Aalto University
School of Science

Combating polarization in social media

Aristides Gionis

Department of Computer Science
Aalto University

Italian information retrieval workshop   IIR 2018
Rome, May 29, 2018

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
- ...

social media : good and bad sides

advantages

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

disadvantages

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

polarization

- political or social polarization

  *the act of separating or making people separate into two groups with completely opposite opinions**

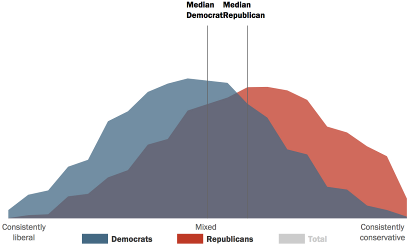
- related term: controversy

  *public discussion and argument about something that many people strongly disagree about**

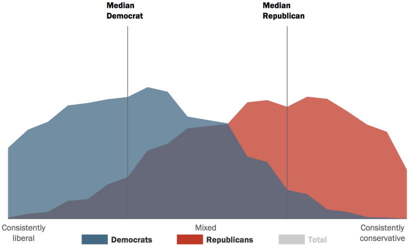
*oxford english dictionary

polarization in US politics

1994



2014



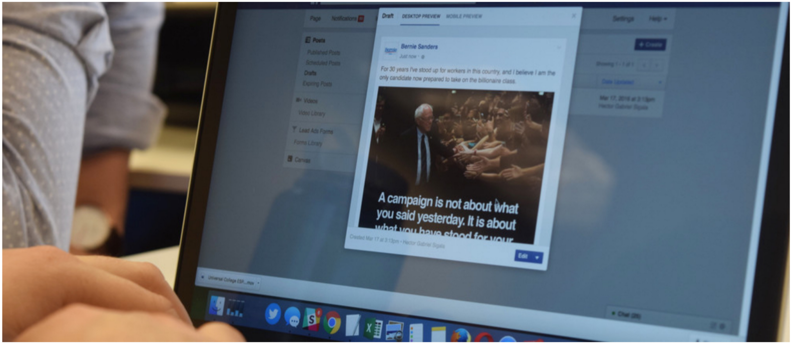
PEW RESEARCH CENTER

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

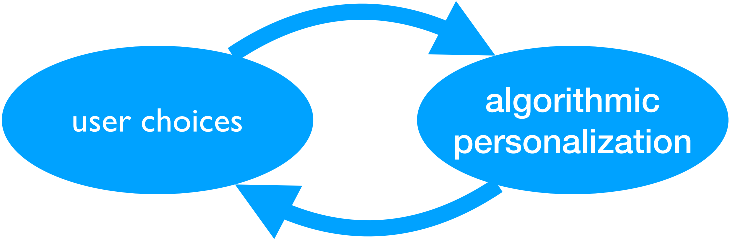


The internet was meant to spread democracy. Could it be having the opposite effect? Image: REUTERS/Melissa Fares

what may cause echo chambers?

- individual biases
  - homophily, confirmation bias,
    cognitive dissonance, selective exposure
- group biases
  - social identity, group polarization,
    in-group favoritism
- system biases
  - algorithmic filtering, algorithmic personalization,
    media bias

the polarization cycle



research questions

- do echo chambers exist?
- can we identify polarized discussions in social media?
- can we design algorithms to help reduce polarization?
- can we design algorithms to moderate online discussions?

research questions

do echo chambers exist?

what is the interplay between content and network?

who are the key players?

K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, *"Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship"*, The Web Conference (WWW) 2018

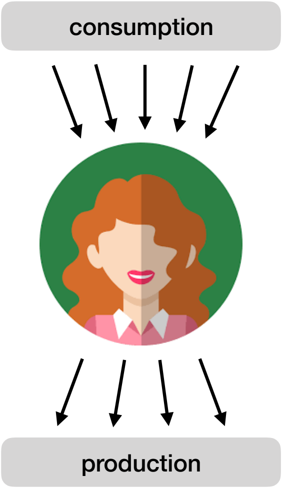
studying echo chambers

- working definition

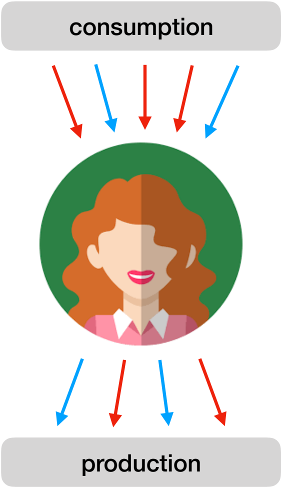
  the political leaning of the content that users receive from the network agrees with that of the content they share

