Polarization on Social Media

Kiran Garimella, Gianmarco De Francisci Morales, Aristides Gionis, Michael Mathioudakis

INSTITUTE
FOR SCIENTIFIC INTERCHANGE
FOUNDATION

Aalto University

UNIVERSITY OF HELSINKI

social media

- people use social media to
  - share information, express opinion, comment, interact, discuss, get personalized news feed

- majority of EU citizens get their news from social media

PEW RESEARCH CENTER 2018

social media : good and bad sides

advantages

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

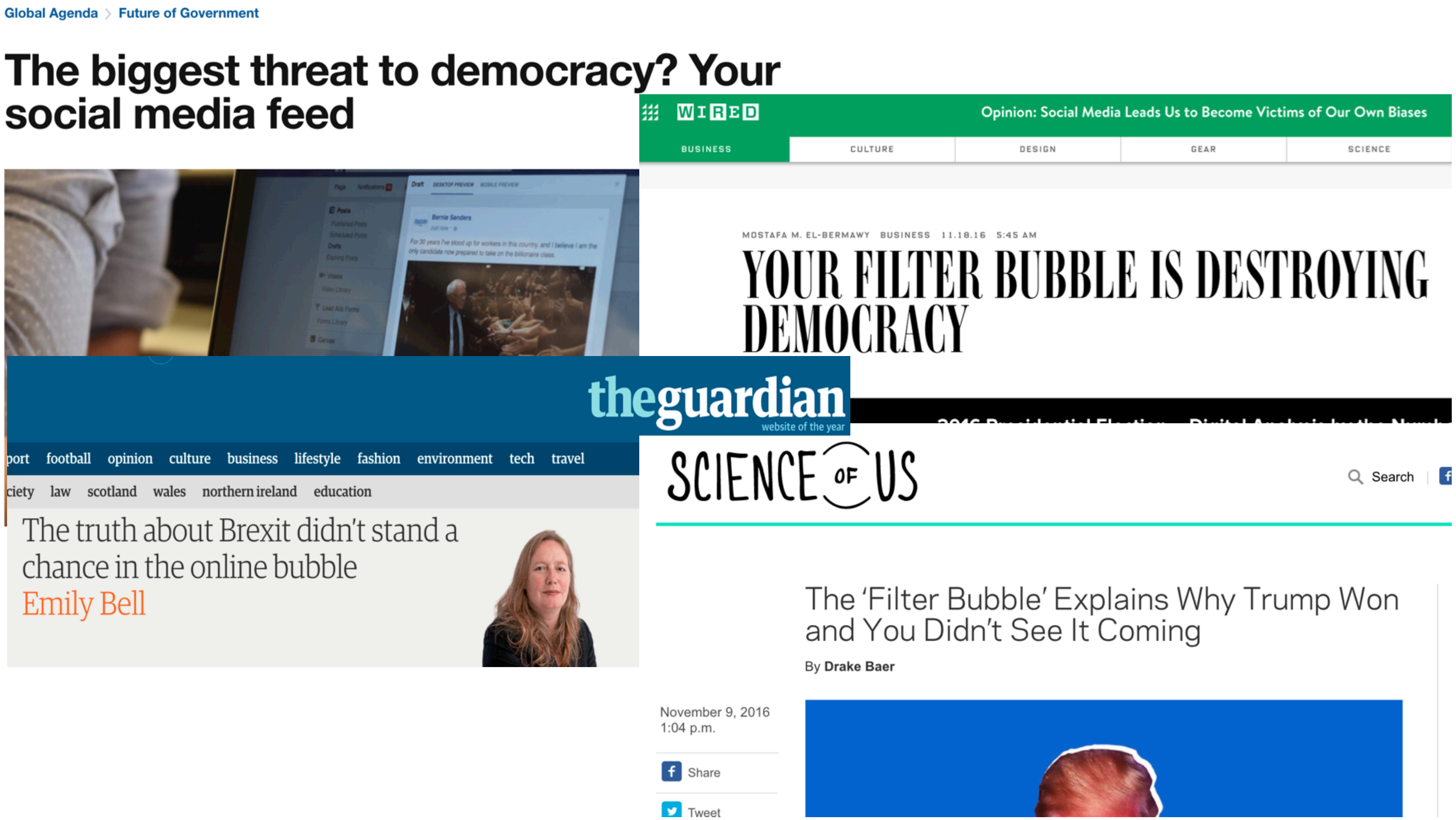
social media : good and bad sides

advantages

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

disadvantages

- harassment
- fake news
- echo chambers
- polarization
- . . .



polarization + online bubble → more polarization

Are 'online bubbles' real?...

...and do they make society more polarized?

What the tutorial is about

- High-level understanding of polarization & related phenomena
- Breadth over depth
- Perspectives from various fields
  - Psychology, social sciences, computer science

What the tutorial is **not** about

- Misinformation / fake news / fact checking

What is polarization?

- The term is used in various domains with similar meaning
- Political polarization (Wikipedia) *"the divergence of political attitudes to ideological extremes."*
  https://en.wikipedia.org/wiki/Polarization_(politics)
- Social polarization *"the segregation within a society that may emerge from income inequality, real-estate fluctuations, economic displacements, etc."*
  https://en.wikipedia.org/wiki/Social_polarization
- Oxford Dictionary *"Division into two sharply contrasting groups or sets of opinions or beliefs."*
  Ref: https://en.oxforddictionaries.com/definition/polarization

Why is it important to study?

- How we handle disagreement is essential to democratic process
  - A large part of the discussion has moved to social media
- Because polarization might be linked to adverse effects
  - Stereotypes
  - Echo chambers
    - Decrease in deliberation
    - Hinders deliberative democracy
- Need to be aware of our biases
  - Sometimes we might not hear opposing views
  - Biases around us (e.g., algorithmic personalization)
- However, not necessarily negative in itself

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Part 2
Exploring Polarization

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Background: Cognitive mechanisms…

- … that manifest when humans are confronted with information that challenges their beliefs
- Polarization involves…
  - … arguments and counter-arguments
  - … evidence that is conflicting or interpreted differently
  - … different points of view – that might challenge our own
- How do we react to opposing opinions / arguments / evidence that challenge their opinion?
  - Do we update our beliefs? How?
  - Are we influenced by the beliefs of others?
  - Do we use evidence to update our beliefs?
  - Or use our beliefs to judge evidence?
- Psychologists & cognitive scientists have studied these questions for long

In this part…

- Part 1: Introduction
- Part 2: Exploring Polarization
  - Cognitive dissonance
  - Why the Web might increase polarization (or not)
  - Studies on the Web
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Cognitive dissonance

- People experience discomfort when presented with information that challenges their beliefs or decisions

    Fischer et al. "The theory of cognitive dissonance: State of the science and directions for future research." 2008.

- Extensively studied behavior, theory first formulated in the 1950's

    Festinger. "A Theory of Cognitive Dissonance." 1957.

Cognitive dissonance

- Cognition: broadly defined
  - Element of knowledge, belief, value
- Dissonance – i.e., lack of harmony or agreement
  - Subjective perception of incompatibility / discrepancy between cognitions
  - Psychological discomfort
  - Motivation to reduce discomfort
- Reduce discomfort by…
  - Adding or highlighting consonant cognitions
  - Removing or downplaying dissonant cognitions

Manifestations of Cognitive Dissonance

- Selective exposure

  Klapper. “The effects of mass communication.” 1960

  - Subjects choose to examine items that agree with their decision

- Biased assimilation

  Lord et al. “Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence.” 1979

  - Subjects find consonant evidence more convincing

- Free choice

  Brehm. “Postdecision changes in the desirability of alternatives.” 1956

  - Spreading-apart of alternatives after decision

- Induced compliance

  Festinger and Carlsmith. “Cognitive consequences of forced compliance.” 1959

  - Subjects justify their decisions a-posteriori, even if they originally disagreed

Selective Exposure to Information

Jonas, Schulz-Hardt, Frey, and Thelen, 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. Journal of personality and social psychology.

- Setting: controlled study

36 students, U.Munich, Germany

1. Difficult question posed to participant
   - 'Difficult': valid, non-trivial options

'should health insurance cover *alternative medicine* methods?'

2. Preliminary decision is made

yes or no

3. Participant is offered additional items of information
   - Items accompanied with description, inspected by participant
   - Description reveals clearly whether item supports decision

(real) articles by experts
8 consonant + 8 dissonant

2-sentence summary

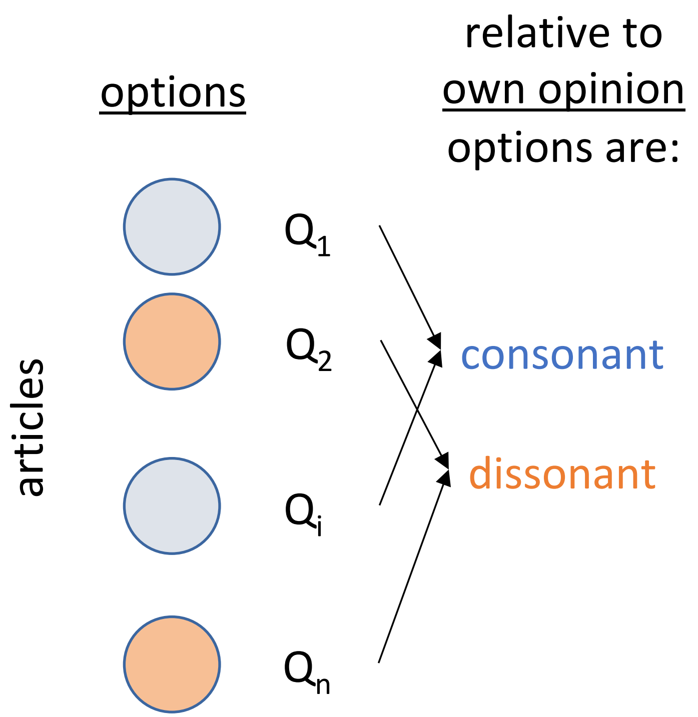
4. Participant chooses some items to examine before final decision

- What items does the participant choose?
- Finding: more items that agree with their decision
  - Confirmation bias

consonant 3.17 ± 1.89
dissonant 2.39 ± 1.79

Selective Exposure to Information

Jonas, Schulz-Hardt, Frey, and Thelen, 2001. Confirmation bias in sequential information search after preliminary decisions: an expansion of dissonance theoretical research on selective exposure to information. Journal of personality and social psychology.



Q: 1 if option is selected, 0 if not

$E[Q_i] = \text{Prob}(\#i \text{ is selected})$

$E[Q_i \mid \#i \text{ is consonant}] > E[Q_j \mid \#j \text{ is dissonant}]$

Biased assimilation

Lord, C.G., Ross, L. and Lepper, M.R., 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of personality and social psychology.

151 students, Stanford

- Setting: controlled study

1. Participant is presented with:
   - a question;
   - two sides for it, in form of study headlines.
2. Participant is asked to give own opinion.
3. Participant is given details of the studies & asked to evaluate how well the study was performed and how convincing it is.

Does capital punishment *deter* crime?

“A study performed by X on data from Y found that capital punishment does [not] deter crime”.

agree or not

- Does the participant find the study well-performed or convincing?
- Finding: More if they agree, less if they disagree.
  - Other finding: polarization increased by end of experiment.

Biased assimilation

Lord, C.G., Ross, L. and Lepper, M.R., 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of personality and social psychology.

Table 1
Evaluations of Prodeterrence and Antideterrence Studies by Proponents and Opponents of Capital Punishment

Study	Proponents	Opponents
Mean ratings of how well the two studies had been conducted		
Prodeterrence	1.5	−2.1
Antideterrence	−1.6	−.3
Difference	3.1	−1.8
Mean ratings of how convincing the two studies were as evidence on the deterrent efficacy of capital punishment		
Prodeterrence	1.4	−2.1
Antideterrence	−1.8	1
Difference	3.2	−2.2

Biased assimilation

Lord, C.G., Ross, L. and Lepper, M.R., 1979. Biased assimilation and attitude polarization: The effects of prior theories on subsequently considered evidence. Journal of personality and social psychology.

	options		relative to own opinion options are:
studies		Q_1	consonant
		Q_2	dissonant

Q: perceived quality of study / convincingness

$E[Q_i \mid \#i \text{ is consonant}] > E[Q_j \mid \#j \text{ is dissonant}]$

Free choice

Brehm, 1956. Postdecision changes in the desirability of alternatives. The Journal of Abnormal and Social Psychology.

- Setting: controlled study
  1. Participant considers a set of items
     - and rates them
  2. Participant is asked to consider 2 of the items
     - Low-dissonance (dissimilar ratings) or high-dissonance (similar ratings)
     - The participant is offered to keep one of the two, and gets to decide
     - Decision is made by participant
  3. Participant is asked to rate the items again

225 students

products, worth 15-30$

rate 1 to 8

high vs low rating

- How does the rating change after the decision?
- Findings:
  - Difference between the two items increases
  - Higher increase with higher dissonance

0.11 for low dissonance
0.79 for high dissonance

Free choice

Brehm, 1956. Postdecision changes in the desirability of alternatives. The Journal of Abnormal and Social Psychology.

	<u>options</u>	relative to <u>own decision</u> options are:
items	Q_1	consonant
	Q_2	dissonant

Q: rating of item

$$E[Q_i^{after} \mid \#i \text{ is consonant}] > E[Q_j^{after} \mid \#j \text{ is dissonant}]$$

$$E[Q_i^{after} \mid \#i \text{ is consonant}] > E[Q_i^{before} \mid \#i \text{ is consonant}]$$

$$E[Q_j^{after} \mid \#j \text{ is dissonant}] < E[Q_j^{before} \mid \#j \text{ is dissonant}]$$

Group biases

- Earlier discussion: bias mechanisms at individual level
- Biases can also manifest at group level
- Social identity complexity
  - Individuals associate themselves with social identities
    - race, religion, gender, class

Roccas, S. and Brewer, M.B., 2002. Social identity complexity. Personality and Social Psychology Review.

- Group polarization
  - The tendency for a group to make decisions that are more extreme than the initial inclination of its members

Sunstein, C.R., 2002. The law of group polarization. Journal of political philosophy.

Wrap-up

- Cognitive dissonance prompts people to expose themselves to confirming information
  - What is consonant or dissonant might also depend on group participation

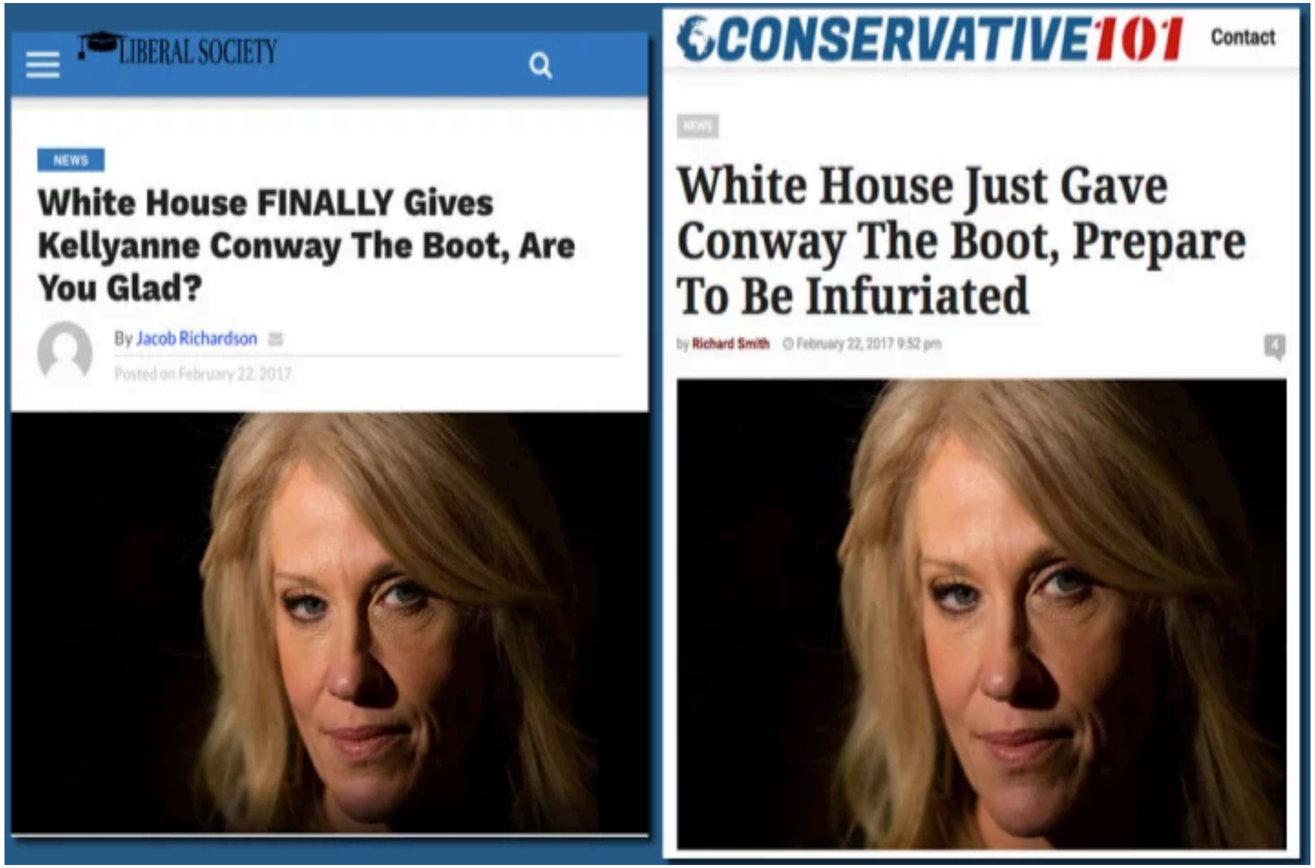
- What could go wrong?

- People share their views on the same platforms they use to consume information
  - Eg: Facebook, Twitter

- If platforms are aware of user views and aim to maximize user satisfaction, what content will they show to users?
  - Why show dissonant content?

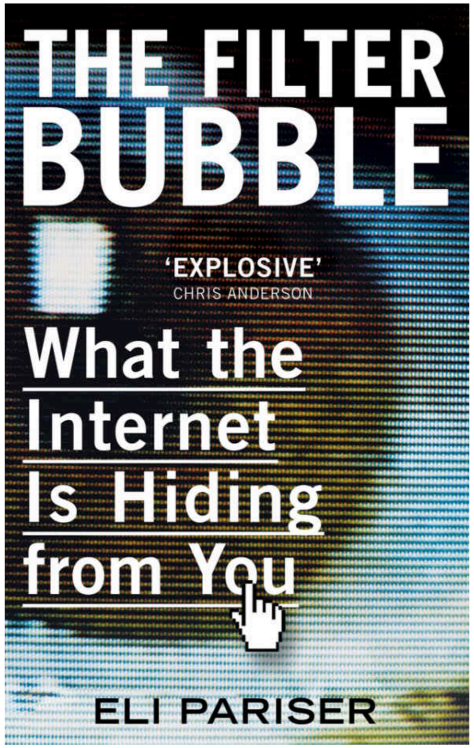
Media bias

- Media present information differently based on their audience

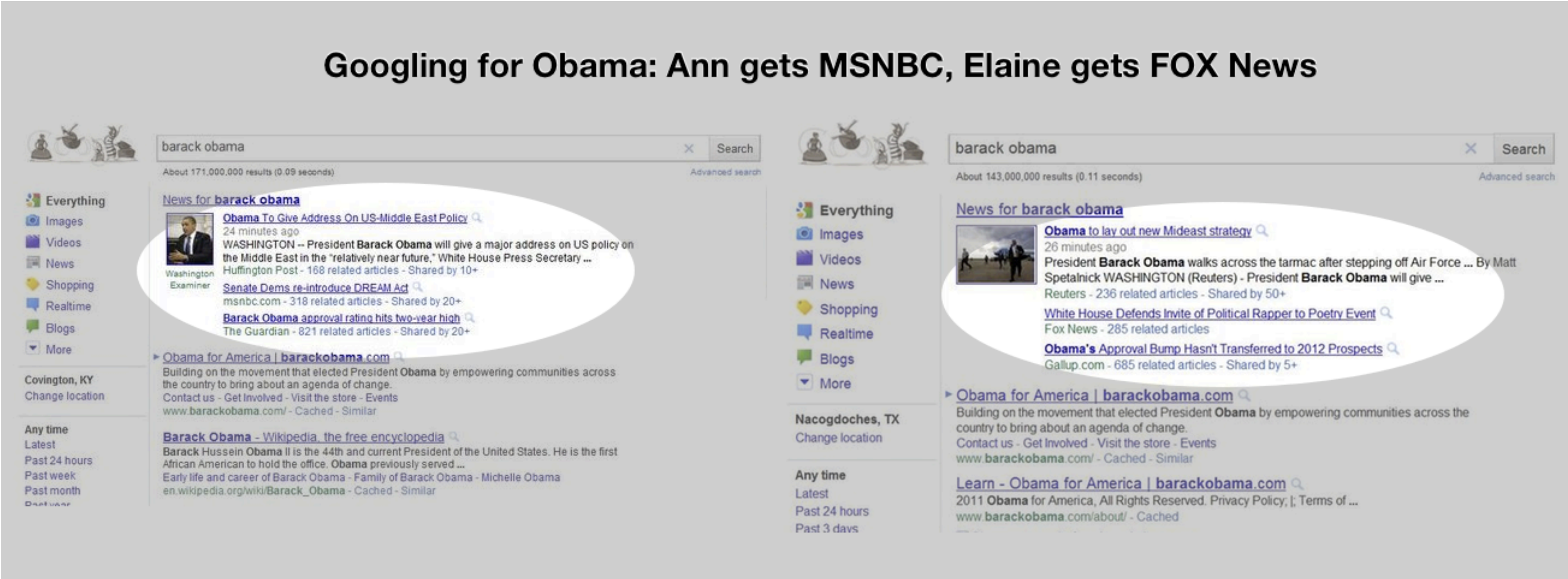


Algorithmic bias

- Online content platforms present information to match individual users
- Algorithmic personalization
  - News
  - Search engines
  - Social media
- Filter bubble
  - We do not see the same content



Filter bubble



Filter bubble

Why the Web might increase polarization

- Increase in available information
- Increase in filtering power
  - People tend to avoid reading conflicting information
- Increase in social feedback (with social media)
  - Homogeneity and group-think reinforced

Echo chambers

Tribal enclaves in which people hear and reinforce their own opinions

In this part…

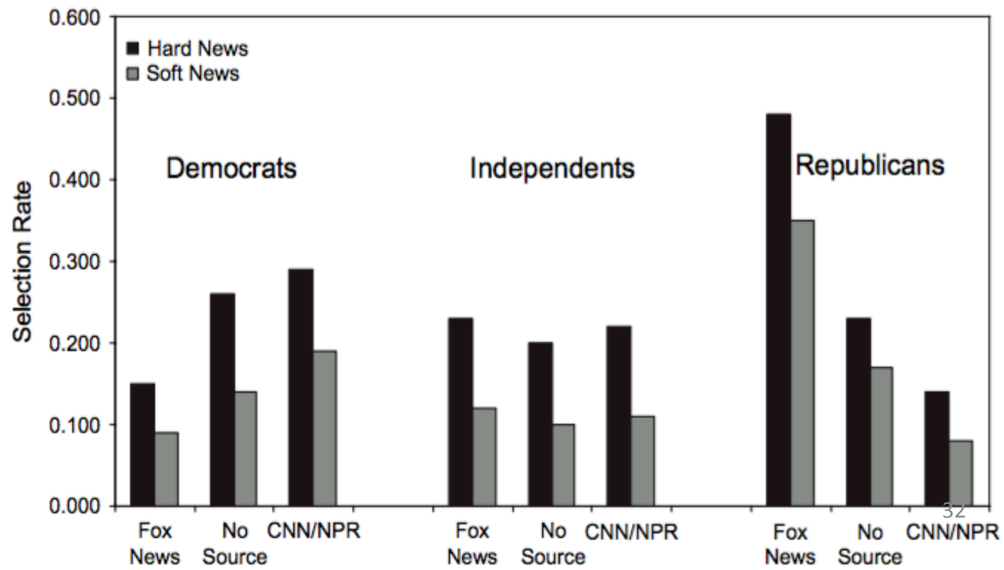
- Part 1: Introduction
- Part 2: Exploring Polarization
  - Cognitive dissonance
  - Why the Web might increase polarization (or not)
  - Studies on the Web
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Ideological Selectivity in Web News

