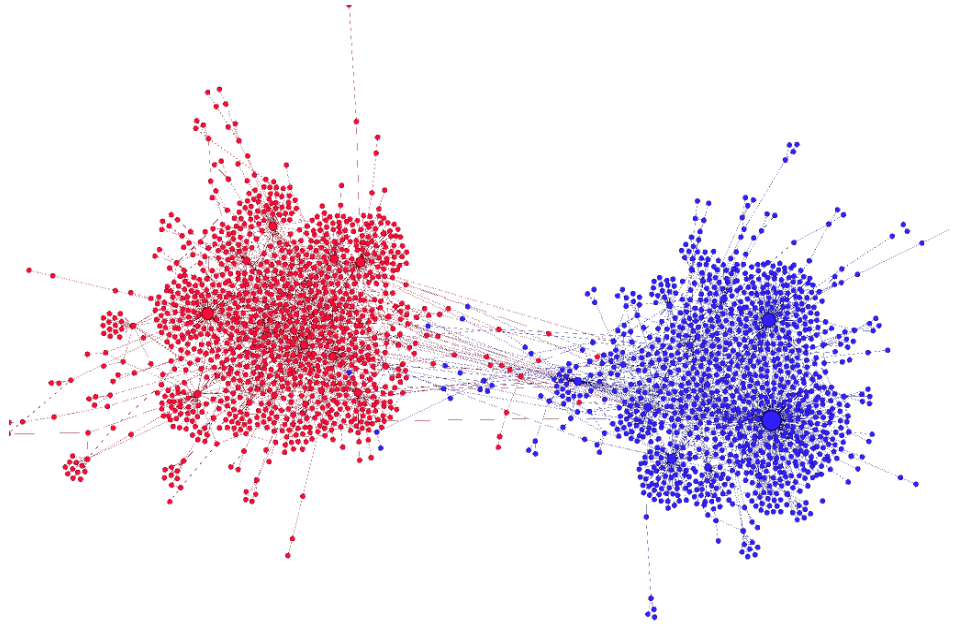


mitigation action 2

user-to-user recommendation

user-to-user recommendation

- social network has clustered structure



- user-to-user recommendation to reduce clustered structure
- e.g., minimize average shortest path length, maximize conductance, etc.
- account for acceptance probabilities

[Garimella et al., WSDM, 2017]

mitigation action 3

balance information exposure

balancing information exposure

- the standard viral-marking setting [Kempe et al. 2003]
  - a social network
  - a model of information propagation
    e.g., the independent-cascade model
  - an action (e.g., meme) propagates in the network
- the influence-maximization problem
  - find k seed nodes to maximize spread
- the standard solution
  - spread is non-decreasing and submodular
  - greedy given $(1 - \frac{1}{e})$ approximation

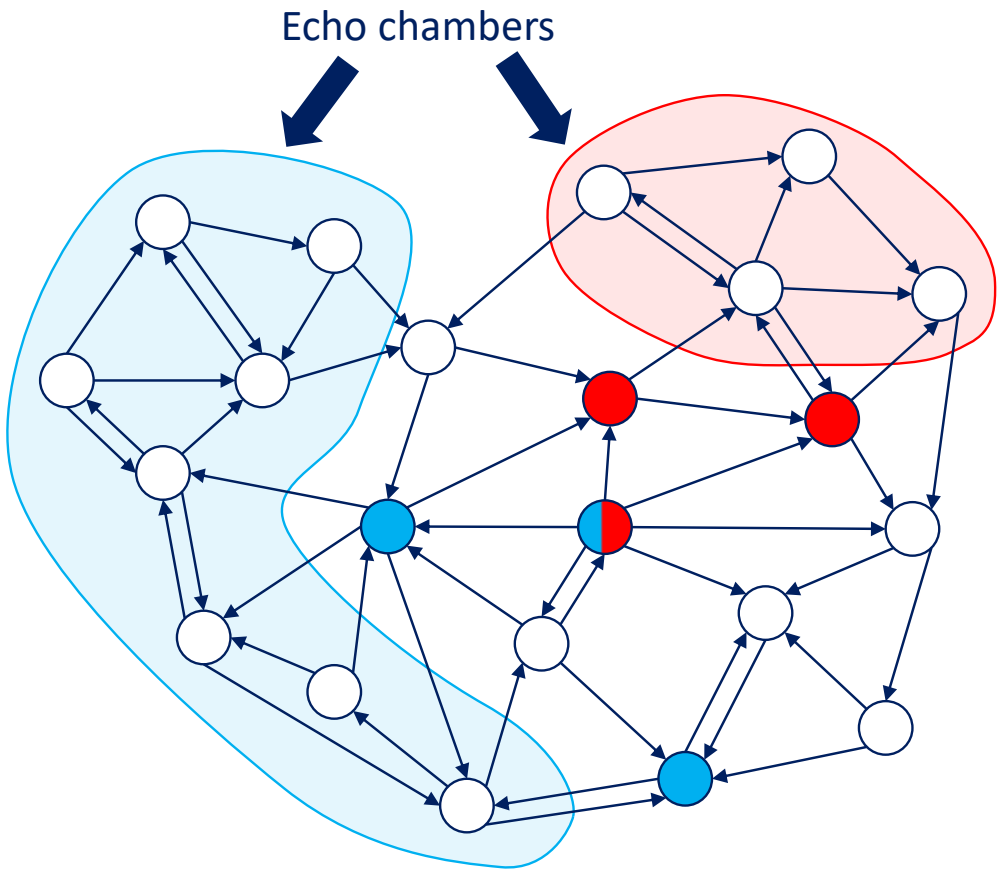
balancing information exposure

- proposed setting
  - a social network and two campaigns
  - seed nodes I_1 and I_2 for the two campaigns
  - a model of information propagation

- the problem of balancing information exposure
  - find additional seeds S_1 and S_2, with $|S_1| + |S_2| \leq k$
  - s.t. minimize # of users who see only one campaign
    or maximize # of users who see both or none

illustration

social discussion on fracking



balancing information exposure : our results

[Garimella et al., NIPS, 2017]

- optimization problem is **NP**-hard
- minimization problem is **NP**-hard to approximate
- maximization problem: objective function non monotone and non submodular
- different models of how the two campaigns propagate
- approximation guarantee $\frac{1}{2}(1 - \frac{1}{e})$

balancing information exposure : example

Side 1
Pro-Choice

Side 2
Pro-Life

Hedge

Pro-Remain

Pro-Leave

summary

- evidence of echo chambers in political discussions on social networks
  - price of bi-partizanship
- actions to mitigate echo chambers
  - improve awareness
  - user recommendation
  - content recommendation

discussion, limitations, future work

- models use mostly network structure
  - language-independent, but
  - incorporating language can help
- simple models
  - two-sided controversies
  - external influence is ignored
  - “follow” does not imply content consumption
  - simple propagation models
- evaluation is challenging, done on few topics
- analysis limited to twitter

thank you!

credits

Kiran Garimella

Gianmarco De Francisci Morales

Michael Mathioudakis

Preethi Lahoti

Nikos Parotsidis

Nikolaj Tatti