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

methodology



methodology



datasets

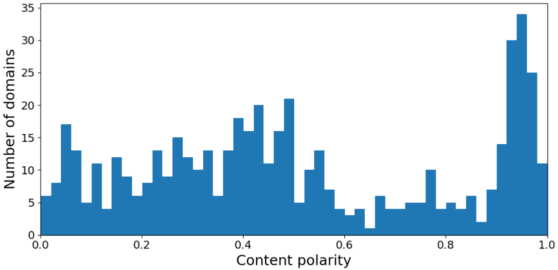
Topic	#Tweets	#Users	Event
guncontrol	19M	7506	Democrat filibuster for gun-control reforms (June 12–18, 2016)[6]
obamacare	39M	8773	Obamacare subsidies preserved in us supreme court ruling (June 22–29, 2015)[7]
abortion	34M	3995	Supreme court strikes down Texas abortion restrictions (June 27–July 3, 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	filtering for these hashtags
#gameofthrones	5M	2159	
#love	3M	2940	
#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

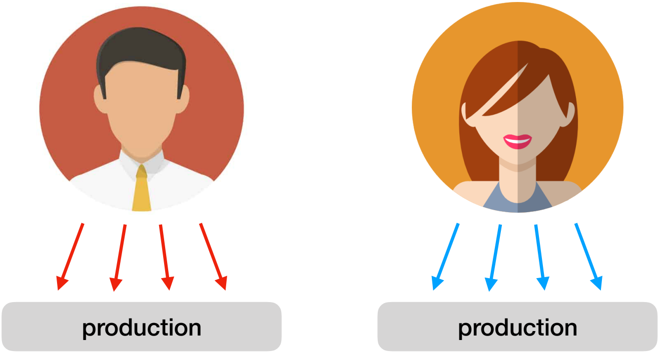
[Bakshy et al., Science, 2015]



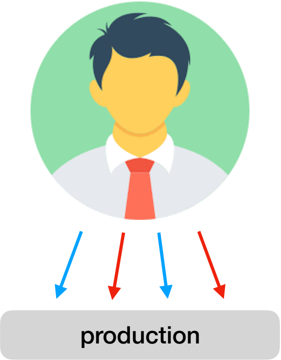
characterize users based on

- production polarity : avg polarity of shared content
- consumption polarity : avg polarity of followees' content