Iyengar, S., & Hahn, K. S. "Red media, blue media: Evidence of ideological selectivity in media use." (2009)

- People prefer to read news from sources close to their leaning
- Finding consistent with selective exposure

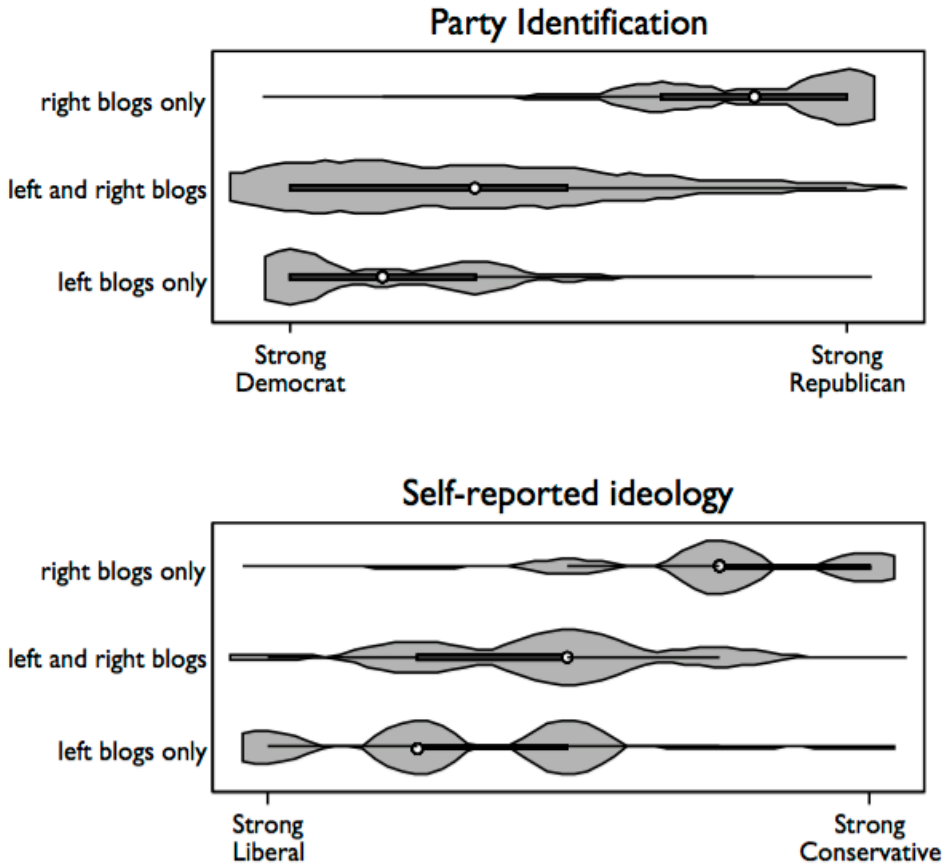
- Online user study with randomized experiments in US
- Headlines for 4 articles, labeled randomly as coming from 4 different sources:
  - Fox News, CNN, NPR, BBC
  - Control group sees same stories with no media logo
- 380 stories, 1020 users
- Tendency to select news based on anticipated agreement as predicted by cognitive dissonance theory
- Effect stronger for hard news



Echo Chambers in Blog Readership

Lawrence, E., Sides, J., & Farrell, H. “Self-segregation or deliberation? Blog readership, participation, and polarization in American politics.” (2010)

- Data from large survey (N=36 000)
- Blog readers are attracted to blogs aligned with their political views (94%)
- Polarization both by party identification and self-reported ideology

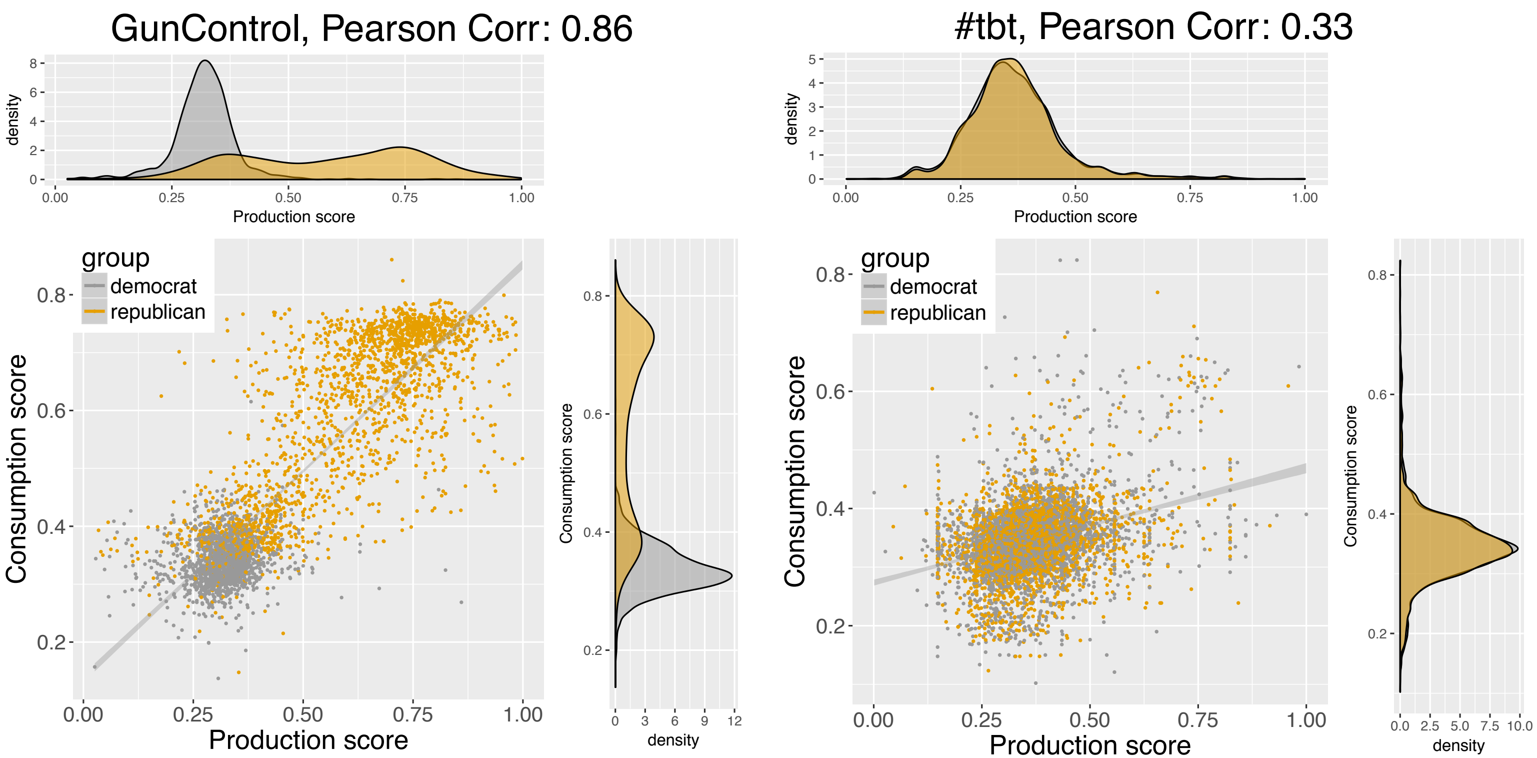


- Finding consistent with selective exposure

Echo Chambers on Twitter

Garimella et.al., “Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship.” WWW2018.

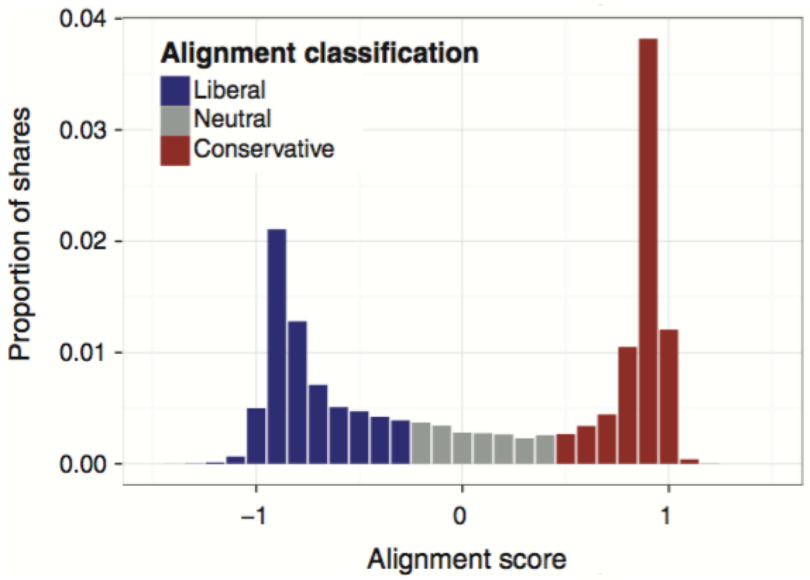
- Fixed set of politically active users
- Set of tweets that mention #topic
- Production vs consumption score
- Main finding: correlation of production and consumption scores
- Finding consistent with selective exposure

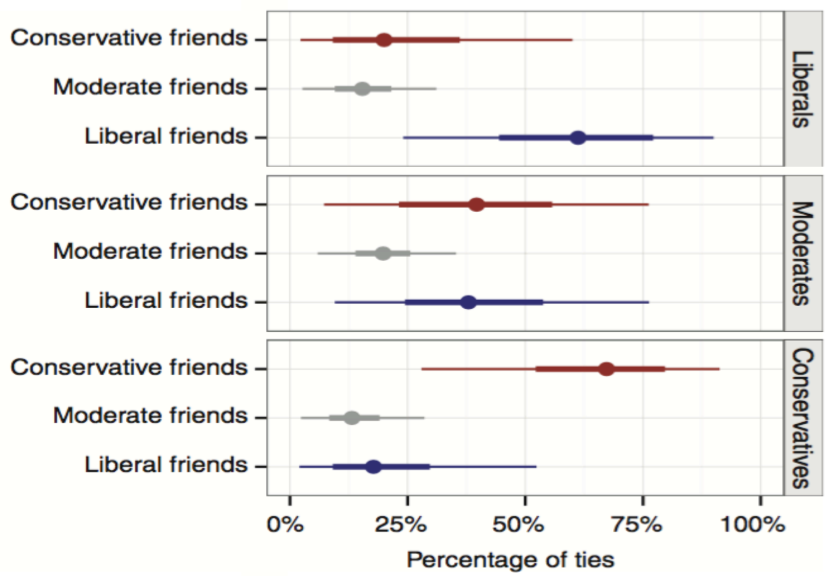


Partisan Exposure on Facebook

Bakshy, E., Messing, S., & Adamic, L. A. "Exposure to ideologically diverse news and opinion on Facebook." (2015)

- US Facebook users with self-reported ideological affiliation
- Analysis on hard news (national news, politics, world affairs)
- Each news item associated with a political alignment
  - Average of the affiliation of users who shared the story
- Cross-cutting news if the alignment of the news and the user differ

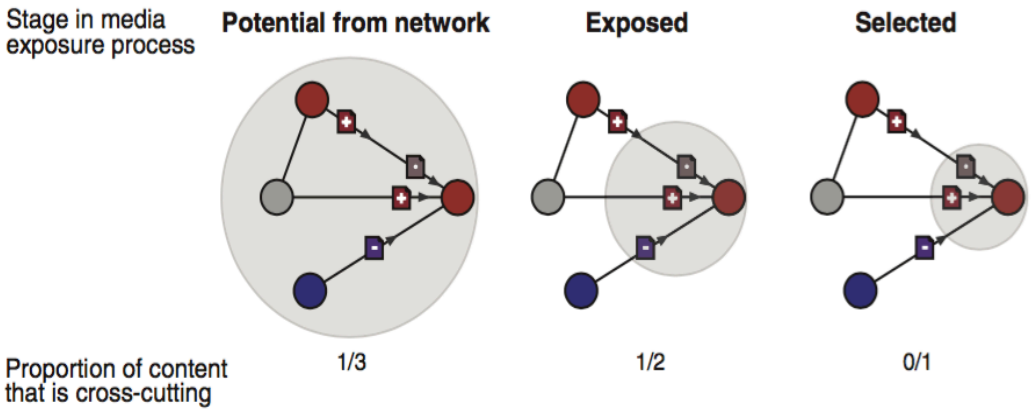


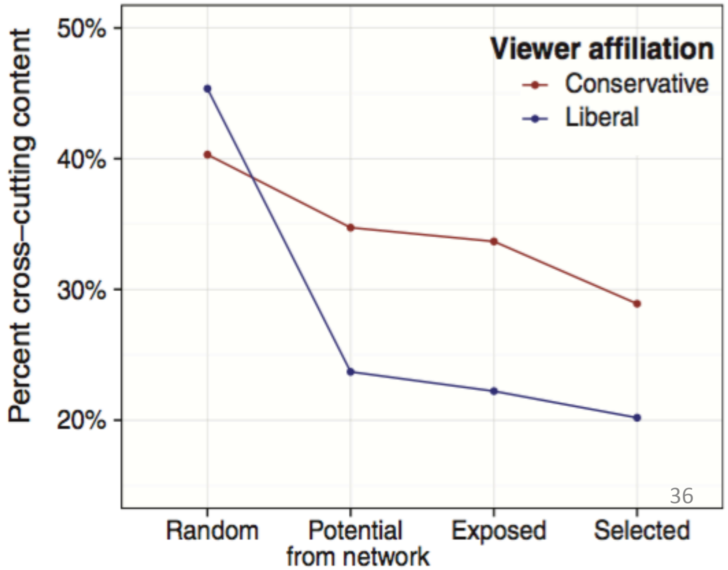


Partisan Exposure on Facebook

Bakshy, E., Messing, S., & Adamic, L. A. "Exposure to ideologically diverse news and opinion on Facebook." (2015)

- Measure the fraction of cross-cutting news among:
  - ones posted in a user's network (potential)
  - ones shown in the user's timeline (exposed)
  - one the user clicked on (selected)
- Compared to random from the whole set, each step reduces the exposure and creates a narrower echo chamber
- Largest reduction from network, rather than algorithmic (filtering), selective exposure still plays a role





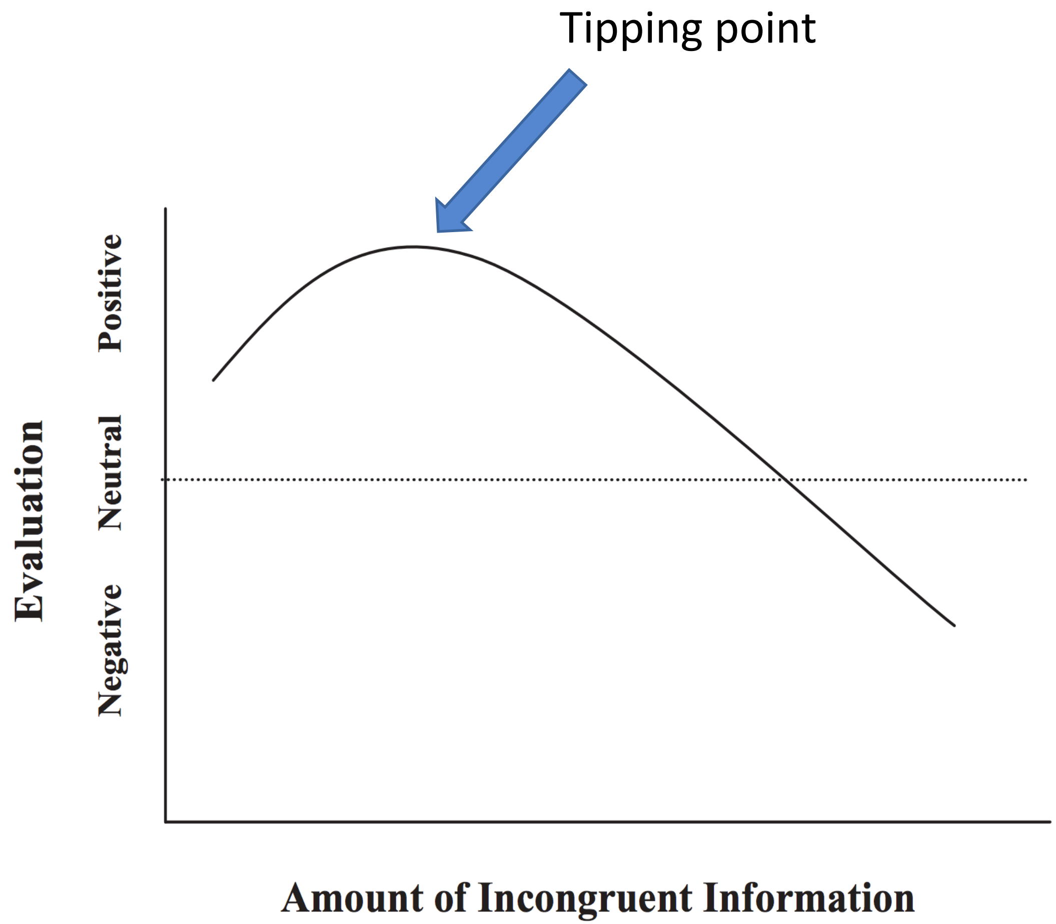
Why the Web might not increase polarization

- Homophily is not observed only for one type of issues (political)
  - The tendency of individuals to associate and bond with similar others
  - Could be based on various facets
    - Gender, age, race, status, religion, geography, beliefs

- Reality kicks in
  - Evidence accumulates at some point

Is there a tipping point?

Redlawsk D, The Affective Tipping Point: Do Motivated Reasoners Ever "Get It"? (2010)



Echo chambers are overstated

Dubois and Blank, The echo chamber is overstated: the moderating effect of political interest and diverse media, 2018

- Data:
  - Nationally representative sample from UK (N=2000)
  - Online and offline news consumption behavior
- Findings:
  - Only a small segment of the population are likely to find themselves in an echo chamber
  - Single media studies are flawed because they do not test the theory in the realistic context of a multiple media environment

The effect of filter bubbles is overstated

Dutton et al. Search and politics: The uses and impacts of search in Britain, France, Germany, Italy, Poland, Spain, and the United States. 2017

- Large scale representative survey (N=14000) from 7 countries.
- The argument that personalization creates filter bubbles is overstated.
- In fact, internet users encounter diverse information across multiple media, which challenges their viewpoints.
- Most users aren't silenced by contrasting views; nor do they silence those with whom they disagree.
- News about fake news has created unjustified levels of concern; people use search to check facts and the validity of information found on social media or the internet.

Internet doesn't accentuate polarization

- Facilitates cross-ideology interactions

  Pablo Barbera. *How Social Media Reduces Mass Political Polarization. Evidence from Germany, Spain, and the U.S*. Psychological science, 2014.

  Kyle A Heatherly, et al*. Filtering out the other side? Cross-cutting and like-minded discussions on social networking sites*. New media & Society, 2017

- Social endorsements more important than partisan source affiliation. Social media facilitates such social endorsements and hence not a cause of polarization

  Messing, S., & Westwood, S. J.. Selective exposure in the age of social media: Endorsements trump partisan source affiliation when selecting news online. Communication Research*. 2014.*

- Fosters potential for deliberation

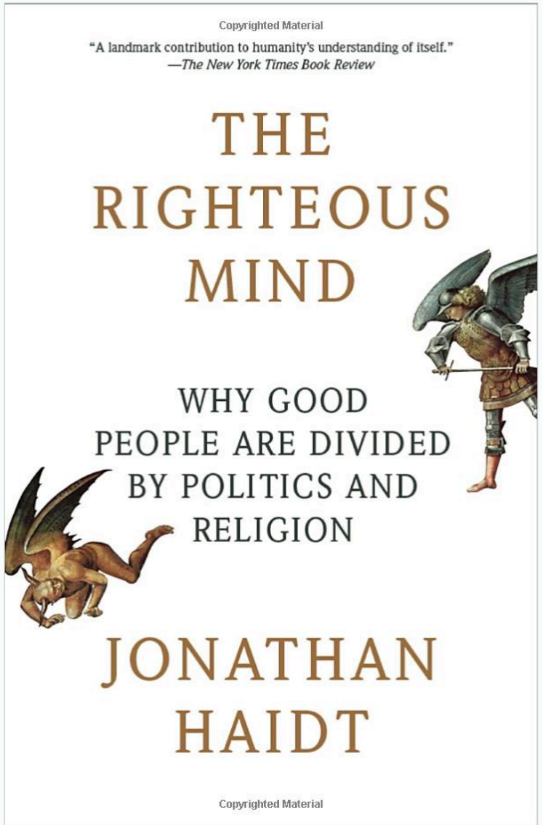
  Wojcieszak, M. E. and D. C. Mutz. "Online Groups and Political Discourse: Do Online Discussion Spaces Facilitate Exposure to Political Disagreement?" In Journal of Communication. 2009.

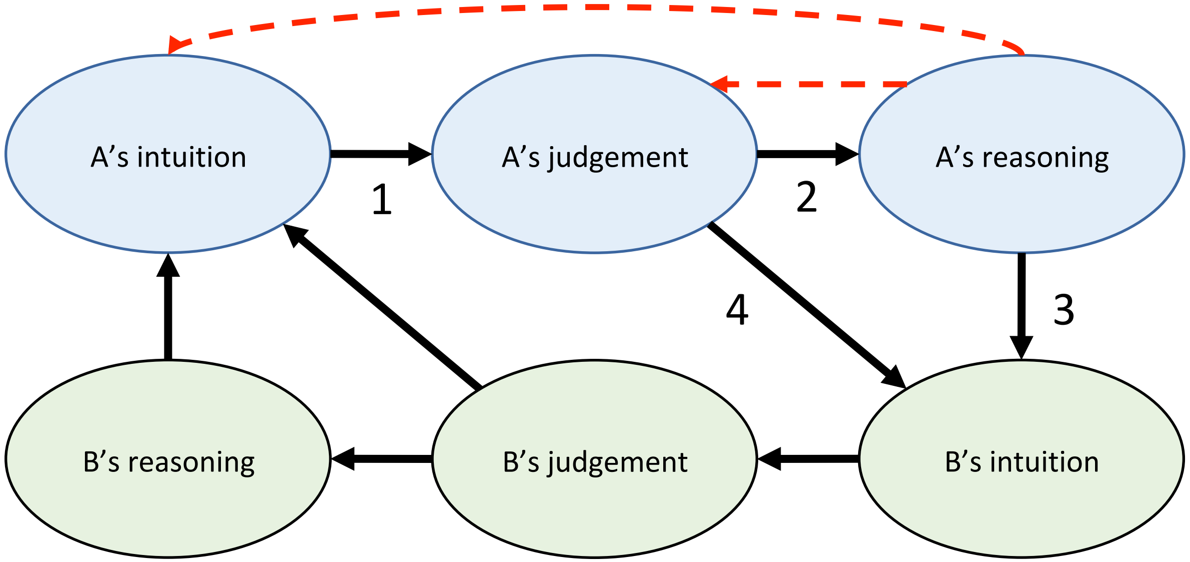
- Polarization is due to user choice and not media

  Bakshy, et al., "Exposure to ideologically diverse news and opinion on Facebook", Science (2016)

  Arceneaux, K., & Johnson, M.. *Changing minds or changing channels? Partisan news in an age of choice*. University of Chicago Press. 2013.