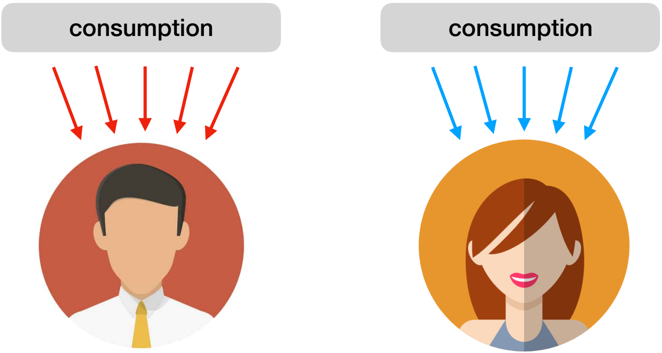
user roles : partisan



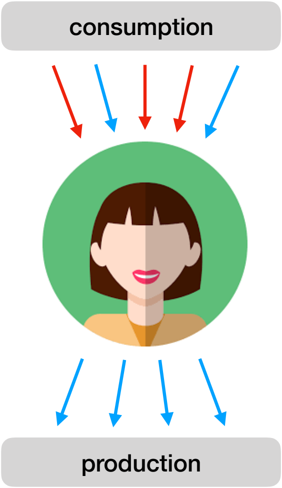
user roles : bi-partisan



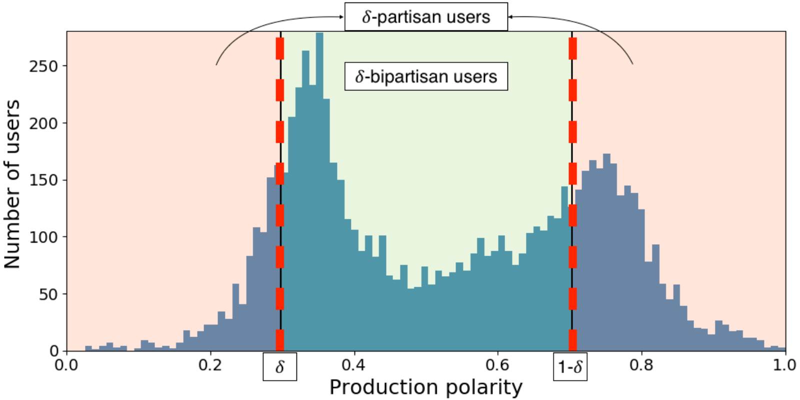
user roles : consumer



user roles : gatekeeper



users — production-polarity distribution



network features

- user polarity (democrat vs. republican)

  [Barberá et al., Psychological Science, 2015]

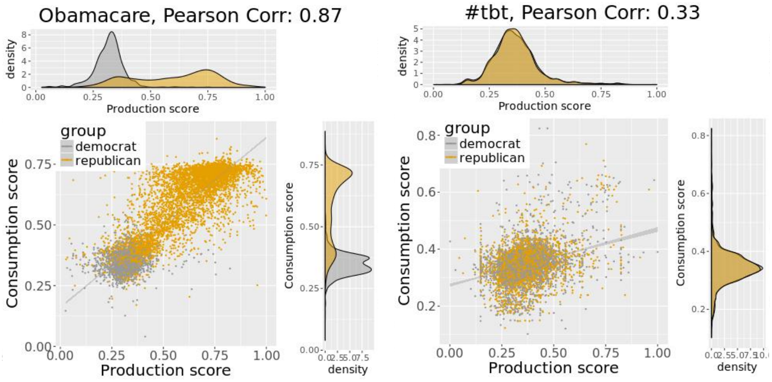
- network centrality : PageRank, in-degree
- clustering coefficient
- retweet ratio
- retweet volume

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. bi-partisans

gatekeepers vs. non gatekeepers

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

partisans vs. bi-partisans

gatekeepers vs. non gatekeepers

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

there is a price to be bi-partisan

prediction

- tweet features
  - n-grams with tf $\cdot$ idf weights
- profile features
  - number of tweets / followers / friends, age on twitter
- network features
  - PageRank, degree, clustering coefficient

prediction

- tweet features
  - n-grams with tf $\cdot$ idf weights
- profile features
  - number of tweets / followers / friends, age on twitter
- network features
  - PageRank, degree, clustering coefficient

predicting partisans (accurasy $\approx$ 0.81)

is easier than

predicting gatekeepers (accurasy $\approx$ 0.68)

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?

research question

can we identify and quantify polarization ?

K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, *"Quantifying controversy in social media"*, ACM WSDM 2016

how can we identify polarization ?

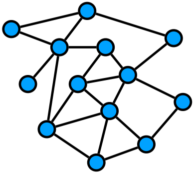
ideas

- content
  - do opposing sides say different things ?
- sentiment
  - do polarized topics exhibit wider range of emotions ?
- interactions
  - do people interact more with their own side ?

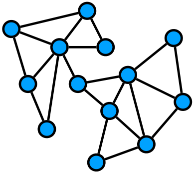
method template

- build an interaction graph
  - try several types

    retweets, replies, connections
- is the interaction graph polarized?
- output polarization score

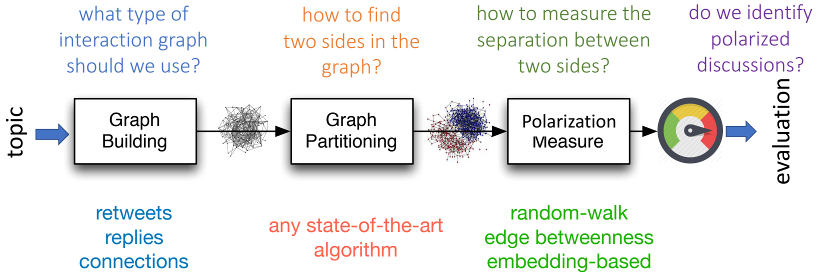


non polarized



polarized
two sides well separated

pipeline


what type of interaction graph should we use?
how to find two sides in the graph?
how to measure the separation between two sides?
do we identify polarized discussions?
topic
Graph Building
Graph Partitioning
Polarization Measure
evaluation
retweets
replies
connections
any state-of-the-art algorithm
random-walk
edge betweenness
embedding-based

random-walk controversy score (RWC)