Human reasoning





<u>Main links</u>
(1) Intuitive judgement
(2) post-hoc reasoning
(3) reasoned persuasion
(4) social persuasion (influence)

<u>Rarely used links</u>
(5) Reasoned judgement
(6) Private reflection

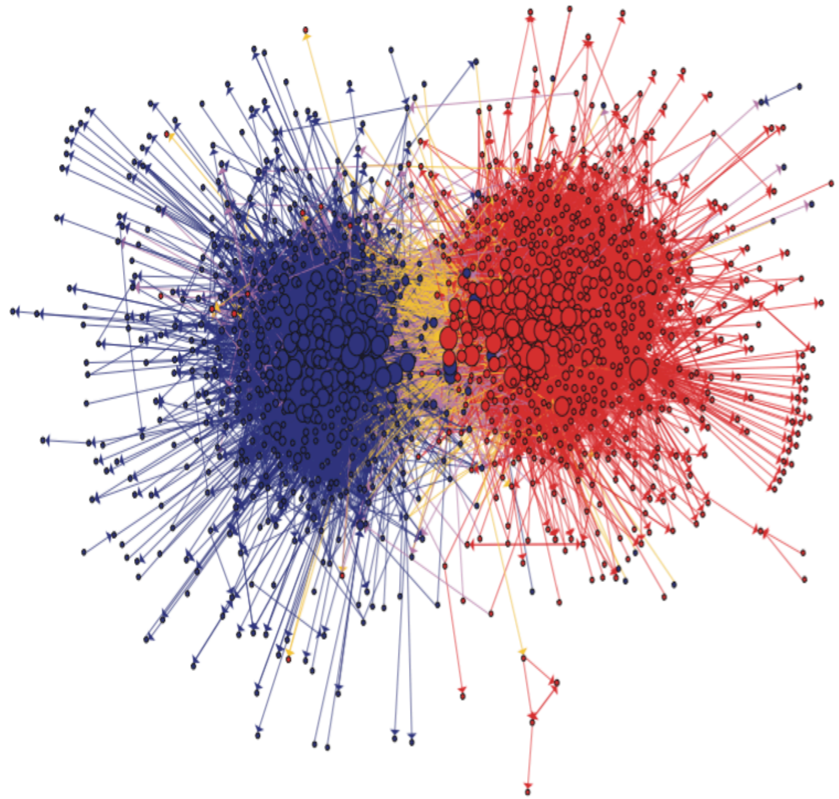
End of Part 2

In what follows, we discuss again:

- Selective exposure
- Biased assimilation
- Algorithmic bias
- Social feedback

Echo Chambers in Blogs

- Studies blog writing
- Political blogs during 2004 US presidential election
- Liberal and conservative blogs link to different news sources (selective exposure + media bias)
- Blogs mostly link internally to the same side (echo chambers due to homophily)
- Conservative blogs link more and more densely within the community
- Cross-community links used to argue (similar to Twitter mentions)



Adamic, L. A., & Glance, N. "The political blogosphere and the 2004 US election: divided they blog." (2005)

Part 3
Polarization Models

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Opinion-formation models

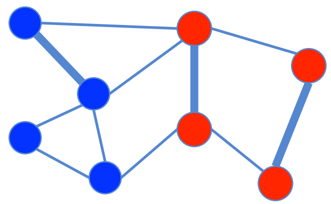
- What process could give rise to consensus or polarization?
- Initial/intrinsic opinions + influence / other dynamic factors
- Rich literature
- Special cases of opinion dynamics
- How are some of the previous concepts captured in opinion-formation models?
- Some opinion-formation models capture polarization – some don't

Outline

- Part 1: Introduction
- Part 2: Exploring polarization
- Part 3: Polarization Models
  - Basic opinion formation model
  - Individual biases
    - Selective exposure
    - Homophily
    - Biased assimilation
  - Group biases
    - Social identity
  - System biases
    - Algorithmic bias
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Conclusions & Future Work

Preliminaries

- Agent-based models (dynamic process in discrete time t)
- Opinions b_i can be continuous or binary ,
  - Opinions change in time $b_i(t)$
- Agents organized in a network G(V,E)
  - Agent i has a set of neighbors N(i)={j | (i,j) in E}



- Influence modeled as weight between agents w_{ij}
  - Weights may change with time or with opinions $w_{ij}(t, b(t))$
  - Usually assume $\Sigma_j w_{ij}=1$ (stochastic matrix)
  - w_{ii} models tendency to keep existing opinion
- Agents interact pairwise (asynchronous) or all-at-once (synchronous)
  - Update their beliefs as a result of the interaction

DeGroot's opinion-formation model

DeGroot. "Reaching a consensus." Journal of the American Statistical Association. 1974.

- DeGroot proposes that individuals update their opinion in each step to the weighted average of their neighbors' opinions and their own opinions in the previous step

$$b_i(t+1) = \sum_j w_{ij} b_j(t)$$

- Social graph models homophily (stronger influence among peers)
- Repeated-averaging process expresses social influence

Polarization in DeGroot's model

Dandekar, Goel, and Lee. "Biased assimilation, homophily, and the dynamics of polarization." PNAS 110.15. 2013.

- Given an opinion vector ***b***, define the network disagreement index (NDI) as

$$NDI(b) = \sum_{i,j} w_{ij}(b_i - b_j)^2$$

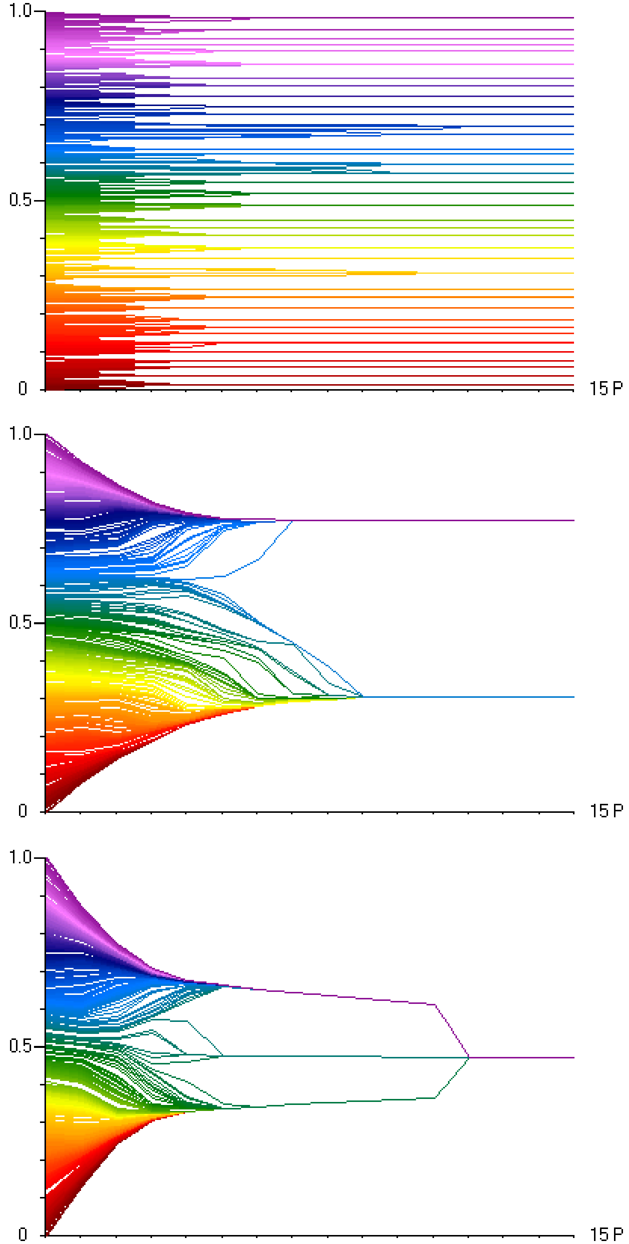
- Each term $w_{ij}(b_i - b_j)^2$ in NDI is disagreement cost imposed upon i and j
- Result: DeGrout's process is not polarizing
  - Disagreement index at time t+1 is no larger than that at time t
- Lemma: $NDI(b(t+1)) \leq NDI(b(t))$

Bounded confidence model (BCM)

Deffuant, Neau, Amblard, and Weisbuch. “Mixing beliefs among interacting agents.” Advances in Complex Systems. 2000.
Krause. “A discrete nonlinear and non–autonomous model of consensus formation.” Communications in difference equations. 2000.

- Agents only interact and update their opinions if the difference between their existing opinions is smaller than a threshold ε
- This threshold models the “openness to discussion”
- Large ε produce consensus, while smaller ε produce polarized opinions
- The threshold ε can be thought as a form of selective exposure

Bounded confidence



ε=0.02

ε=0.3

ε=0.5

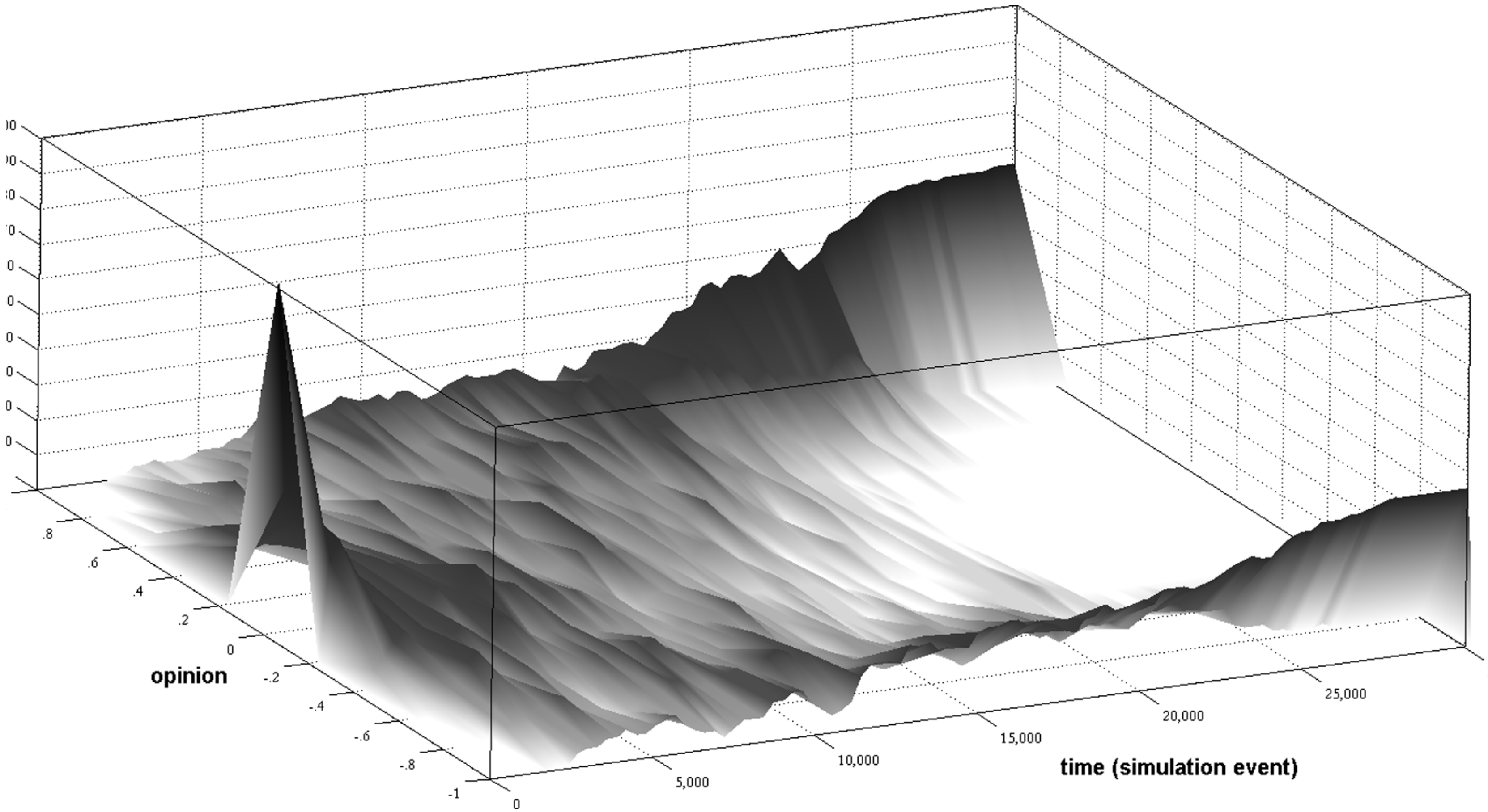
Homophily in interaction

Mäs and Flache. "Differentiation without distancing. Explaining bi-polarization of opinions without negative influence." PloS one. 2013.

- Due to homophily, individuals with opinions leaning towards the same pole of the opinion spectrum interact more likely with each other than with those who lean towards opposite poles
- Non-uniform probability of selecting j as a pair for i, depending on the distance between their opinions
- Modulated by h, the strength of homophily / selective exposure

$$p_i(j) = \frac{\left(1 - \frac{1}{2}\left|b_i - b_j\right|\right)^h}{\sum_{k \neq i}\left(1 - \frac{1}{2}\left|b_i - b_k\right|\right)^h}$$

Effects of homophily



Biased assimilation can lead to polarization

Dandekar, Goel, and Lee. “Biased assimilation, homophily, and the dynamics of polarization.” PNAS 110.15. 2013.

- Modify DeGroot’s model to explicitly incorporate biased assimilation
- In particular, modify weighted average to be non-linear
  - Neighbors with similar opinions are weighted more
  - So, opinions of individuals are reinforced by like-minded neighbors
- Opinion formation model with biased assimilation can lead to polarization
- Under certain conditions:
  - Opinion of moderate individuals can go to extremes (0 or 1)
  - Network disagreement index can increase with time t

Biased assimilation can lead to polarization

Dandekar, Goel, and Lee. "Biased assimilation, homophily, and the dynamics of polarization." PNAS 110.15. 2013.

- Update for agent i interacting with agent j ($b_i \in [0,1]$)

$$\frac{w_{ii} b_i + b_i^{\beta} s_i}{w_{ii} + b_i^{\beta} s_i + (1 - b_i)^{\beta} (1 - s_i)}$$

- β = strength of the bias (β=0 is DeGroot's model)
- s_i : weighted opinion of neighbors
- Model polarizing for β>1

Other variants

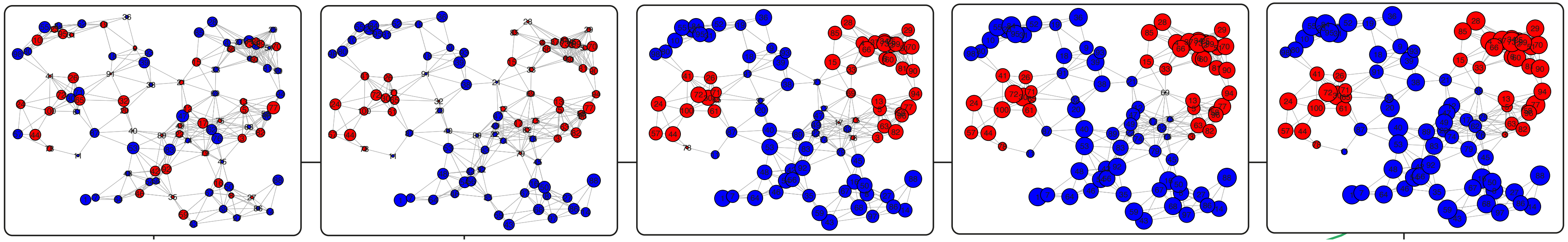
- Axelrod’s model
  - Axelrod. “The dissemination of culture: A model with local convergence and global polarization.” Journal of conflict resolution. 1997.
  - Vector beliefs, interaction probability ≈ similarity, square lattice network
- Negative social influence (BCM)
  - Flache, and Macy. “Small worlds and cultural polarization.” The Journal of Mathematical Sociology.2011.
  - Scarce empirical evidence of this phenomenon
- Media influence (BCM)
  - Quattrociocchi, Caldarelli, Scala. “Opinion dynamics on interacting networks: media competition and social influence.” Scientific reports. 2014.
  - Media opinion evolves via signed media relationship network (friend/foe)
- Propaganda and extremism (BCM)
  - Timothy. “How does propaganda influence the opinion dynamics of a population?” arXiv:1703.10138. 2017.
  - Agents create pockets of radicalization (pro and against the propaganda)
- Internal belief networks
  - Rodriguez, Bollen, and Ahn. “Collective dynamics of belief evolution under cognitive coherence and social conformity.” PloS one. 2016.
  - Minimize cognitive dissonance (unstable triads), both internally and across peers

Polarization via Social Feedback

Banisch and Olbrich. "Opinion Polarization by Learning from Social Feedback." arXiv:1704.02890. 2017.

- Different mechanism of opinion formation based on social feedback, and related to social identity
- Agents form opinion via social feedback on their expressed opinion
- $Q_i(b,t) \in \Re$ captures how well opinion b is received by the social network of i at time t
- Updates: $Q_i(b,t+1) = (1-\lambda)Q_i(b,t) + \lambda b_i(t) b_j(t)$
- Agent i expresses opinion with highest Q_i (with conviction ΔQ_i) with some small deviation ε (exploration rate)
- Captures group polarization: agents in homogeneous neighborhood approach maximal conviction, even if initially weakly convinced

Polarization via Social Feedback



Polarization via Algorithmic Bias

Sîrbu, Pedreschi, Giannotti, and Kertész. "Algorithmic bias amplifies opinion polarization: A bounded confidence model." arXiv: 1803.02111. 2018.

- Modify bounded confidence model to include algorithmic bias γ
- Enhanced probability of picking a pair whose opinion is within threshold ε

$$p_i(j) = \frac{\left|b_i - b_j\right|^{-\gamma}}{\sum_{k \neq i} \left|b_i - b_k\right|^{-\gamma}}$$

- Models online media which suggests interaction with similar peers
- Increased tendency towards polarization
  - Polarization in cases where the original model produces consensus
- Slower convergence towards consensus

Polarization via Algorithmic Bias

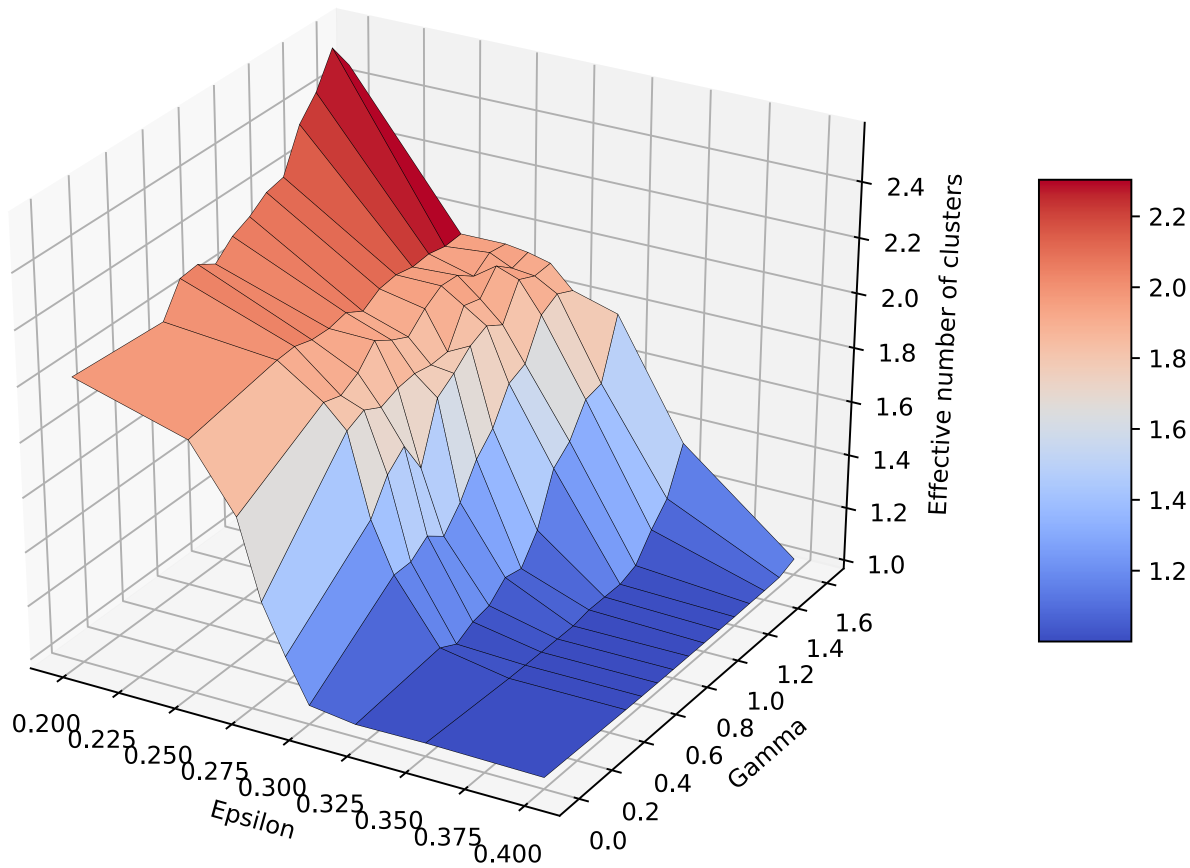
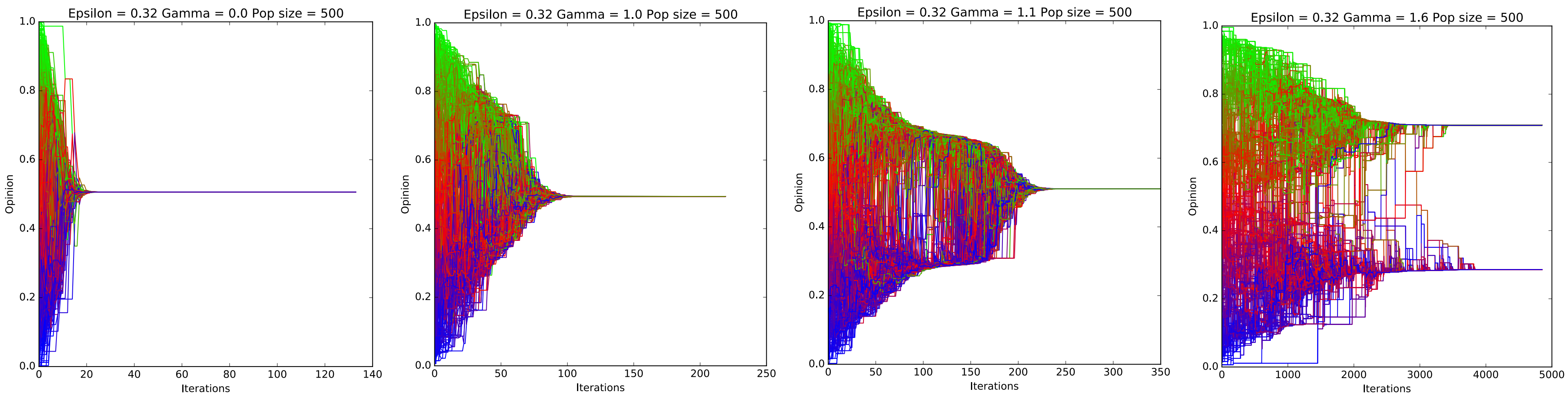


Figure 1: **Number of clusters obtained for various ε and γ.**

Polarization via Algorithmic Bias



same ε = 0.32, increasing γ (0 .. 1.6)

Part 4
Measuring Polarization

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
  - Identifying and Quantifying
  - Content vs. Network based methods
  - User polarization
- Part 5: Mitigating Polarization
- Part 6: Challenges and Directions for Future Research

Why do we want to do this?