- assume graph is partitioned in two sides, A and B
- consider a random walk that started at a random node and finished in a hub in $Y \in \{A, B\}$
- probability that random walk started in $X \in \{A, B\}$

$$P_{XY} = \Pr(\text{r.w. started in } X \mid \text{r.w. finished in } Y)$$

random-walk controversy score (RWC)

- assume graph is partitioned in two sides, A and B
- consider a random walk that started at a random node and finished in a hub in $Y \in \{A, B\}$
- probability that random walk started in $X \in \{A, B\}$

$$P_{XY} = \Pr(\text{r.w. started in } X \mid \text{r.w. finished in } Y)$$

- random-walk controversy score (RWC)

$$\text{RWC} = P_{AA}P_{BB} - P_{AB}P_{BA}$$

random-walk controversy score (RWC)

- assume graph is partitioned in two sides, A and B
- consider a random walk that started at a random node and finished in a hub in $Y \in \{A, B\}$
- probability that random walk started in $X \in \{A, B\}$

$$P_{XY} = \Pr(\text{r.w. started in } X \mid \text{r.w. finished in } Y)$$

- random-walk controversy score (RWC)

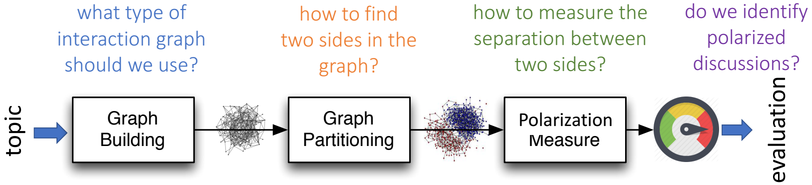
$$\text{RWC} = P_{AA}P_{BB} - P_{AB}P_{BA}$$

does not depend on cluster sizes and relative in-degrees

evaluation

- annotate polarized and non-polarized topics
- polarized
  - indian beefban, nemtsov protests, netanyahu US congress speech, baltimore riots, ukraine
- non-polarized
  - germanwings plane crash, sxsw, mother's day, jurassic world movie, national kissing day
- evaluate different settings on ground truth