- To model and understand social processes
- To reduce polarization
- To create a balanced news diet
- To design recommender systems

Identifying polarized topics

- Can we identify a polarized discussion?
- How polarized is a discussion
  - Axioms of polarization
- Distribution over some “attribute” (Likert scale)

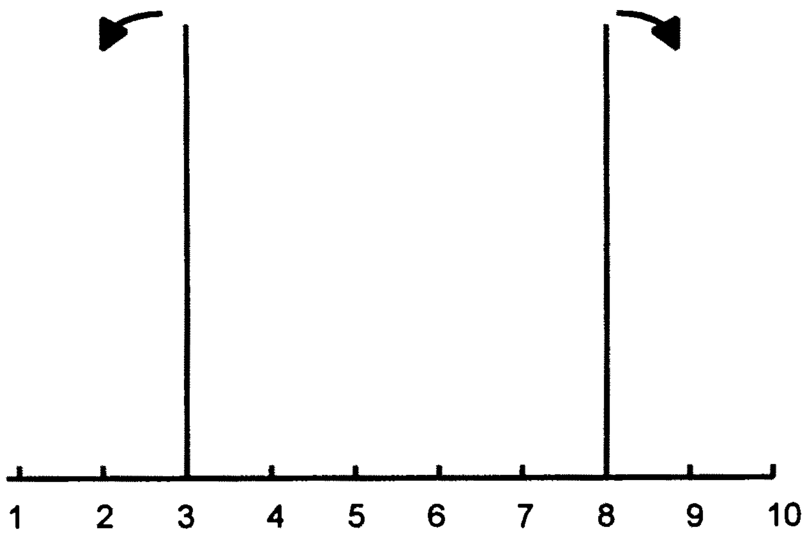


FIGURE 1A

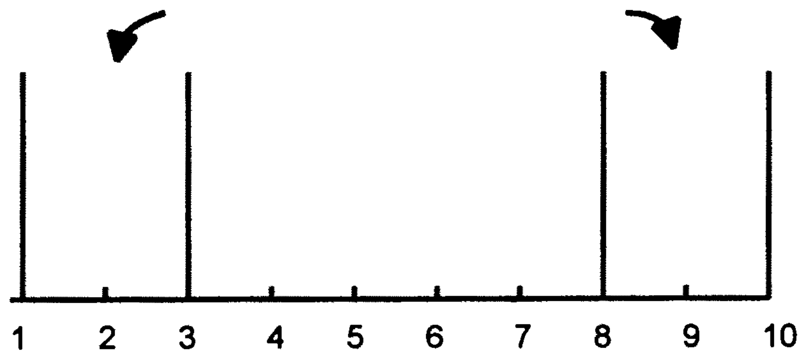


FIGURE 1B

Measurement of polarization

Esteban and Ray. “On the measurement of polarization.” Econometrica. 1994.

- On the Measurement of Polarization
  - Axioms of polarization
  - “clustered distribution”
  - Related but different from economic inequality
    - Similar to Gini coefficient (Lorenz curve)
    - Takes into account antagonism (pairwise difference)

FEATURE 1: There must be a high degree of homogeneity *within* each group.

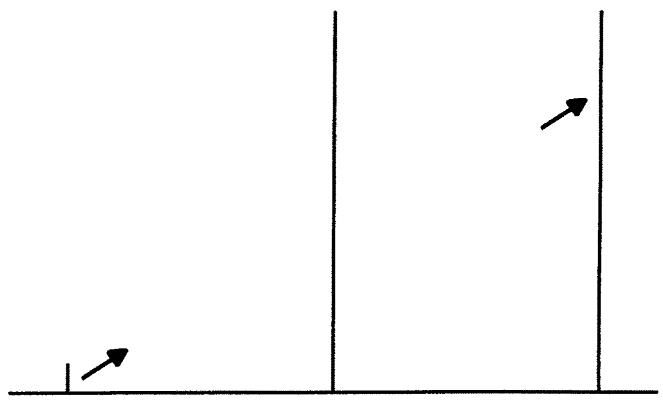
FEATURE 2: There must be a high degree of heterogeneity *across* groups.

FEATURE 3: There must be a small number of significantly sized groups. In particular, groups of insignificant size (e.g., isolated individuals) carry little weight.

Measurement of polarization

Esteban and Ray. “On the measurement of polarization.” Econometrica. 1994.

- An example where polarization measure is different than economic inequality measure



- Consider a small move of mass from extreme left to extreme right
  - (Arguably) polarization increases
  - Economic inequality decreases

Defining polarization is hard

Bremson, Grim, Singer, Fisher, Berger, Sack, and Flocken. "Disambiguation of social polarization concepts and measures." Journal of Mathematical Sociology. 2016.

- Definition of polarization:
  - nine senses
  - different definitions based on the domain
- Distribution of attitudes
  - histogram of the number of individuals holding a specific attitude value along the spectrum

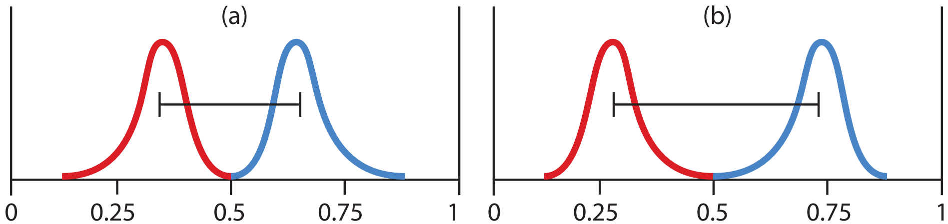


Figure 2. Distribution (b) shows greater polarization in the sense of spread than does distribution (a).

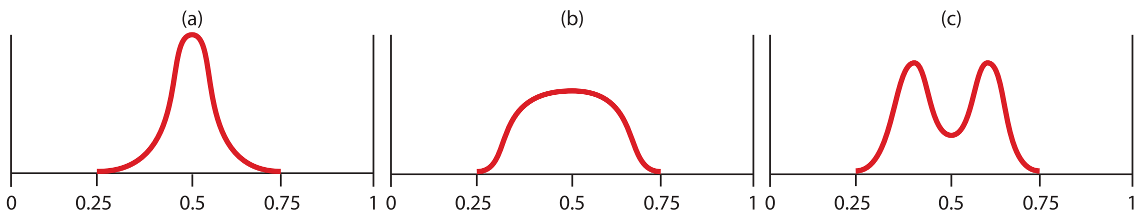


Figure 3. Distribution (c) shows greater polarization in the sense of dispersion than does distribution (b), which is greater than distribution (a).

Identifying Polarization - Content

Sentiment variance in news

Choi, Jung , and Myaeng. “Identifying controversial issues and their sub-topics in news articles.” PAW-ISI 2010.
Garimella, De Francisci Morales, Gionis, and Mathioudakis. “Quantifying Controversy in Social Media.” WSDM 2016.
Klenner, Amsler, Hollenstein, and Faaß. “Verb Polarity Frames: a New Resource and its Application in Target-specific Polarity Classification.” KONVENS 2014.

- Controversial topic - a concept that invokes conflicting sentiments
- Subtopic - factor that gives a particular sentiment (positive or negative)
- Assumption - a controversial topic receives contrasting sentiment
  - positive vs. negative feelings, pros vs. cons, rightness vs. wrongness in their judgments
- Similar results observed by
  - Garimella et al. WSDM 2016
  - Klenner et al. KONVENS 2014

Sentiment variance

- Method:
  - Identify candidate entities (noun phrases)
  - Compute sentiment in sentences involving these entities
  - Controversial if positive_sentiment + negative_sentiment > δ and |positive - negative| > Υ

Issue: ***Afghanistan War***

September 11	positive	The Afghanistan war launched after the September 11 …
	negative	The Afghanistan war was of revenge by the Americans for September 11
Troops	negative	Most Americans oppose sending more troops to Afghanistan war …
Weapons of mass destruction	negative	The Afghanistan war is perilous because of weapons of mass destruction …
Obama	positive	Obama supports the Afghanistan war …

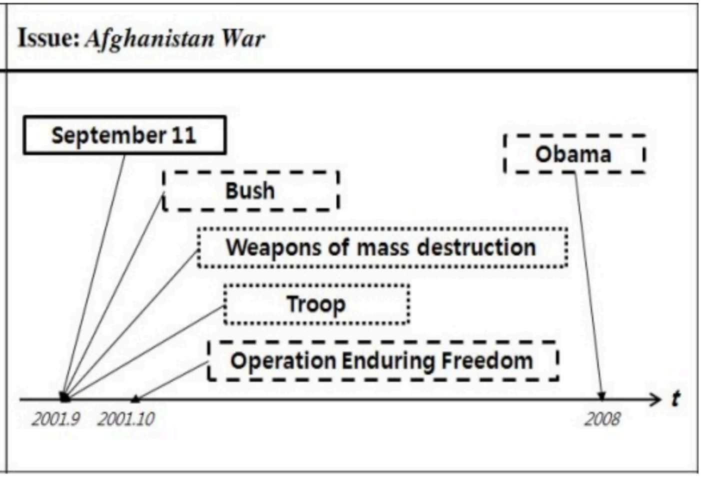
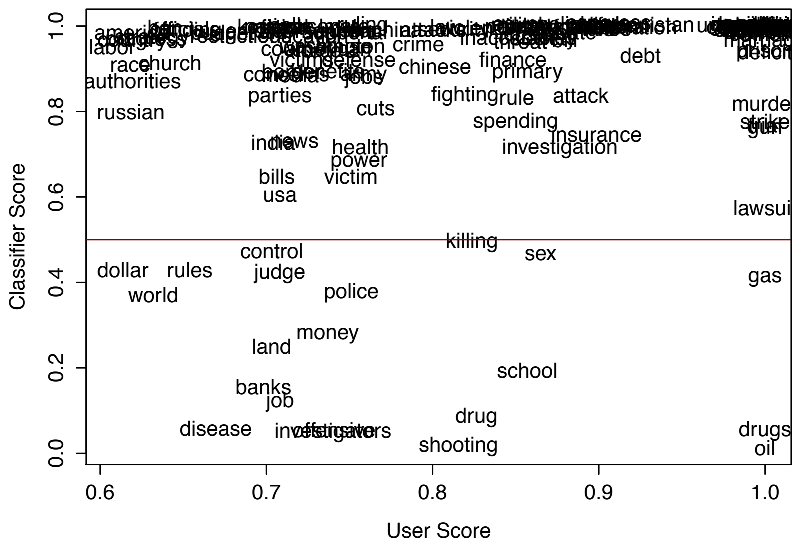


Fig. 1. A summary of the sentiment-generating subtopics for an issue "*Afghanistan War*"

Controversy language in news

Mejova, Zhang, Diakopoulos, and Castillo. “Controversy and Sentiment in Online News.” C+J Symposium 2014.

- Controversy lexicon
- Controversial topics have:
  - strongly biased terms
  - more negative terms
  - fewer strongly emotional terms
- “we show that we can indicate to what extent an issue is controversial, by comparing it with other issues in terms of how they are portrayed across different media.”



(b) Controversial words; correctly classified words appear above the horizontal line.

Figure 2: Scores of controversial and non-controversial words including classification errors. “User score” is the confidence with which the manual labeling was done (with at least 7 annotators per element), while “classifier score” is the output of the classifier on the training data.

Detecting controversy on the Web

Dori-Hacohen and Allan. "Detecting Controversy on the Web." CIKM 2013.
Jang, Foley, and Allan. "Probabilistic Approaches to Controversy Detection." CIKM 2016.

- Find out if a Web page discusses a (known) controversial topic
- Map topics (named entities) in a Web page to Wikipedia articles
  - A Web page is controversial if it is similar to a controversial Wikipedia article
  - E.g., If a news article mentions Abortion it is controversial
- Related:
  - There is a lot of work on identifying controversial topics on Wikipedia
  - Edit wars, hyperlink structure, etc.
- Related:
  - Jang et al. show that in addition to this, language models can be built to directly detect controversy

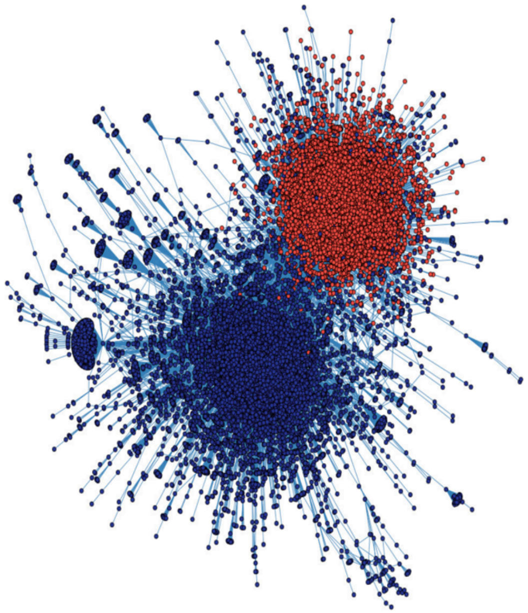
Identifying polarization - Network

- Methods based on network structure
  - Social media, hyperlinks
- Twitter
  - Retweet
  - Reply
  - Social (follow)
- Idea: Controversial topics have a clustered structure in their discussions

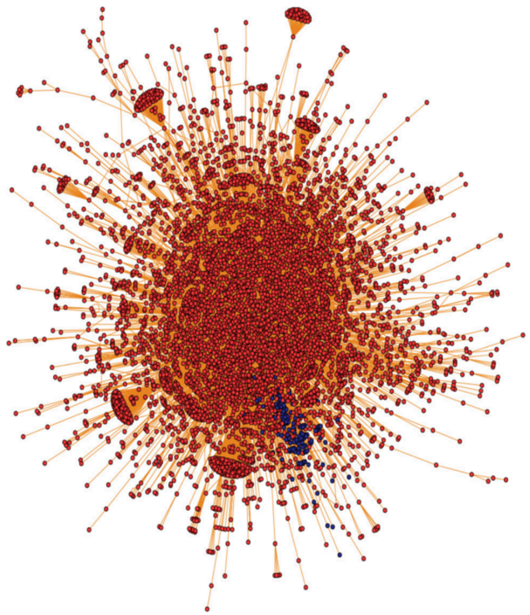
Political polarization on Twitter

Conover, Ratkiewicz, Francisco, Gonçalves, Menczer, and Flammini. "Political Polarization on Twitter." ICWSM 2011.

- Retweet network for political hashtags has a bi-clustered structure
  - Retweet network exhibits a highly modular structure, segregating users into two homogenous communities corresponding to the political left and right
- Users mention/reply to others from their opposing viewpoint



Retweet

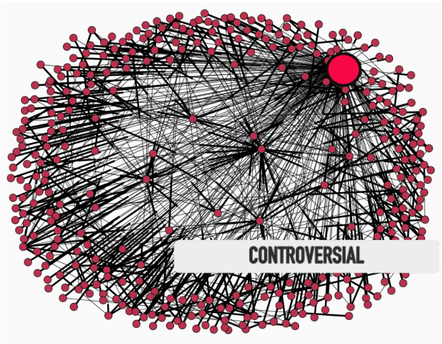


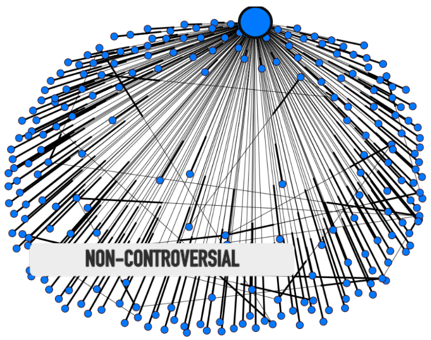
Reply

Motif-based approach

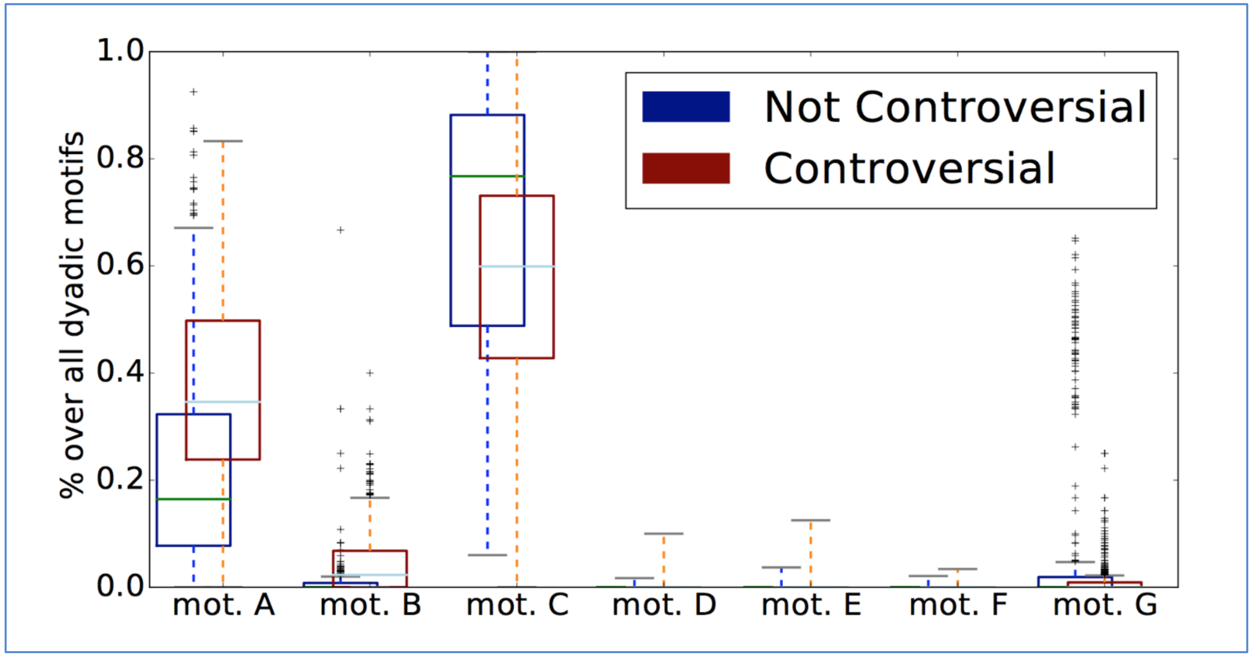
Coletto, Garimella, Luchesse, and Gionis. “A Motif-based Approach for Identifying Controversy.” OSNEM. 2017.

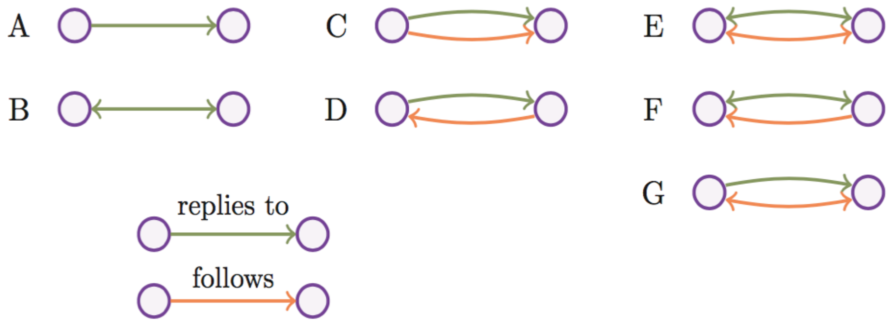
- Define reply trees
- Identify frequency of motifs in these trees
- Take into account also social graph
  - follower information





Motifs


1.0
0.8
0.6
0.4
0.2
0.0
% over all dyadic motifs
Not Controversial
Controversial
mot. A
mot. B
mot. C
mot. D
mot. E
mot. F
mot. G


A
B
C
D
E
F
G
replies to
follows

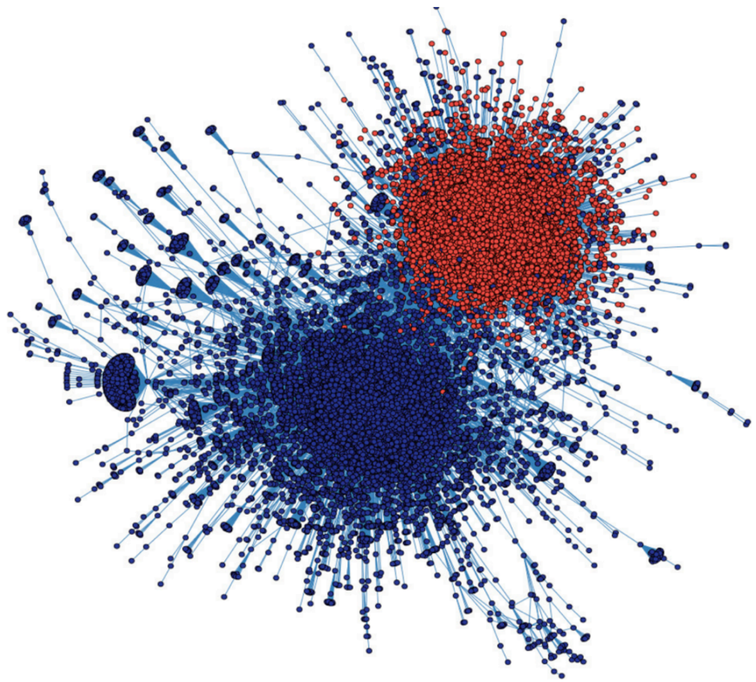
Quantifying polarization

- Identifying vs. Quantifying
- Defining what polarized/controversial is hard/subjective
- Quantifying might help to get a sense of the degree
- Basic idea:
  - Interactions have a clustered structure
  - Can we measure how well clustered the interactions are?

Quantifying polarization

Conover, Ratkiewicz, Francisco, Gonçalves, Menczer, and Flammini. "Political Polarization on Twitter." ICWSM 2011.

- Modularity:
  - the fraction of the edges that fall within the given groups minus the expected fraction if edges were distributed at random
  - Compares the number of edges inside a cluster with the expected on a random graph
  - Captures the strength of division of a network into modules

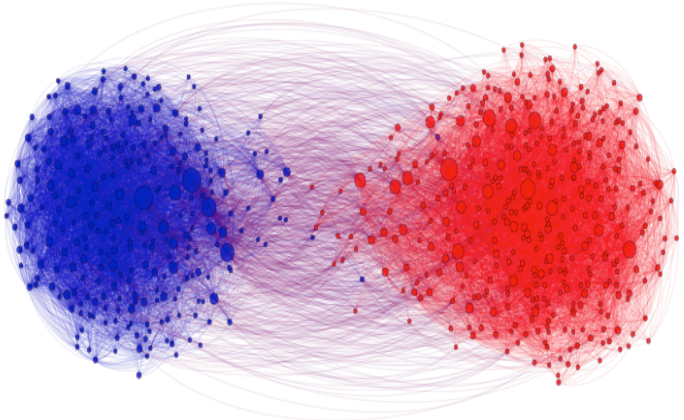


Modularity: 0.48

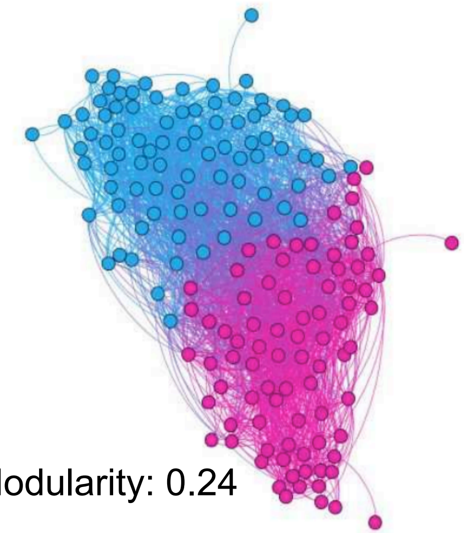
Modularity is not a direct measure of polarization

Guerra, Meira, Cardie, and Kleinberg. “A Measure of Polarization on Social Media Networks Based on Community Boundaries.” ICWSM 2013.

- We want to capture the in-group vs out-group interaction
- Sensitive to the size of the graph and partitions
- Not “monotone”
    - Strengthening of internal ties can decrease modularity
- How much modularity indicates polarization?