best performing setting



- retweet graph
- RWC

other good settings: edge betweenness score

sentiment variance

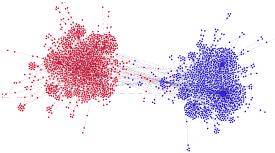
example of results

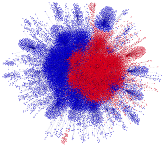
high RWC

low RWC

polarized topics

nemtsov
protests

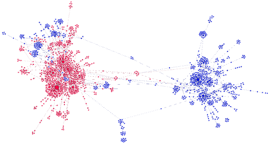


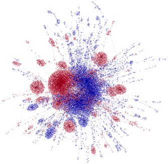


germanwings
plane crash

non-polarized topics

indian
beef ban



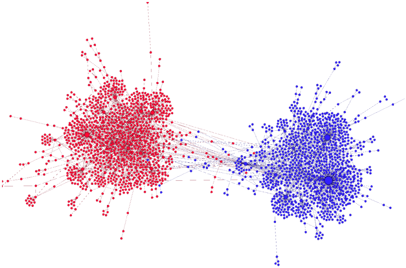


sxsw
conference

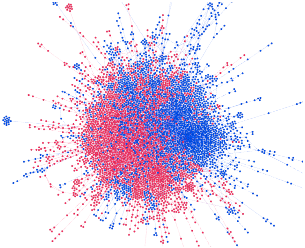
using retweet graph

example of results

nemtsov protests



retweets



replies

research questions

design algorithms to help reducing polarization

design algorithms to moderate online discussions

mitigation action I

improve awareness

P. Lahoti, K. Garimella, A. Gionis, *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*, ACM WSDM 2018

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

intuition

- map users and content in a joint latent ideology space

such that

- 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

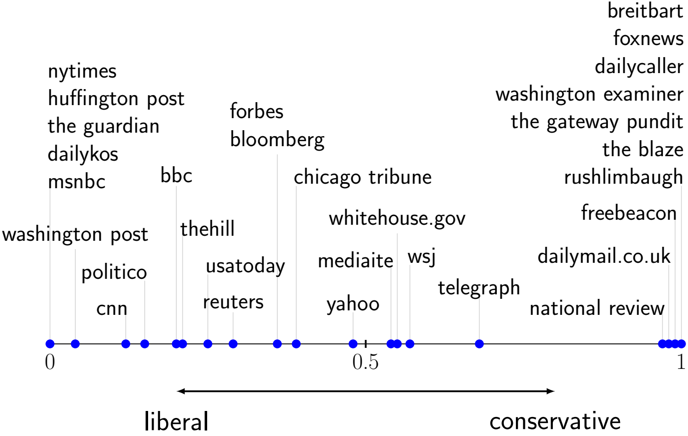
$$\mathbf{A} \approx \mathbf{U}\mathbf{H}_u\mathbf{U}^T \quad \text{and} \quad \mathbf{C} \approx \mathbf{U}\mathbf{H}_v\mathbf{V}^T$$

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

in practice

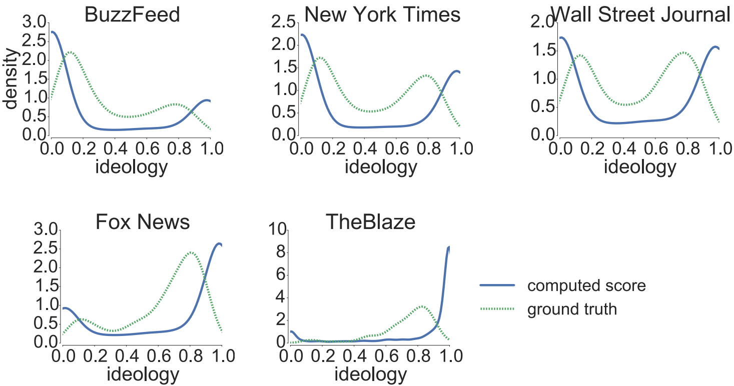
- twitter data from 2011 to 2016, focusing on controversial topics (gun control, abortion, obamacare)
- 6 391 users and 19 million tweets
- user matrix **A** represents follow graph
- content items represent url hostnames
- gather ground-truth polarity scores
  - content polarity [Bakshy et al., 2015]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

content ideology scores



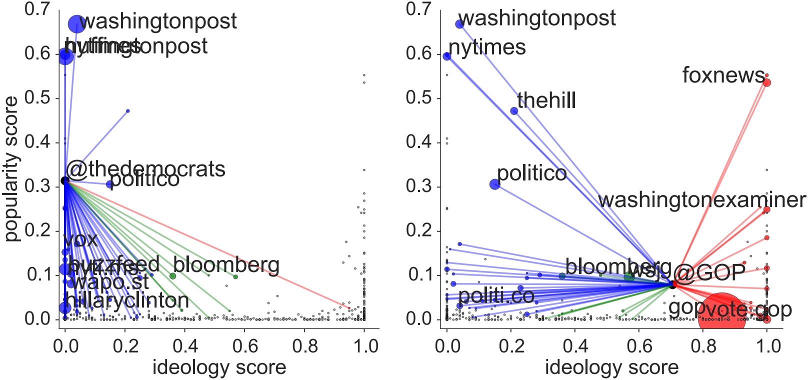
correlation with ground-truth scores 0.82

audience ideology scores



correlation of user ideology scores with ground-truth 0.90

visualizing the information bubble



@thedemocrats

@gop

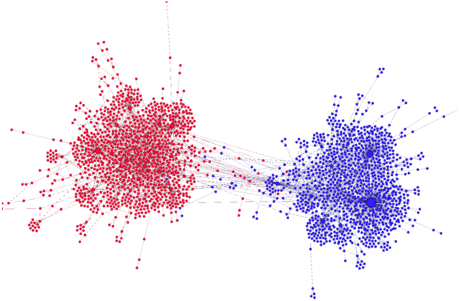
mitigation action II

user-to-user recommendation

K. Garimella, G. De Francisci Morales, A. Gionis, M. Mathioudakis, *"Reducing controversy by connecting opposing views"*, ACM WSDM 2017

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.

reducing polarization

how can we bridge the divide?

- assuming
  - polarization score measured by RWC

$\Rightarrow$ we want to reduce RWC

- problem
  - add k edges that maximally reduce RWC

reducing polarization

- greedy algorithm
  - find the single best edge to reduce RWC
  - repeat k times
- inefficient
  - computing RWC requires $\mathcal{O}(\textsc{mmult}(n))$
    faster in practice with iterative computation
  - still, greedy requires $\mathcal{O}(n^2 \cdot k \cdot \textsc{mmult}(n))$
- improvements
  - consider adding edges only between hubs
  - incremental RWC computation using
    Sherman-Morrison formula

reducing polarization

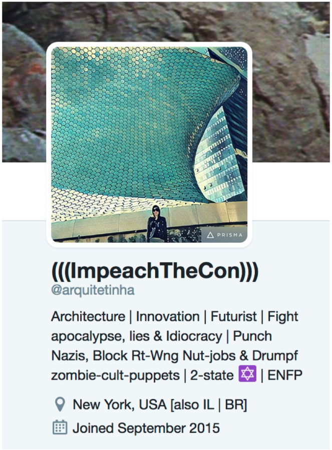
- what does it mean “add k edges”?
- answer: recommendations
- but many recommendations are unlikely to be materialized

  no point recommending D. Trump to retweet H. Clinton

- incorporate probability of accepting a recommendation
  - compute user polarity, and
  - acceptance probability as a function of user polarity

reducing polarization : real example

polarity=-.99



polarity=.95

reducing polarization : real example

polarity=-.99

polarity=.15

reducing polarization : results

	obamacare		guncontrol	
	node1	node2	node1	node2
ROV	mittromney	barackobama	ghostpanther	barackobama
	realdonaldtrump	truthteam2012	mmflint	robdelaney
	barackobama	drudge_report	miafarrow	chuckwoolery
	barackobama	paulryanvp	realalexjones	barackobama
	michelebachmann	barackobama	goldiehawn	jedediahbila
ROV-AP	kksheld	ezraklein	chuckwoolery	csgv
	lolgop	romneyresponse	liamkfisher	miafarrow
	irritatedwoman	motherjones	csgv	dloesch
	hcan	romneyresponse	jonlovett	spreadbutter
	klsouth	dennisdmz	drmartyfox	huffpostpol

mitigation action III

balance information exposure

K. Garimella, A. Gionis, N. Parotsidis, N. Tatti, *"Balancing information exposure in social networks"*, NIPS 2017

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

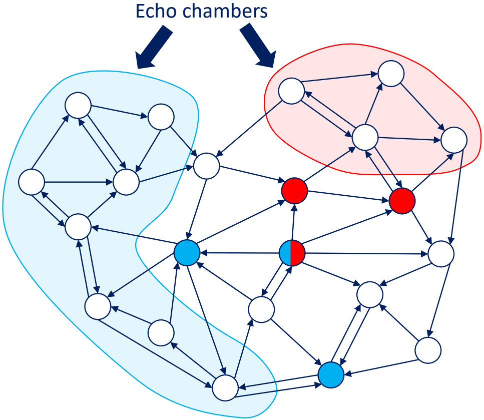
balancing information exposure

- 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

- 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

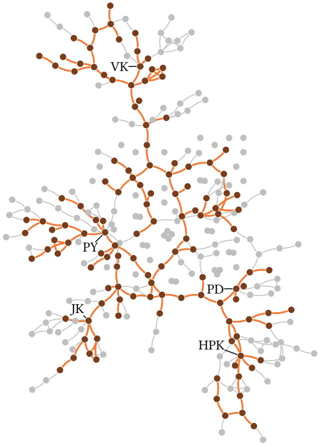
- evidence of echo chambers in social networks
  - price of bi-partizanship
- quantifying polarization in social media
  - random-walk controversy score
- actions to mitigate echo chambers
  - improve awareness
  - user recommendation
  - content recommendation

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

Q & A



credits

Kiran
Garimella

Gianmarco
De Francisci Morales

Michael
Mathioudakis

Preethi
Lahoti

Nikos
Parotsidis

Nikolaj
Tatti