Modularity: 0.42

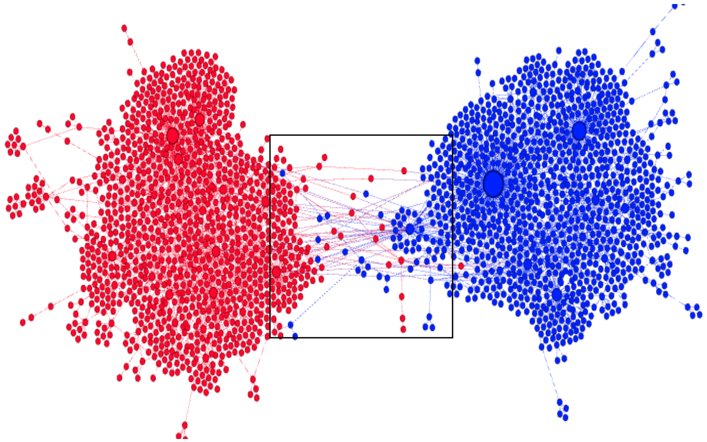


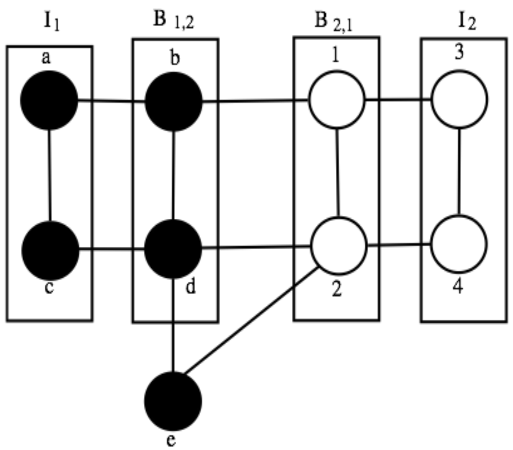
Modularity: 0.24

Community boundary

Guerra, Meira, Cardie, and Kleinberg. "A Measure of Polarization on Social Media Networks Based on Community Boundaries." ICWSM 2013.

- Boundary node:
  - have at least one edge that connecting to the other community
  - have at least one edge connecting to a member of its community which does not link to the other community
- $P(v) = d_{internal}(v)/(d_{external}(v) + d_{internal}(v)) - 0.5$
- $P(v) > 0 \rightarrow$ v prefers internal connections (antagonism?)
- $P(v) < 0 \rightarrow$ v prefers connections with members of the other group
- Polarization measure: average P(v) value over all boundary nodes

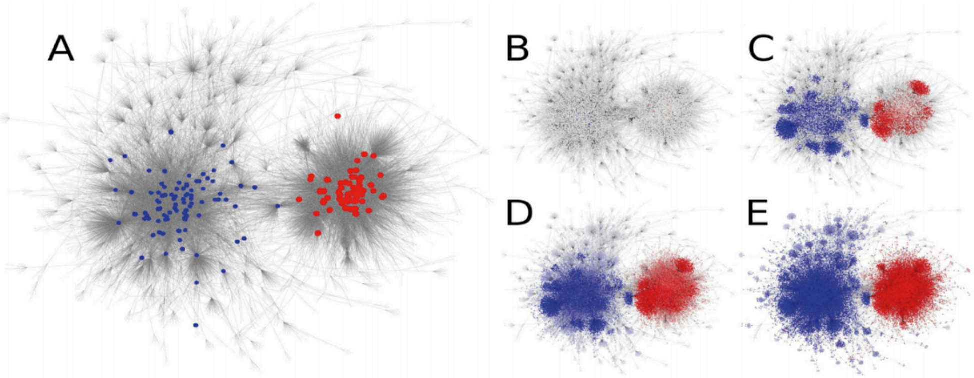




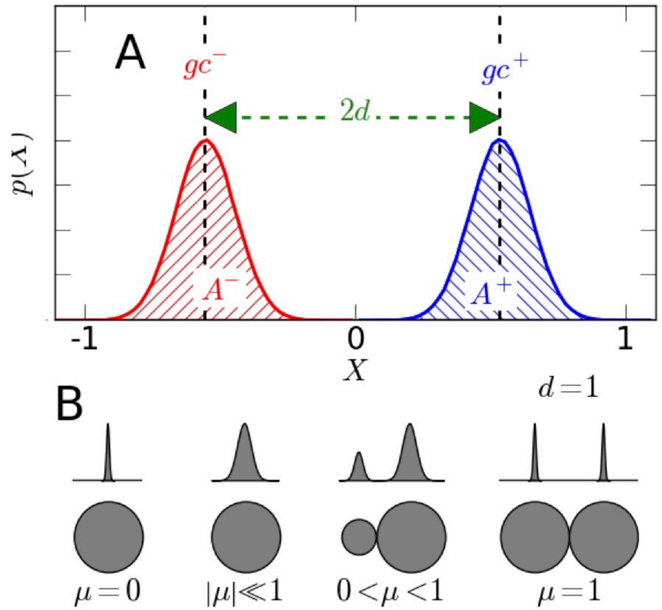
Label propagation

Morales, Borondo, Losada, and Benito. “Measuring political polarization: Twitter shows the two sides of Venezuela.” Chaos. 2015.

- Opinion formation:
  - Identify a set of ‘seed’ users and propagate until convergence



- Measure: distance between distributions
  - “Dipole moment”
  - Accounts for the mass of the population

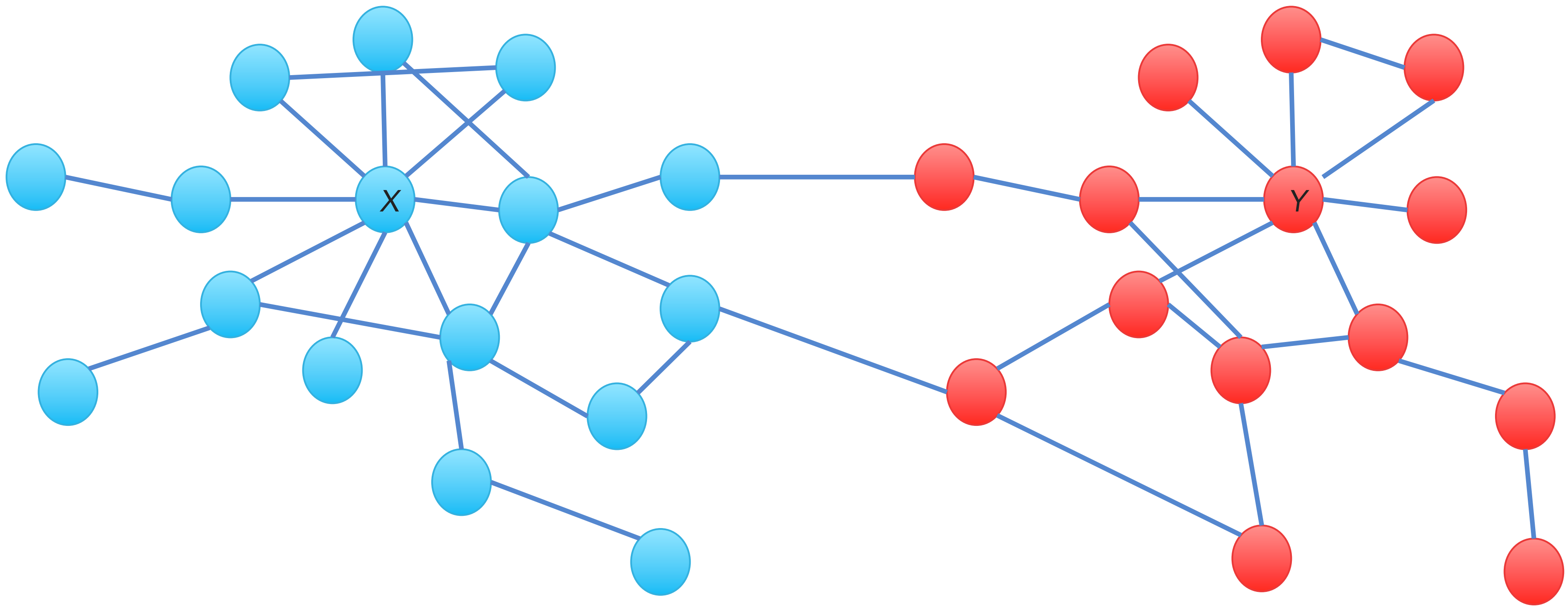


Based on information flow

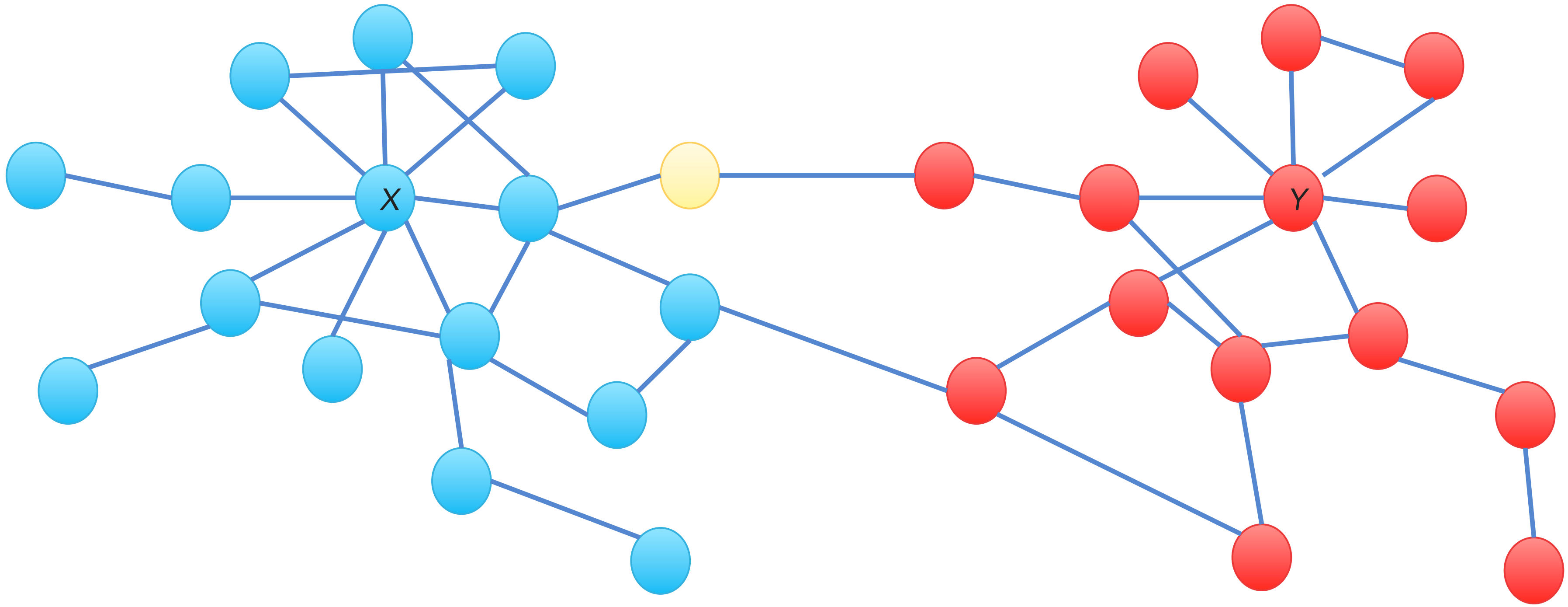
Garimella, De Francisci Morales, Gionis, and Mathioudakis. "Quantifying Controversy in Social Media." WSDM 2016.

- Random walk controversy measure (RWC)
  - Authoritative users exist on both sides of the controversy
  - How likely a random user on either side is to be exposed to authoritative content from the opposing side
- Works on both the retweet graph and the social graph
- Requires a partition of the graph

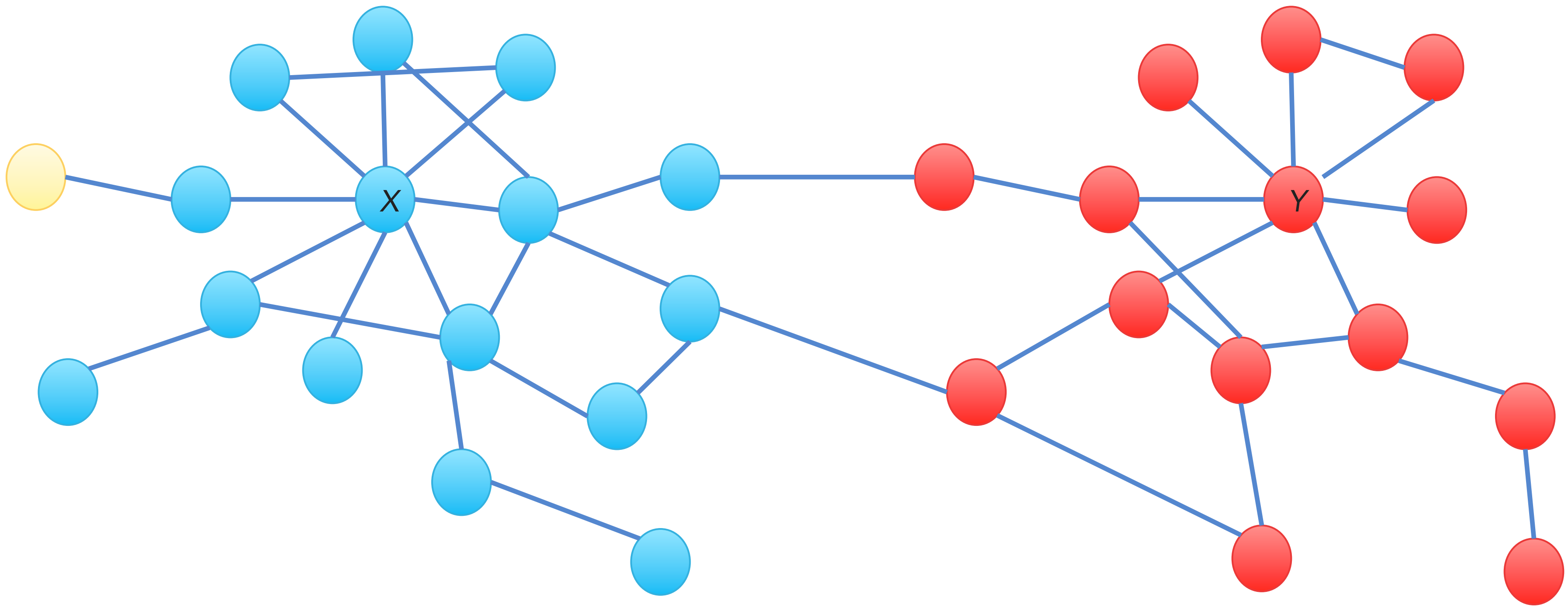
Random walk controversy score



Random walk controversy score



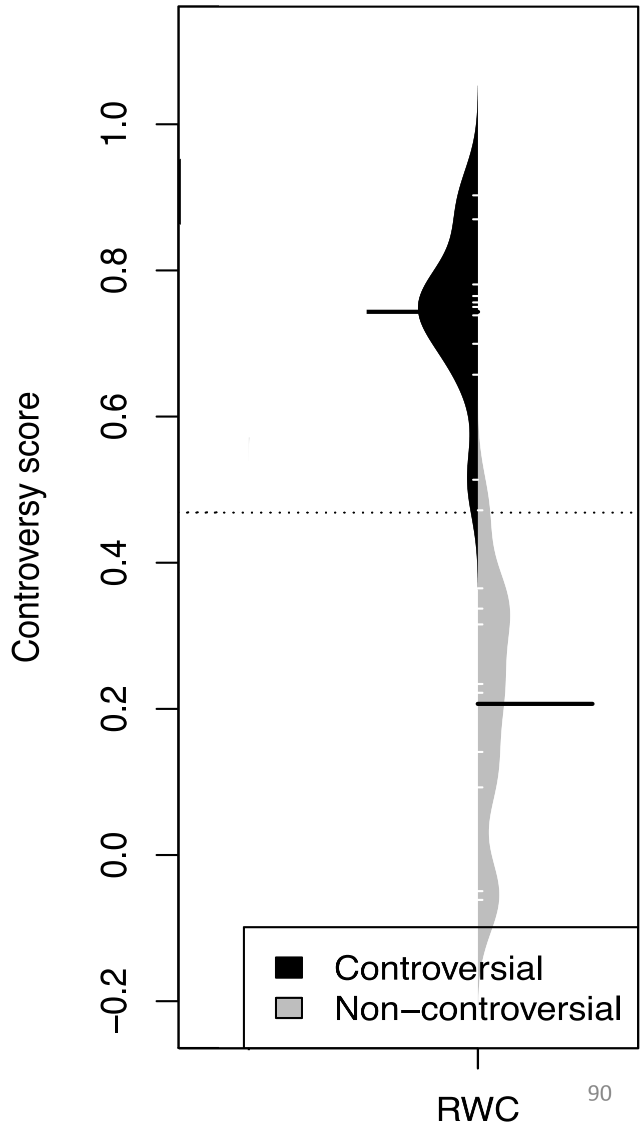
Random walk controversy score



Random walk controversy score (RWC)

- P_{AB} = probability that random walk started in cluster A, given that it ended in a hub of cluster B

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



User level polarization

- Can we find how a user will lean on a controversial topic?
- Mostly – “can we identify political affiliation of users on Twitter?”
  - Several papers look at social network of users
    - Following known political figures (e.g., Obama, Trump) or news outlets with known political leaning (Fox News, NYT)

Bayesian ideal point estimation

Barbera. "Birds of the Same Feather Tweet Together. Bayesian Ideal Point Estimation Using Twitter Data." Psychological science. 2013.

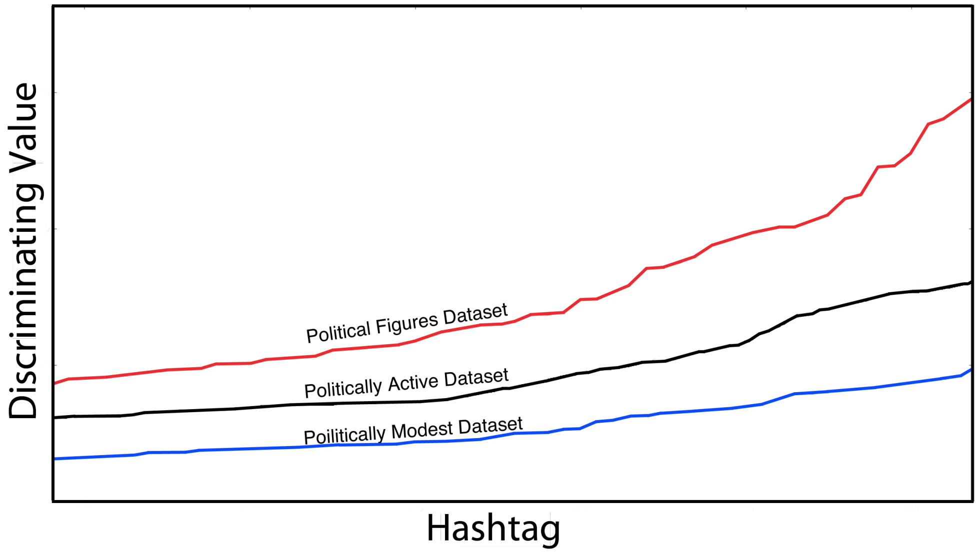
- Ideal point estimation (continuous) vs ideology/polarity (binary)
- Assumption: Twitter users prefer to follow politicians whose position on the latent ideological dimension are similar to theirs
    - Parameters to control for popularity of the politician and activity of the user

User level - content

Conover, Gonçalves, Ratkiewicz, Flammini, and Menczer. "Predicting the political alignment of Twitter users." PASSAT 2011.
Cohen and Ruths. "Classifying Political Orientation on Twitter: It's Not Easy!" ICWSM 2013.

- Binary classifier based on several features: text, hashtags, clusters
- However, Cohen and Ruths show that its not as simple and depends on who you measure the polarity for and what you train on
  - Use of loaded (political) hashtags to create dataset biases the result
  - Politically "modest" users are much harder to classify

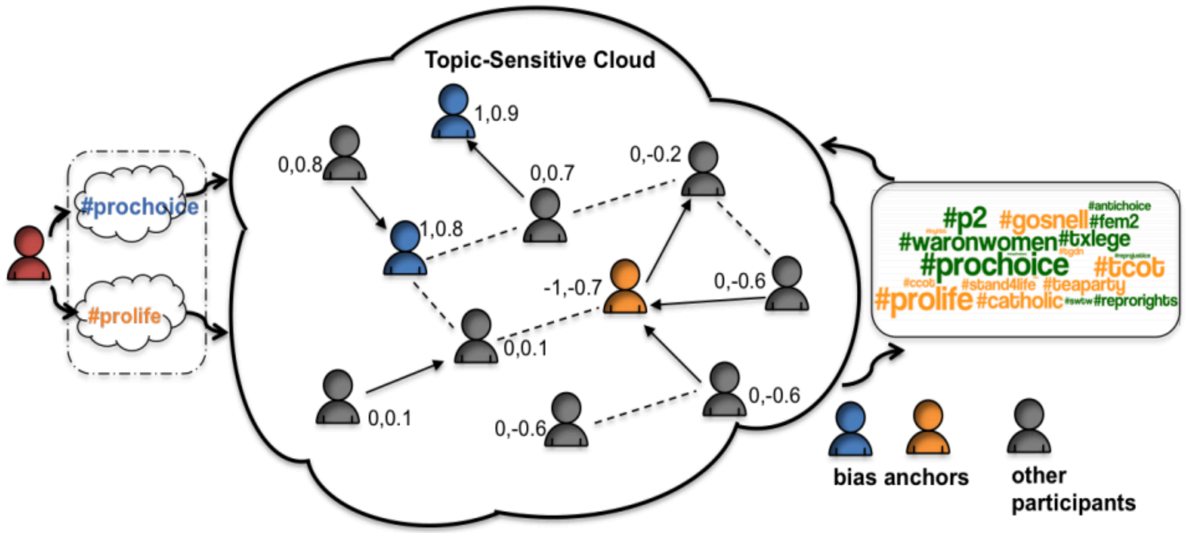
Features	Conf. matrix	Accuracy
Full-Text	$\begin{bmatrix} 266 & 107 \\ 75 & 431 \end{bmatrix}$	79.2%
Hashtags	$\begin{bmatrix} 331 & 42 \\ 41 & 465 \end{bmatrix}$	90.8%
Clusters	$\begin{bmatrix} 367 & 6 \\ 38 & 468 \end{bmatrix}$	94.9%
Clusters + Tags	$\begin{bmatrix} 366 & 7 \\ 38 & 468 \end{bmatrix}$	94.9%



Combine content and network

Lu, Caverlee, and Niu. “Biaswatch: A lightweight system for discovering and tracking topic-sensitive opinion bias in social media.” CIKM 2015.

- Not necessarily political affiliation, but bias towards a polarizing topic
- Method:
  - Start with hand-picked seed hashtags (e.g. #prochoice vs #prolife)
  - Find bias anchors (partisan users) and other biased hashtags
  - Construct a user similarity network based on content and retweets
  - Propagate bias on this user similarity network
  - Correct for noise



Random walk-based approach

Garimella, De Francisci Morales, Gionis, and Mathioudakis. "Quantifying Controversy on Social Media." Transactions on Social Computing. 2018.

- Language and topic independent (not necessarily politics)
- No content features needed
- 2 variants
  - Direct extension of RWC

$$RWC^{user}(u, X) = \frac{Pr[\text{start} = u \mid \text{end} = X^+]}{Pr[\text{start} = u \mid \text{end} = X^+] + Pr[\text{start} = u \mid \text{end} = Y^+]}.$$

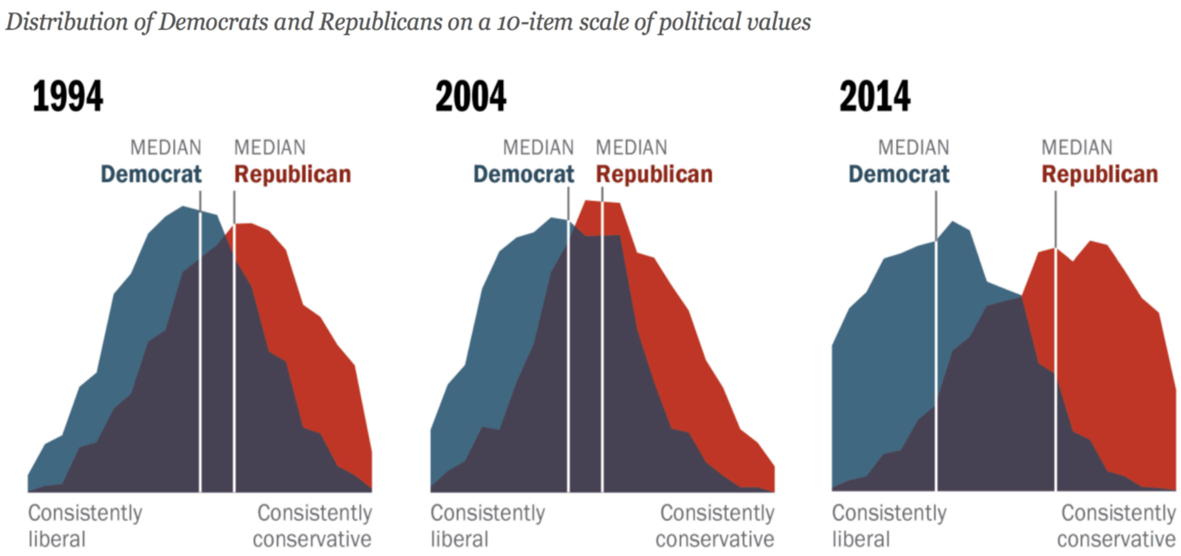
  - Expected hitting time
    - For a user u, find the expected number of steps in a random walk to hit an influential node from the X side (and Y side)
    - Rank all the users according to this measure, $\rho^X(u) \in [0,1]$ is the rank for X
    - The polarity of u is the difference in ranks

$$\rho(u) = \rho^X(u) - \rho^Y(u) \in (-1, 1).$$

Polarization over time

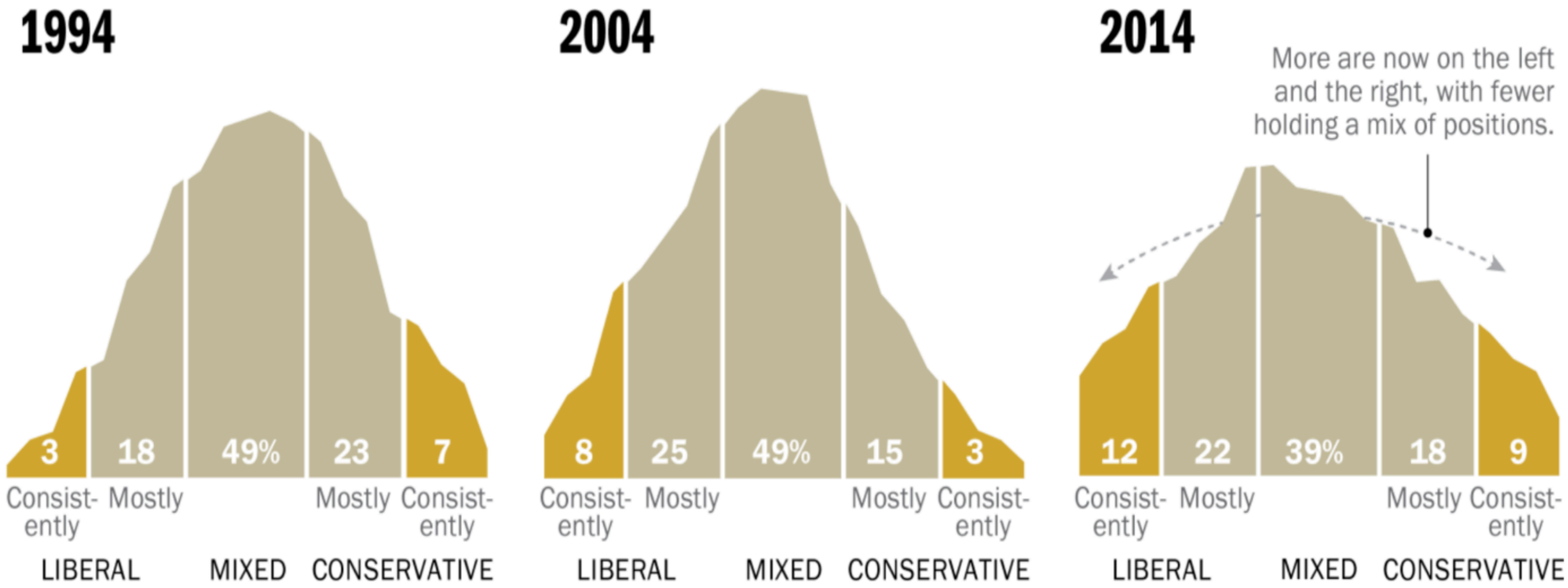
Political Polarization in the American Public

- ~10k adults nationwide
- 10 political values questions



Pew Research Center, “Political Polarization in the American Public” (2014)

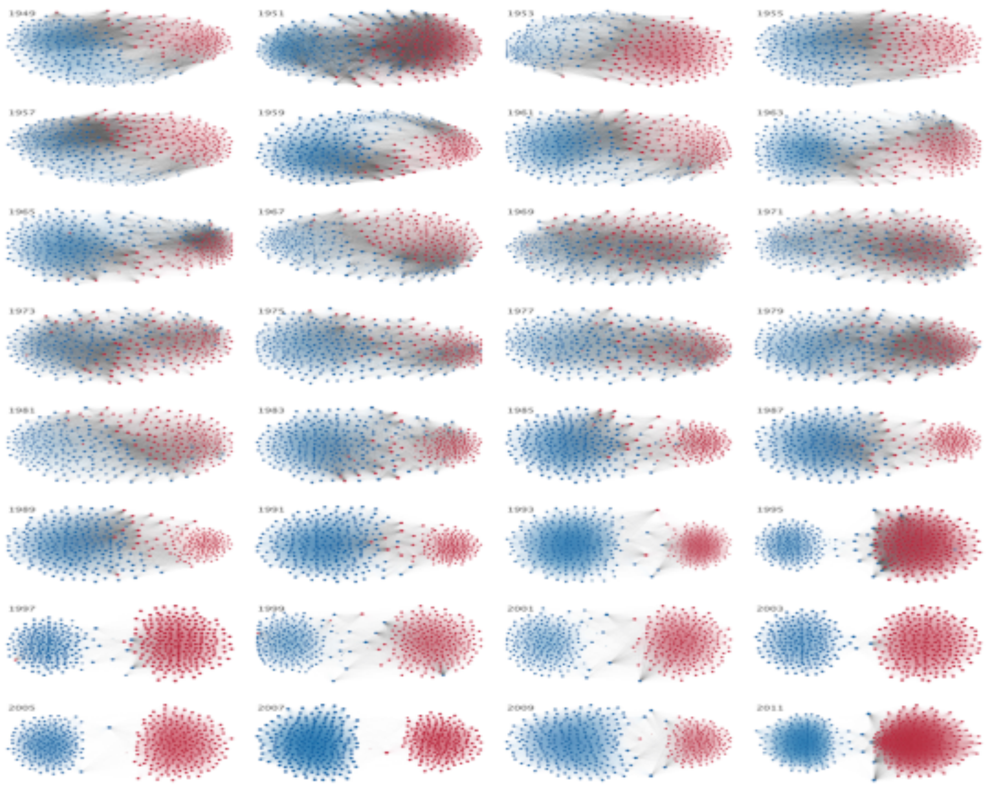
Political Polarization in the American Public



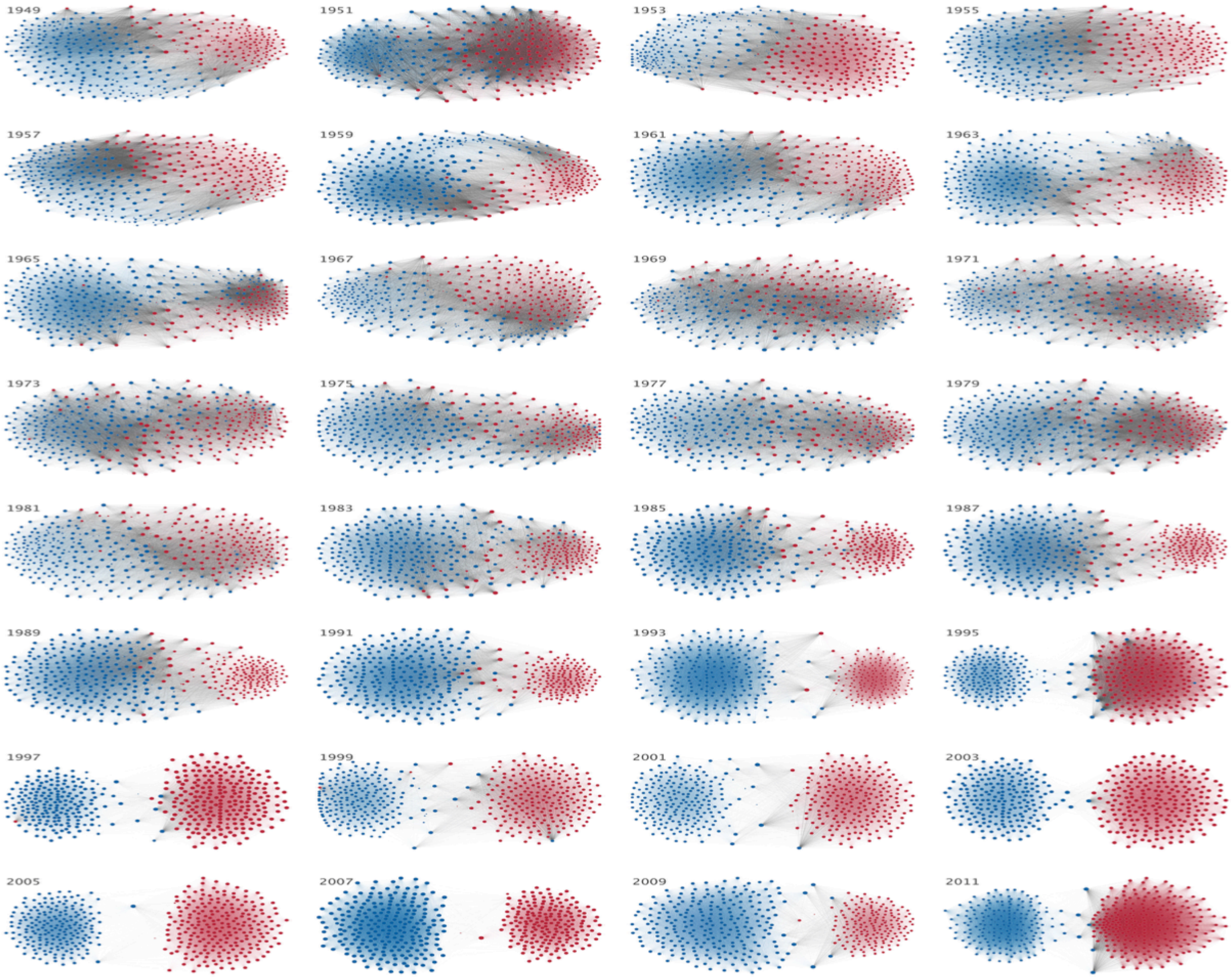
Political Polarization in the American Public

- People with less interest in politics are less involved
- People with higher interest are more involved and more polarized
  - these people vote and hence matter the most
- Polarized politics = polarized everything

Partisanship of US House of Representatives



Andris C., Lee D., Hamilton M., Martino M., Gunning C., Selden J.. "The Rise of Partisanship and Super-Cooperators in the U.S. House of Representatives." (2015)


1949
1951
1953
1955
1957
1959
1961
1963
1965
1967
1969
1971
1973
1975
1977
1979
1981
1983
1985
1987
1989
1991
1993
1995
1997
1999
2001
2003
2005
2007
2009
2011

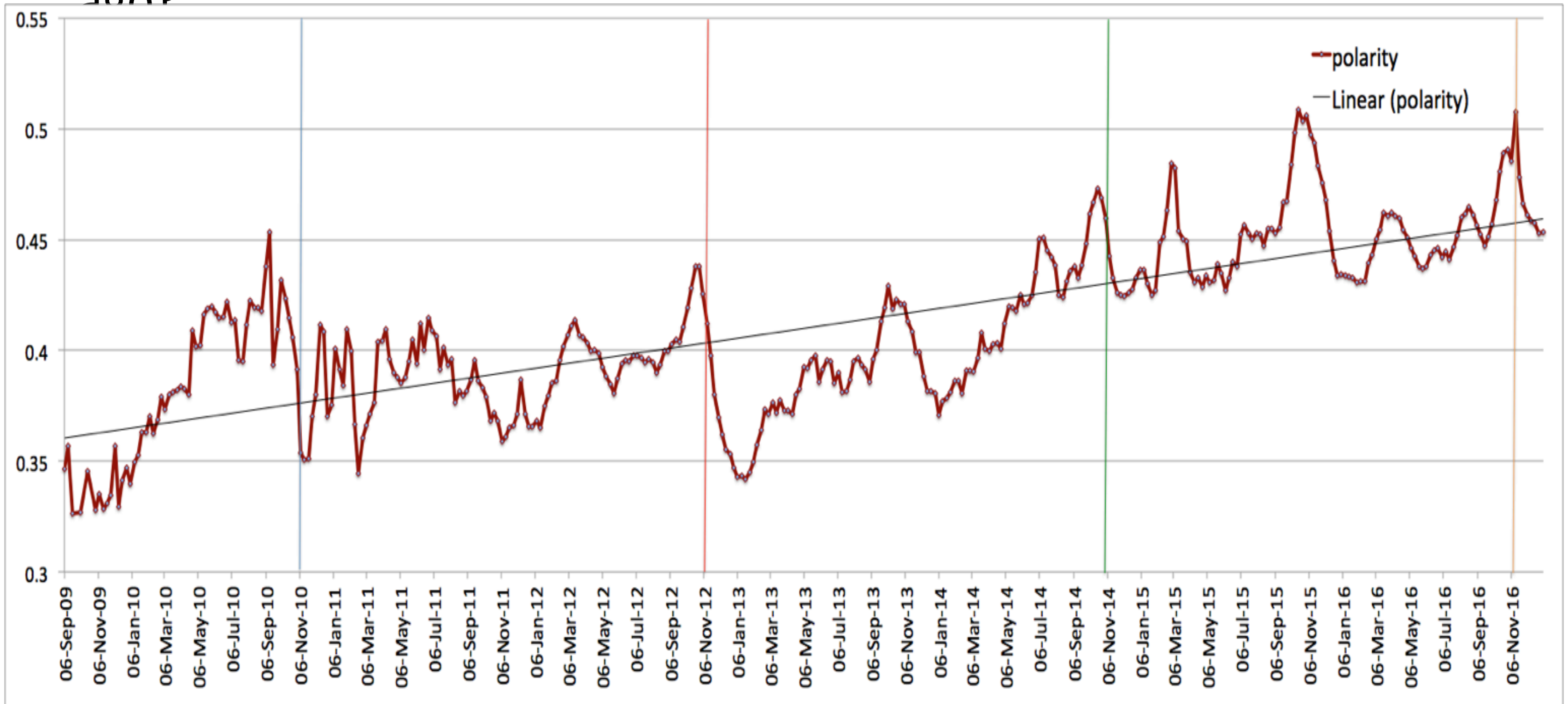
Long-term trends in polarization on Twitter

- Are twitter users more/less likely to follow/retweet political figures/media accounts from both sides now compared to 8 years ago?
- Are users more/less likely to use biased content? (hashtags)

Garimella, K., & Weber, I. "A long term Analysis of Polarisation on Twitter." (2017)

Long-term trends in polarization on Twitter

- Are twitter users more/less likely to follow/retweet political figures/media accounts from both sides now compared to 8 years ago?



Garimella, K., & Weber, I. "A long term Analysis of Polarisation on Twitter." (2017)

Summary

- What is polarization
- Methods to identify polarized topics
  - And quantify the degree of polarization
  - From content and network
- Methods to identify the polarity of users
  - From content and network

Part 5

Mitigating polarization

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
  - Motivation
  - Tools for bursting the filter bubble
  - Learning and visualizing the ideology space
  - Algorithmic mediation
  - Why is the problem hard?
- Part 6: Challenges and Directions for Future Research

In this part of the tutorial

- Social engineering for mitigating polarization
- The main idea:

  nudge people to stay in touch with the opposing-side view

- We discuss
  - Existing tools
  - Research approaches
  - User-interface issues

Motivation

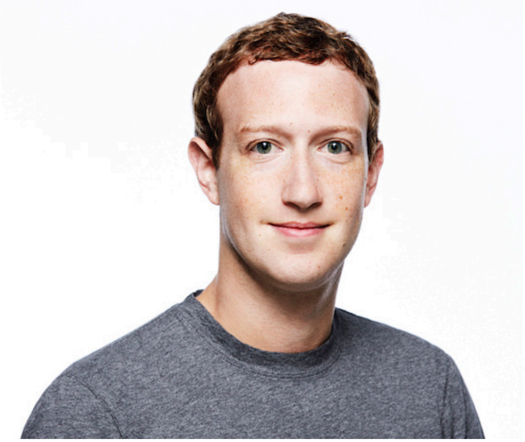
Acknowledging the problem of information silos

"The internet has exacerbated phenomenon of people having conversations in their own silos."
"If you're liberal, then you're on MSNBC. If you're a conservative, you're on Fox News."

Barack Obama, 24 April 2017

"Filter bubbles are a serious problem with news."

Bill Gates, 21 February 2017



"The two most discussed concerns this past year were about diversity of viewpoints we see (filter bubbles) and accuracy of information (fake news)."

Mark Zuckerberg, 16 February 2017

Bursting filter bubbles

- Online filter bubbles are seen as a problem
- Often people are not aware of the other side
- Other times they are aware but do not want to know or care
- Initiatives to counter the issue
- Different objectives
  - Inform users of the biases in their news/information diets
  - Allow users to see other viewpoints (outside their bubble)
  - Correct misinformation

How to mitigate the problem?
Offering an opposing view, e.g., fact checking

- Fact-checks influence people's assessments of negative political ads (Fridkin et al. 2015)
  - limitation: a single experiment

- longitudinal effects of exposure to fact-checking (Nyhan and Reifler 2015)
  - in general, the public has positive views of fact-checking
  - exposure to fact-checks helps people become better informed
  - attitude becomes more positive for well-informed, educated
  - viewed more favorably by Democrats than Republicans

Fridkin, Kenney, and Wintersieck. "*Liar, liar, pants on fire: How fact-checking influences citizens' reactions to negative advertising*." Political Communication 32.1 (2015)

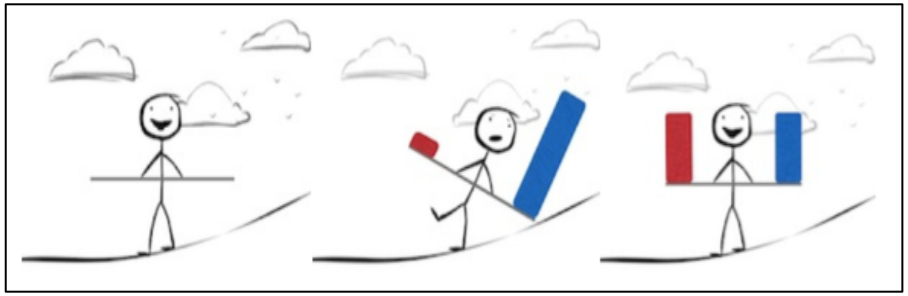
Nyhan, and Reifler. "*Estimating fact-checking's effects: Evidence from a long term experiment during campaign*", Unpublished manuscript (2015).

Social media can support political deliberation and depolarization

- (Semaan et al. 2014) conducted a small-scale study where people could use multiple platforms where they
  - were serendipitously exposed to diverse political information
  - constructed diverse information feeds
  - disseminated diverse information
  - engaged in reasoned political discussions with diverse audiences
  - purposefully sought diverse information and discussants
- Interviews confirm that online discussions lead to depolarization (people changing their views)

Semaan, Robertson, Douglas, Maruyama, "*Social media supporting political deliberation across multiple public spheres: Towards Depolarization*", CSCW 2014

Balancer



- Browser (chrome) extension that augment *Digg* and *Reddit*
- Monitors news articles visited by user
- Reports left-vs-right balance
- Leanings computed by curated lists of websites
- Does awareness improve balance?
  - Study finds small improvement

S. Munson, S. Lee, P. Resnick. *"Encouraging reading of diverse political viewpoints with a browser widget".* ICWSM 2013

In the following slides

- Tools to burst filter bubbles
  - Sandboxes in popular websites
- Learning and visualizing the ideology space
- Algorithmic mediation / recommendations
  - What to recommend?
    - Users-to-follow vs. content
  - How to find meaningful recommendations?
    - Utilize existing metrics (polarization, opinion, diffusion)
    - Make recommendations to alleviate polarization problem, according to such metrics
- Why is the problem hard?

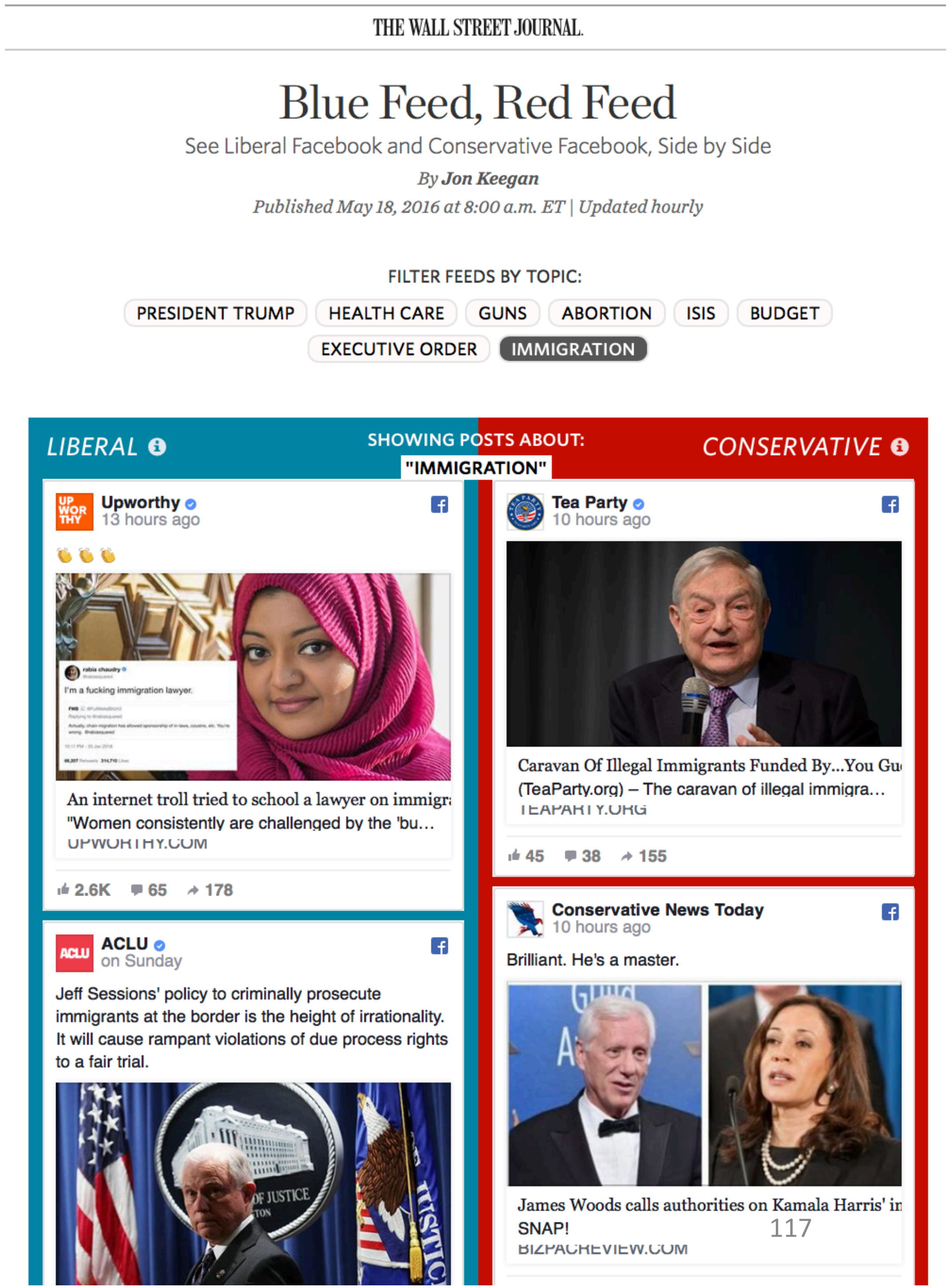
Tools for bursting the filter bubble

Wall Street Journal

Blue Feed, Red Feed

Curated by the newspaper

Aims to show how different the facebook feed can be for different users



Support The Guardian

Subscribe Find a job Sign in Search

International edition

The Guardian

News Opinion Sport Culture Lifestyle More

World ▶ Europe US Americas Asia Australia Middle East Africa Inequality Cities Global development

Burst your bubble

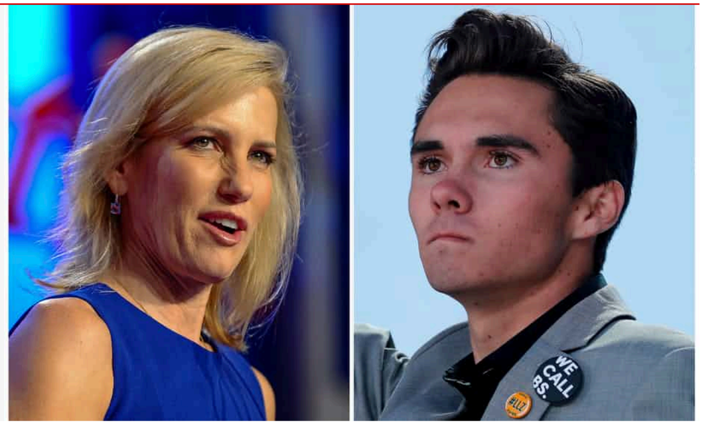
The Guardian's weekly guide to conservative articles worth reading to expand your thinking

2 April 2018

Laura Ingraham is a victim of a totalitarian campaign from the left, apparently

The American right have revealed a vision of free speech that is very expansive for conservatives but far less accommodating for those who disagree with them

7:41 PM 694



Burst your vubble by the guardian

The Guardian is left-wing

The column shows selected conservative articles from around the web

Escape your bubble

Browser (chrome) extension

Asks you which type of people you would like to be more accepting to

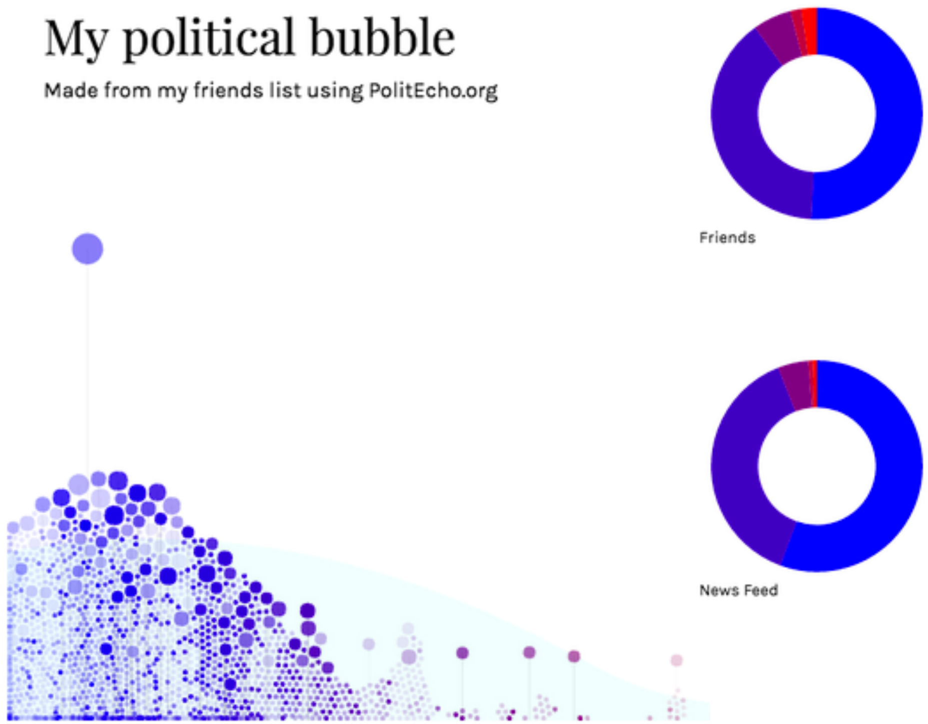
App inserts human-curated, positive articles and images into Facebook News Feed, which paint those you would like to be more accepting of in a positive light

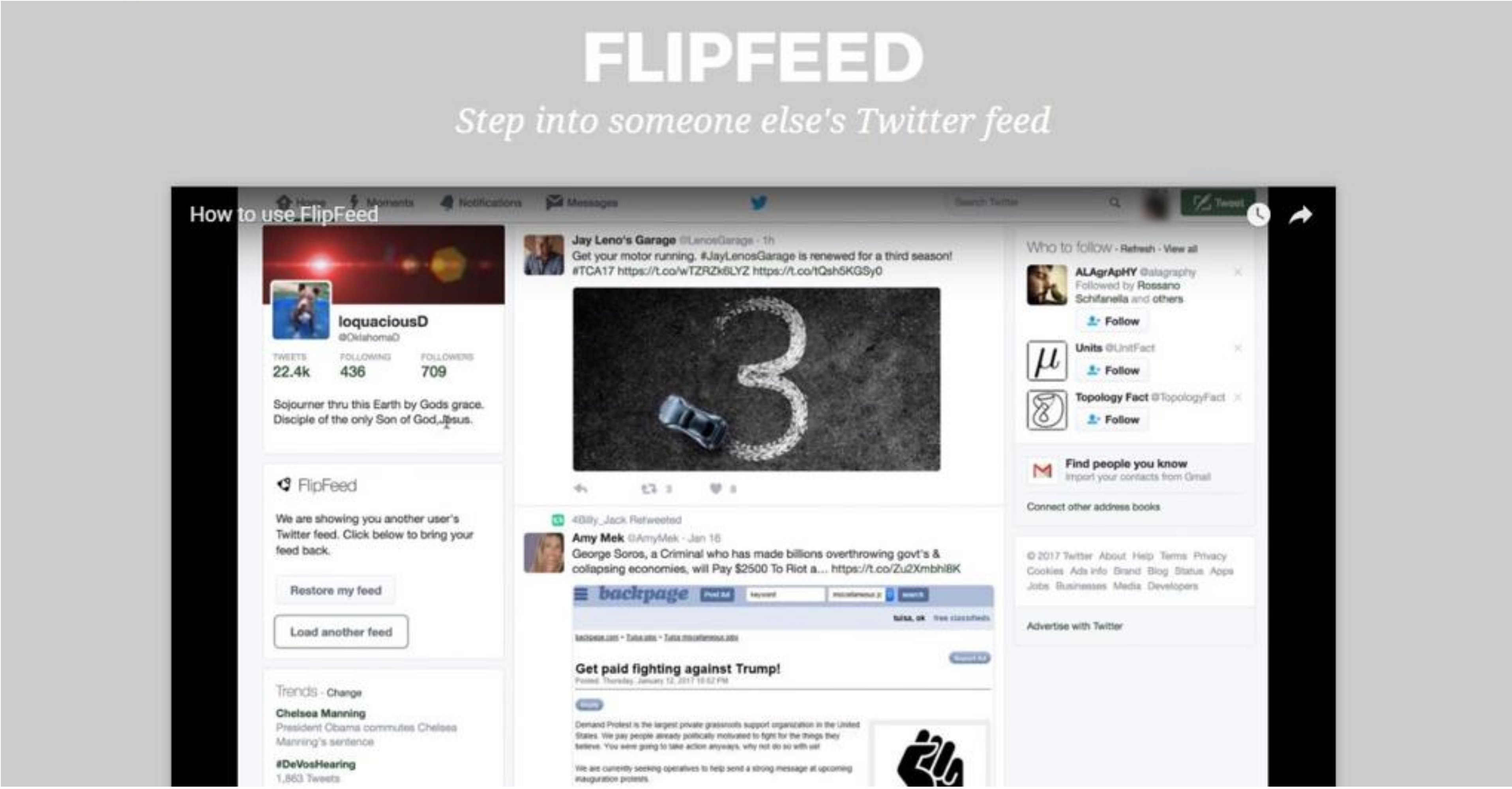
politecho.org

Browser (chrome) extension

Shows political distribution of own Facebook feed vs. that of friends

Compares liked political pages with a reference set of political pages

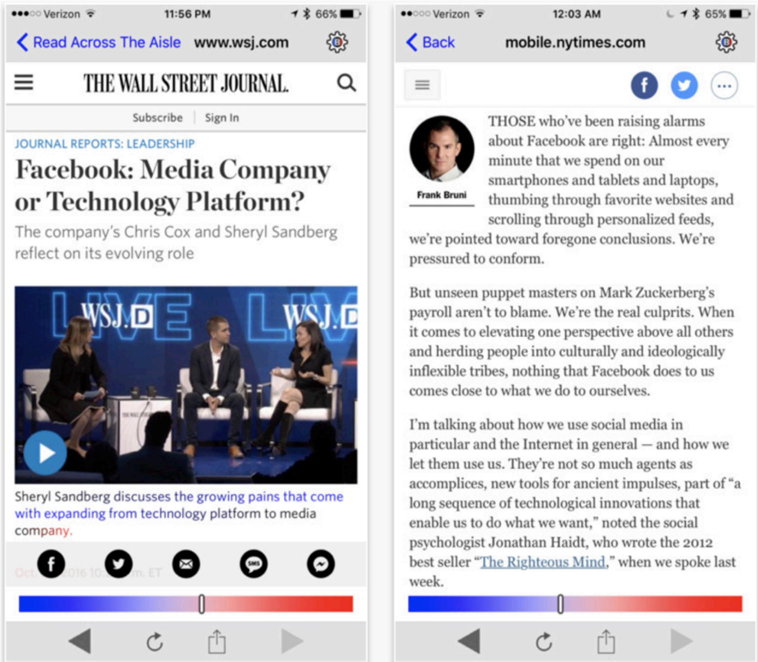




Browser (chrome) extension

Allows Twitter users to see a feed that resembles that of another user who has been pre-classified as right- or left-leaning

Laboratory for Social Machines at MIT Media Lab



Mobile (iPhone) app and chrome extension

News reader for select sources

Keeps track of personal reading history

Informs user of news diet bias

Learning and visualizing the ideology space

Improve awareness

- Develop data-driven methods that allow users to perceive their news diet
- Visualize/navigate in the underlying ideology space,

  their position, the accounts they follow, the news they read

- Task : learn latent ideological space of users and content
- Joint non-negative matrix factorization
  - User-user follow graph
  - User-content share graph

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

Case study

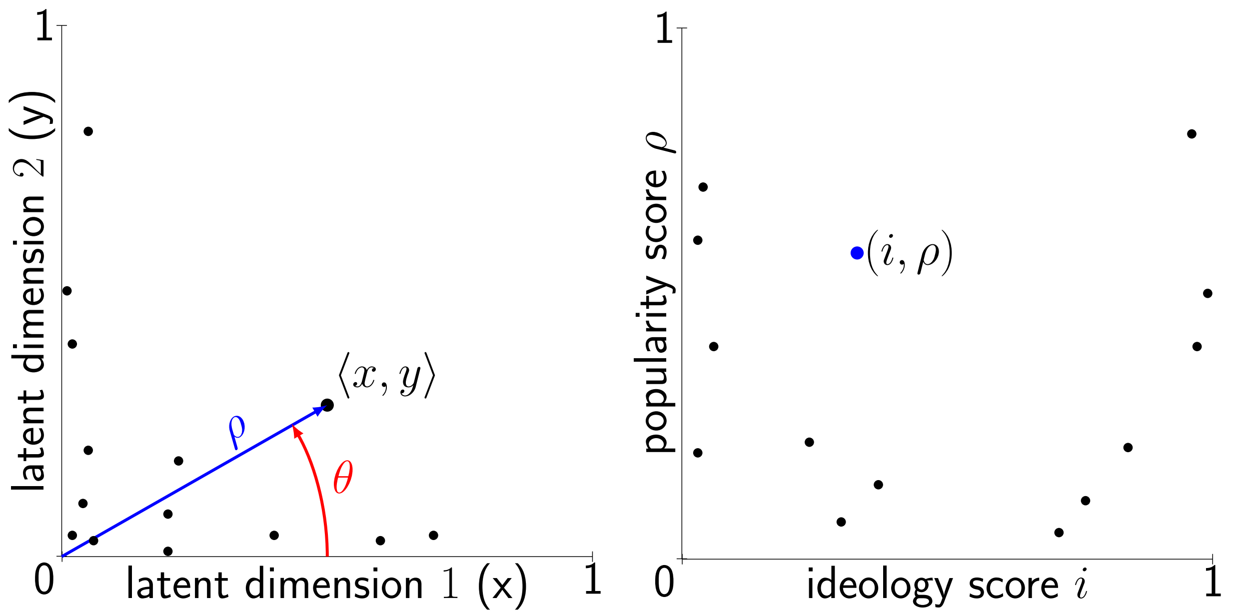
- Twitter data from 2011 to 2016, focusing on controversial topics
  (gun control, abortion, obamacare)
- 6391 users and 19 million tweets
- gather ground-truth polarity scores
  - content polarity (Bakshy et al., 2015)
  - user polarity (Barberá et al., 2015)

Barberá, Jost, Nagler, Tucker, and R. Bonneau. *"Tweeting from left to right: Is online political communication more than an echo chamber?"* Psychological science, 2015

Bakshy, Messing, and Adamic. "*Exposure to ideologically diverse news and opinion on facebook*". Science, 2015

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

Learned ideology latent space

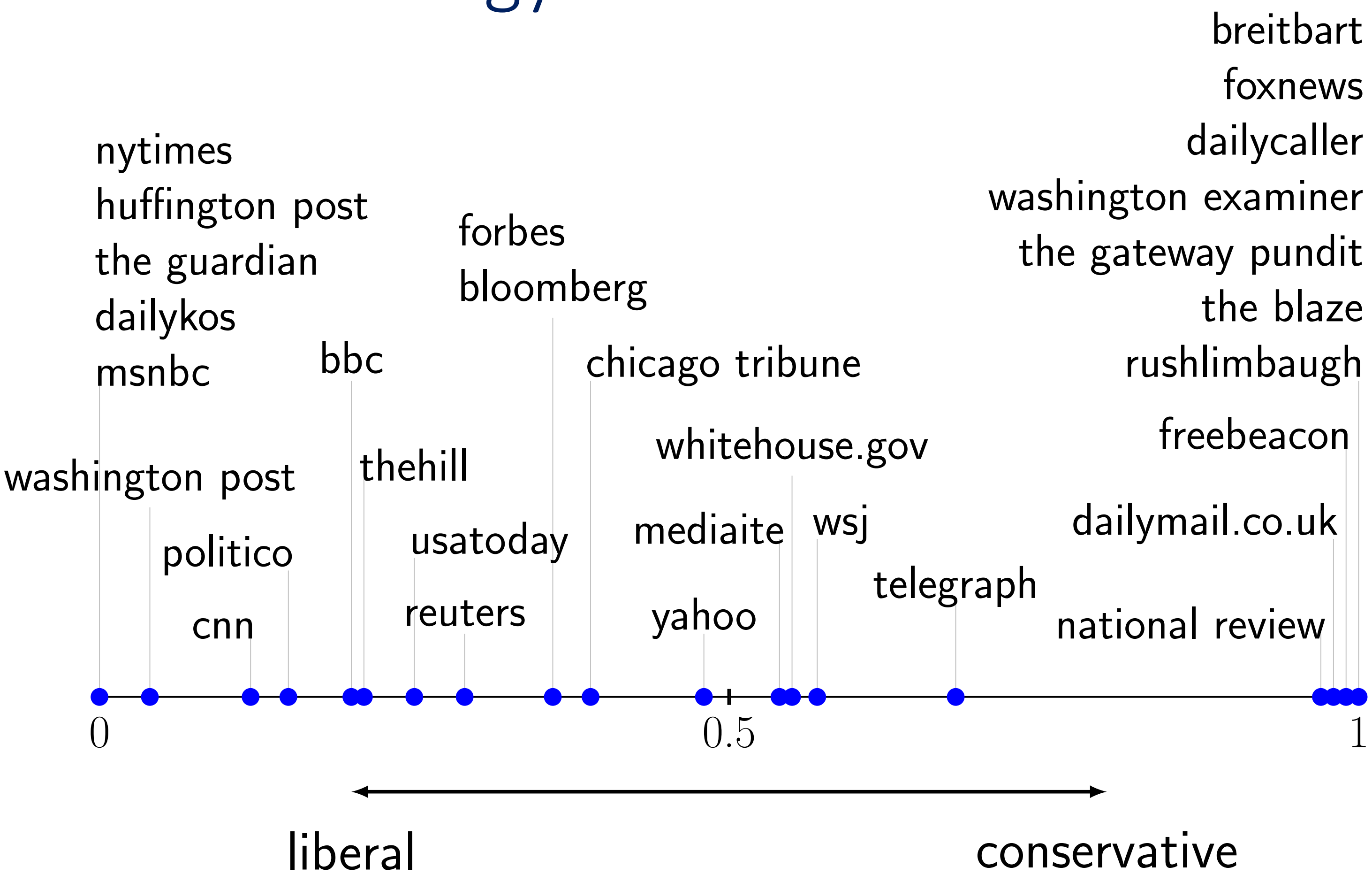


(a) original latent space **(b) transformed coordinate space**

Figure 1: Projection of a subset data points in the learned ideology latent space and the transformed ideology-popularity coordinate space.

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

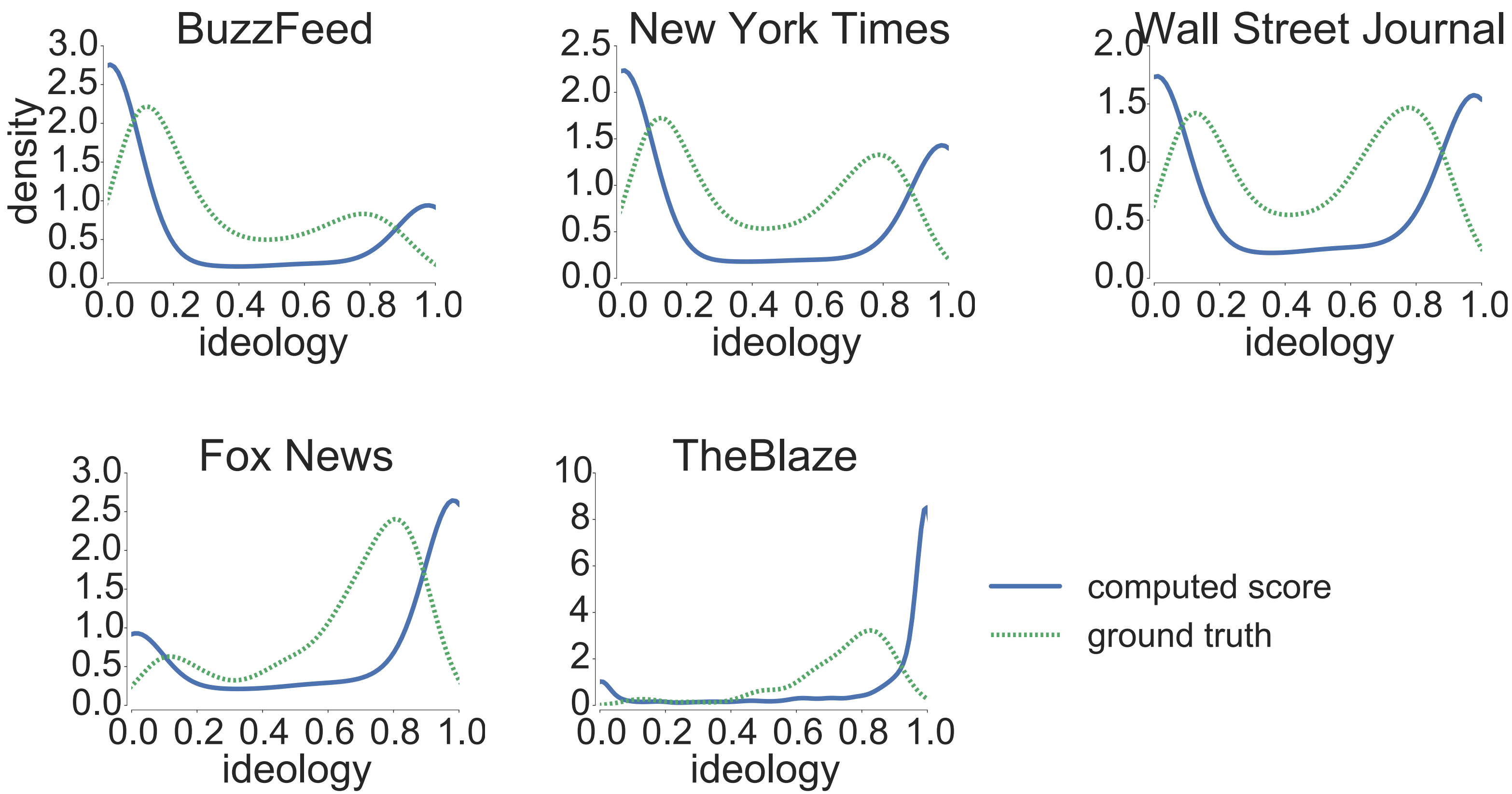
Content ideology scores



correlation with ground-truth scores 0.82

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

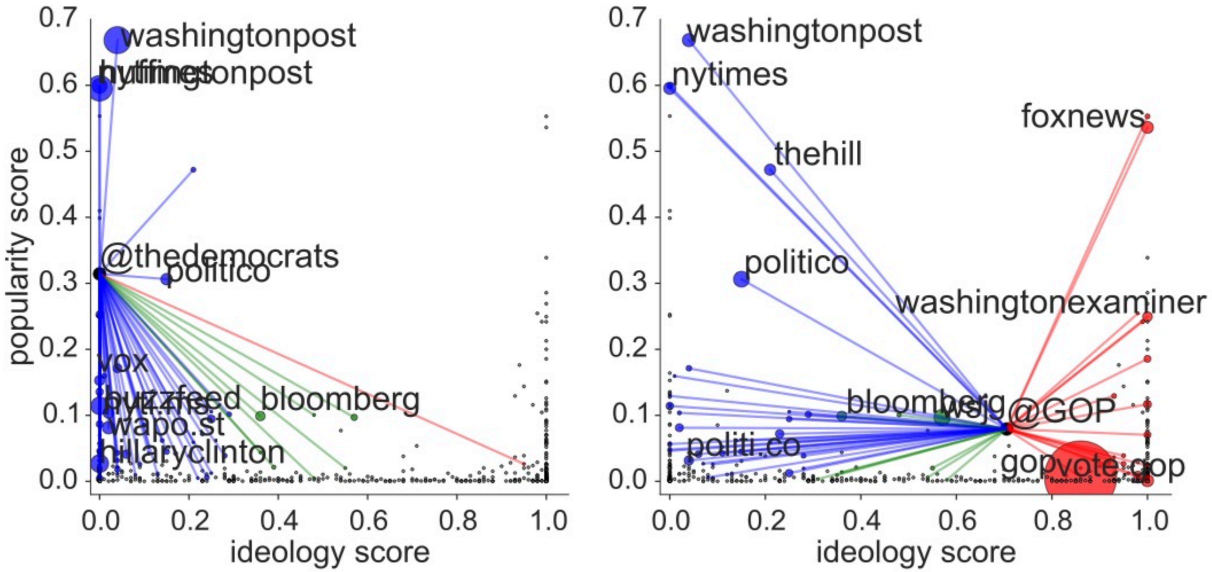
Audience ideology scores



correlation of user ideology scores with ground-truth 0.90

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

Visualizing the information bubble



(a) Democratic Party **(b) Republican Party**

Figure 4: Ideological position of @thedemocrats and @gop (black dots) and their content engagement. Points in the grey are the sources that the user never interacted with.

P. Lahoti, et al. *"Joint non-negative matrix factorization for learning ideological leaning on twitter"*. WSDM 2018

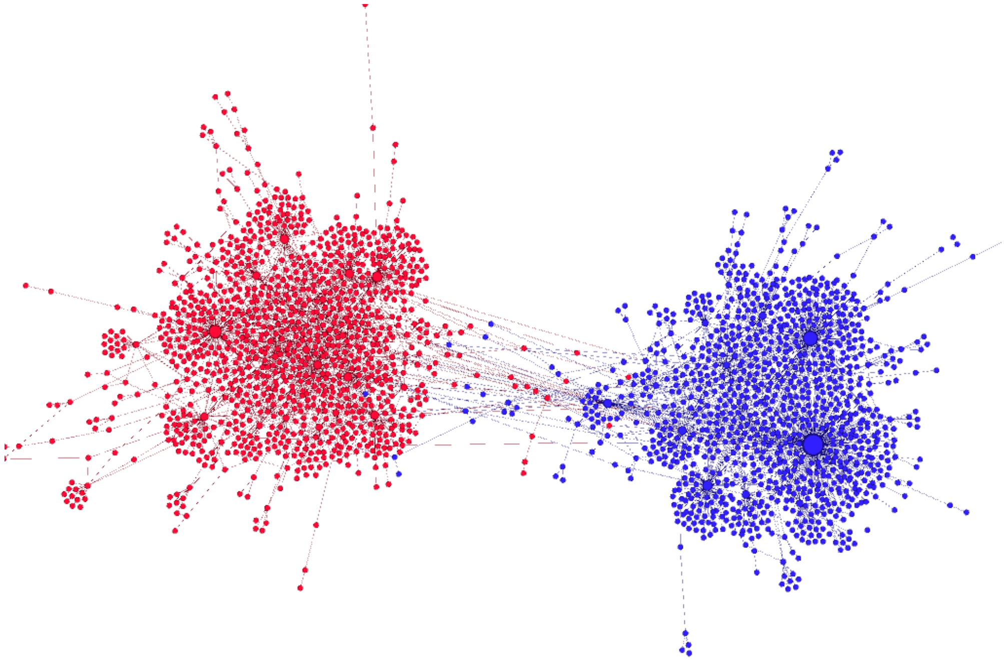
Algorithmic mediation

Algorithmic mediation / recommendations

- Task : make a recommendation helping to reduce polarization
- Different approaches driven by polarization metrics
  - Pick a favorite metric : RWC, opinion diversity, influence-based
  - Compute recommendation that reduces polarization according to the selected metric
  - Account for recommendation acceptance probability
- Another dimension: What to recommend? User vs. content

1. Recommendations based on RWC

- *Recall* : random-walk controversy score
- Quantifies the degree of polarization of a given topic
- Based on the structure of the retweet graph of the topic



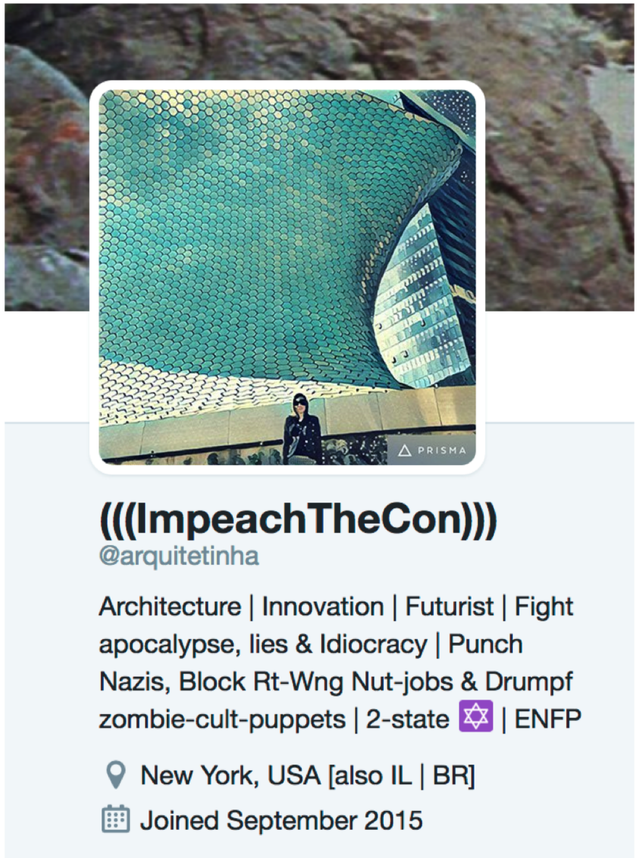
1. Recommendations based on RWC

- Assuming : polarization is measured by RWC
- Problem : add k edges to maximally reduce RWC
- Enhance greedy with efficient incremental computation
- Edge additions are interpreted as recommendations
- Incorporate probability of accepting a recommendation
  - compute user polarity, and
  - acceptance probability as a function of user polarity

K. Garimella et al. *“Reducing controversy by connecting opposing views”*. WSDM 2017

Reducing polarization : real example

Polarity = −0.99



Polarity = 0.95

Reducing polarization : real example

Polarity = −0.99

Polarity = 0.15

Reducing polarization : results

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

K. Garimella et al. *"Reducing controversy by connecting opposing views"*. WSDM 2017

2. Reducing polarization and disagreement based on an opinion-formation model

- Assume Friedkin-Johnsen opinion formation model
  - Agent i has innate opinion s_i and expressed opinion z_i
  - z_i is determined by the opinion-formation model
- Polarization index $P = \sum_{u \in V} z_u^2$
- Problem : set the opinion of at most k agents to 0 so as to minimize the polarization index
- Problem shown to be NP-hard
- Suggested a greedy method and compared against baselines

A. Matakos et al. *"Measuring and moderating opinion polarization in social networks"*. DMKD 2017

3. Reducing polarization and disagreement based on an opinion-formation model

- Assume Friedkin-Johnsen opinion formation model
  - Agent i has innate opinion s_i and expressed opinion z_i
  - z_i is determined by the opinion-formation model

- Disagreement index $$D = \sum_{(u,v) \in E} w_{uv}(z_u - z_v)^2$$

- Polarization index $$P = \sum_{u \in V} z_u^2$$

- Polarization and disagreement index $$I = P + D$$

Musco, Musco, and Tsourakakis, "*Minimizing Polarization and Disagreement in Social Networks*". WWW 2018

3. Reducing polarization and disagreement based on an opinion-formation model

- Problem 1 : given agents with their innate opinions, what is the optimal graph topology that minimizes polarization and disagreement?
  - (optimization over the space of all possible graphs)
- Problem 2 : given a network G, of agents with their own innate opinions, how should we change the initial opinions, for a maximum total change in opinion mass, so as to minimize polarization and disagreement?
- Both problems are convex – solvable in polynomial time

Musco, Musco, and Tsourakakis, "*Minimizing Polarization and Disagreement in Social Networks*". WWW 2018

4. Recommendations based on information propagation models

- *Recall* the classic viral-marketing setting
  - Given a social network and a propagation model
    e.g., 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 gives (1−1/e) approximation

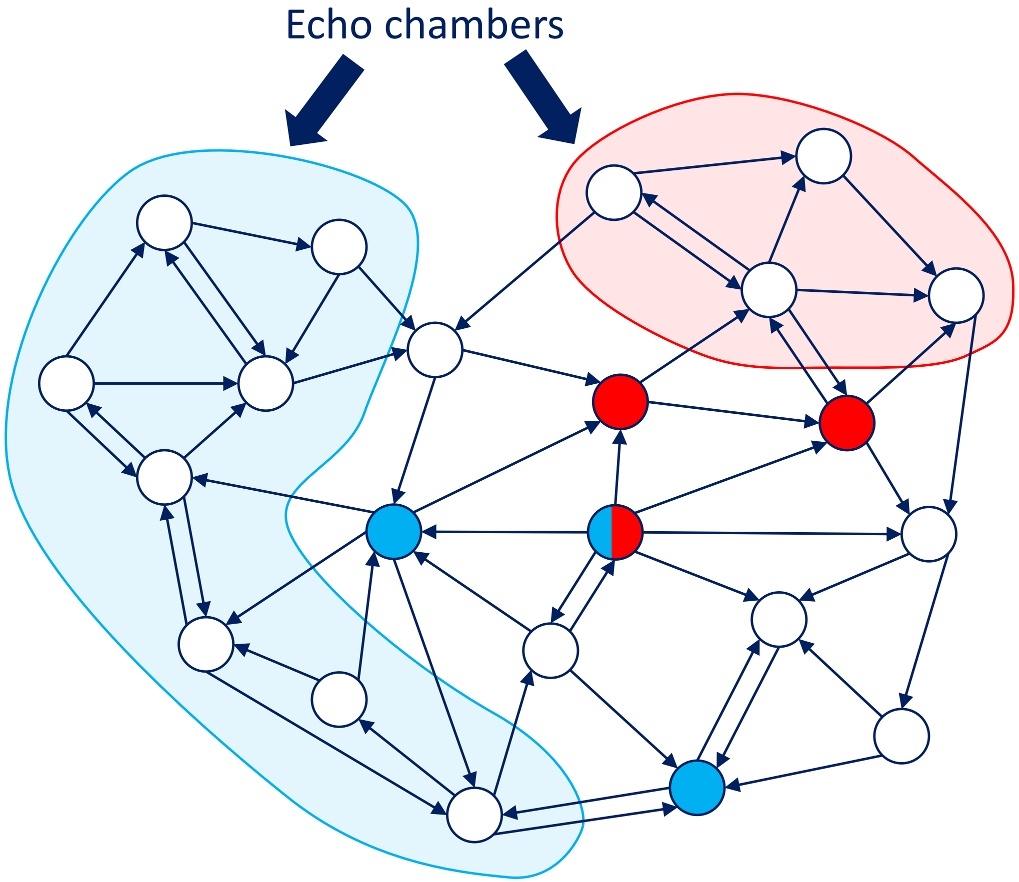
Balancing information exposure

- 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

K. Garimella et al. *"Balancing information exposure in social networks"*. NIPS 2017

Illustration

Online discussion on fracking



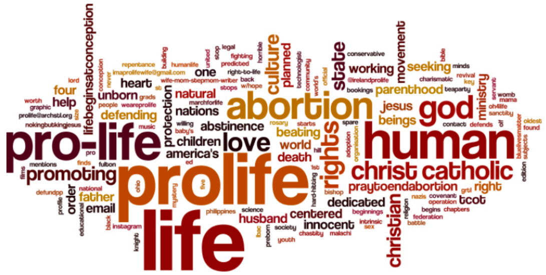
Balancing information exposure : results

- Optimization problem is **NP**-hard
- Minimization problem is **NP**-hard to approximate
- Maximization problem
  - objective function non monotone and non submodular
- Two models of how the campaigns propagate
- Approximation guarantee ½ (1−1/e)

K. Garimella et al. *"Balancing information exposure in social networks"*. NIPS 2017

Balancing information exposure : example

Side 1	Side 2	Hedge
Pro-Choice	*Pro-Life*	



Pro-Remain

Pro-Leave

K. Garimella et al. *"Balancing information exposure in social networks"*. NIPS 2017

Maximizing diversity

- An alternative approach is to make recommendations so as to maximize diversity in the social network
- What is diversity and how to measure it?
- At a user level : recommend diverse content
- At a network level : make recommendations so that friends see different content
  - Motivation : friends can discuss/debate
- Combinations of user and network diversity
- Another consideration : model propagation effects, or not

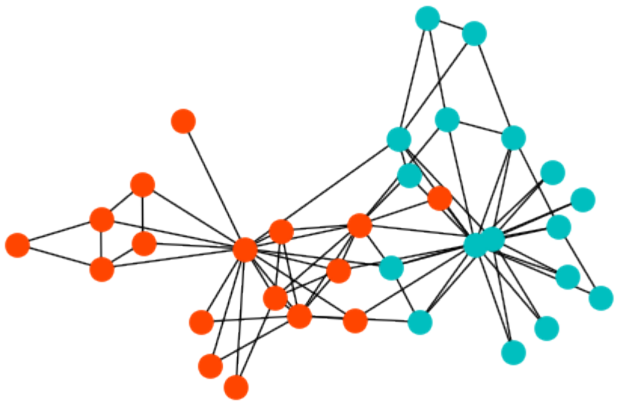
5. Maximizing diversity : *"tell me something my friends do not know"*

- Motivation : make recommendations to users in a social network so that friends see different content
- Make a small number of recommendations (k)
  - Why ? Intervene as little as possible
- A simple formulation that captures the essence of this problem
  - Social network graph G=(V,E)
  - Graph nodes have values +1 or -1
    - Corresponding to what kind of content they see
  - Select k nodes to swap their values so as to maximize the number of edges having different values at their endpoints, i.e., edges having values (+1,-1)

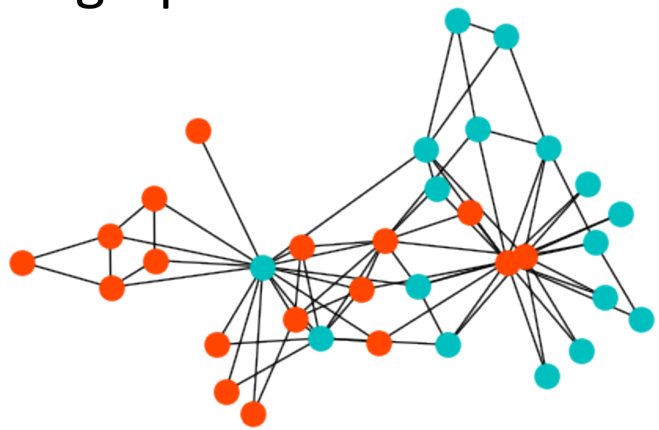
A. Matakos et al. "*Tell me something my friends do not know: Diversity maximization in social networks*". ICDM 2018

5. Maximizing diversity : *“tell me something my friends do not know”*

- Toy example in the “karate club” graph



(a) Echo-chamber graph



(b) Graph with diversified expo- sure

- Optimal solution for k=4

A. Matakos et al. “*Tell me something my friends do not know: Diversity maximization in social networks*”. ICDM 2018

5. Maximizing diversity : *"tell me something my friends do not know"*

- Problem complexity, formulation, and proposed solutions
- Problem is **NP**-hard (generalization of max-cut)
  - also **NP**-hard to approximate
- Problem formulation : non convex 0-1 quadratic problem
  - an instance of quadratic knapsack (QK)
- Proposed solutions
  1. SDP relaxation, and rounding inspired by QK solutions
  2. Glover's linearization, solve LP, round
  3. Greedy
  4. Obtain exact solution by mixed-integer quadratic programming (not scalable)

A. Matakos et al. "*Tell me something my friends do not know: Diversity maximization in social networks*". ICDM 2018

5. Maximizing diversity :
"tell me something my friends do not know"

TABLE II: Solution quality and bounds from the relaxations

Dataset	k	IQP	SDP-Relax	Glover	I-Greedy	S-Greedy
Karate	$0.1\,n$	46	46 (46.43)	46 (52.28)	46	46
	$0.2\,n$	56	56 (59.13)	54 (69.05)	56	51
	n	61	61 (63.48)	52 (78.00)	57	51
Karate-D	$0.1\,n$	50	50 (53.69)	49 (65.85)	50	50
	$0.2\,n$	55	55 (60.74)	50 (79.84)	53	52
	n	61	61 (63.89)	50 (93.00)	55	48
Books	$0.1\,n$	207	207 (207.81)	207 (235.90)	207	207
	$0.2\,n$	264	262 (272.26)	249 (330.01)	264	248
	n	309	306 (318.43)	267 (447.00)	298	253
Books-D	$0.1\,n$	265	262 (272.95)	249 (328.32)	263	252
	$0.2\,n$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	n	309	307 (318.44)	282 (497.50)	286	243
Twitter100	$0.1\,n$	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
	$0.2\,n$	599	599 (601.38)	589 (791.55)	599	592
	n	–	793 (804.21)	733 (1 406.00)	790	647
Twitter100-D	$0.1\,n$	743	742 (752.94)	722 (917.99)	742	715
	$0.2\,n$	–	757 (775.53)	729 (1 071.32)	761	715
	n	–	793 (804.21)	775 (1 496.00)	766	737
Blogs	$0.1\,n$	–	–	9 878 (12 659.06)	9 879	8 889
Elections	$0.01n$	–	–	–	117 950	117 482
Twitter	$0.001n$	–	–	–	–	1 678 753

A. Matakos et al. "*Tell me something my friends do not know: Diversity maximization in social networks*". ICDM 2018

5. Maximizing diversity : *"tell me something my friends do not know"*

- Future directions
- Consider more realistic problem formulations
    - Continuous user leaning score
    - Continuous recommendation leaning score
    - Probability of accepting a recommendation

A. Matakos et al. "*Tell me something my friends do not know: Diversity maximization in social networks*". ICDM 2018

6. Maximizing diversity of exposure

- Motivation : recommend content to users of a social network
- Recommended content may be shared among users, creating possible cascades
- We want users to be exposed to diverse content
- Make a small number of recommendations
    - Why ? Intervene as little as possible
  - Make at most k recommendations in total
  - Make at most k_v recommendations to user v

C. Aslay et al. "*Maximizing the diversity of exposure in a social network*". ICDM 2018

6. Maximizing diversity of exposure

- Problem formulation inspired influence maximization setting
- Consider social graph G=(V,E)
- Assume users have known leaning score s(v)
- Assume set of content items I, each with leaning score s(i)
- Items propagate according to the independence cascade model
  - Influence probabilities are known
- We want to recommend k items to k users
  - (find an assignment from items to users)
- Goal : maximize the diversity score

$$\sum_{v \in V} \left(\max_{i \in E(v)} s(i) - \min_{i \in E(v)} s(i) \right)$$

- E(v) : set of items that v is exposed (considering also cascades)

6. Maximizing diversity of exposure -- results

- Diversity function is submodular
- Greedy algorithm provides ½ approximation
  - Maximizing a submodular function under partition matroid constraints
  - (recommend at most k_v items to user v)
- But computation required in the greedy step by standard Monte-Carlo simulations is prohibitively expensive
- Adapt recently-developed techniques for the influence-maximization problem to obtain highly scalable algorithm
  - Generalize the idea of reverse-reachable sets
  - Estimate the sample size required by greedy using martingales

C. Aslay et al. "*Maximizing the diversity of exposure in a social network*". ICDM 2018

Why is the problem hard?

Belief echoes

- User study (Nyhan and Reifler 2010)
  - Three different political topics:
  - Iraq and WMD, tax cuts, stem cell research
- Findings:
  - Corrective information often fail to reduce misperceptions
  - It may actually strengthen misperceptions among ideological subgroups (backfire effect)
- Similar study and supporting evidence by (Thorson 2016)
  - Exposure to a piece of misinformation can shape a person’s attitudes despite the fact that she recognizes it is false.

Nyhan and Reifler. "*When corrections fail: The persistence of political misperceptions*." Political Behavior 32.2 (2010)

Thorson, "*Belief echoes: The persistent effects of corrected misinformation*." Political Communication 33.3 (2016)

Backfire effect

- Recent study (Bail et al, 2018)
- Surveyed a large sample (N=1652) of politically active twitter users, Democrats and Republicans
- Paid them to follow a Twitter bot for one month that exposed them to content of opposing political ideologies.
- Resurveyed after 1 month
- Finding:
  - Republicans who followed a liberal Twitter bot became substantially more conservative post-treatment
  - Democrats who followed a conservative Twitter bot became slightly more liberal post-treatment

Bail, et al “Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media.” (2018)

Selective partisan sharing

- Partisans do not like fact checking that challenges their views
  (Shin and Thorson 2017)
- Analyzed fact-checking tweets of the 2012 campaign
- Finding 1: Parisans selectively share fact-checking results
  - they hand-pick and promote fact-checking tweets that serve their view
- Finding 2: Fact-checkers receive hostility from the side that is negatively affected by fact-checking
- Fact-checking messages are diffused in a biased manner

Shin and Thorson. "*Partisan selective sharing: The biased diffusion of fact-checking messages on social media*." Journal of Communication 67.2 (2017)

Language can play an important role

- A study on Bing search engine (US, July 2012)
- Found an echo chamber effect
  - Users clicking on news outlets of similar leaning to their own
- But the effect can be mitigated via language
  - higher chance to click of an article of opposite viewpoint if the language model was similar to their side's language model

Yom-Tov, Dumais, and Guo. "*Promoting civil discourse through search engine diversity.*" Social Science Computer Review 32.2 (2014)

Other features can have an effect

- Indicators of the position of the source (Liao and Fu, 2014)
  - Make users aware of other users' position
- Indicators help users who are motivated to acquire more accurate information to decrease their selective exposure
- No effect to users with low motivation

- Credibility of a source, or the expertise of a user, increases the chances of other users believing in the content
  (Vydiswaran et al., 2015)

Liao and Fu, "*Can You Hear Me Now? Mitigating the Echo Chamber Effect by Source Position Indicators*", CSCW 2014

Vydiswaran, Zhai, Roth, and Pirolli, "*Overcoming Bias to Learn about Controversial Topics*", JAIST 2015

Interface can play an important role

- Visualize the topics and tweets of users (Graells-Garrido 2013)
- Organically embed tweets of users with opposing view
- Opposing users may interact, on the basis of good first impression

Graells-Garrido, Lalmas, and Quercia. "*Data portraits: connecting people of opposing views*", arxiv, 2013

Intermediate topics could serve as a middle ground for discussion

- #prochoice, #prolife users
- hardly interact with #prochoice users in a debate context
- although those users could engage in conversation about other interests, such as #musicmonday
- Study verified that such intermediate topics exist
  (Graells-Garrido 2014)

Graells-Garrido, Lalmas, and Quercia. "*People of opposing views can share common interests*", WWW 2014

Summary

- Reducing polarization is beneficial
- Several initiatives and tools for bursting filter bubbles
- Algorithmic mediation / recommendations
- Connecting people with opposing views is also a psychological challenge
  - Many studies have shown that small details matter

Part 6

Challenges and directions for future research

Outline

- Part 1: Introduction
- Part 2: Exploring Polarization
- Part 3: Polarization Models
- Part 4: Measuring Polarization
- Part 5: Mitigating Polarization
- Part 6: Future Research

Wrap-up

- Polarization is an active area of research
  - associated phenomena – filter bubbles and echo chambers
  - inter-disciplinary field
    - psychology, political & social science, statistics & computer science
- We saw efforts to:
  - study instances of polarization on the Web
  - point out mechanisms behind polarization
  - quantify polarization and measure algorithmically from web activity
  - mitigate its negative effects

Addressed some of the following aspects…

- Definition of polarization and related terms
- Psychological and social theories underlying polarized settings
- Why the Web might increase polarization
- How to model polarized social interactions
- Case studies on user activity on the Web
- How to quantify the degree of polarization
- How to reduce polarization / open echo chambers
- Is social media causing polarization to increase?

Challenges

- Many studies with conflicting results
- For instance, studies have supported the following claims:
    - Polarization is not increasing in the society
    - Polarization is increasing but not among the young
    - Internet is not becoming more segregated
    - News consumption is not polarized
    - The effect of filter bubbles is overstated
    - Personalization is not bad
    - Backfire effect does not exist

Why these contradictions?

- Different definitions (of polarization, echo chambers, filter bubbles)
- Different datasets
- Different populations
- Reporting bias

There is still a large chunk of people who are not interested

- Most people pay little or no attention to politics
- The audiences of Fox news and MSNBC are only 2-3 million at most, out of 300 million Americans compared that to audiences for entertainment shows like Big Bang Theory, The Walking Dead, etc. which are in the 10s of millions.
- Similar traffic numbers for Breitbart (10 million) vs. main stream news like NYT and Washington post (70-100 million).

Future research directions

- Measure the extent of polarization and other phenomena
- Modeling
  - different user roles
  - how users react to content from different sides
- Psychological / design challenges
  - users might react negatively to seeing content they do not choose
- Echo chambers
  - do users get out of their echo chambers?
  - combine offline and online data
  - real life consequences of polarization

Future research directions

- Biases in data
  - Representativeness
  - US bias
  - impact of bots
- Ethics of bubble bursting
- Should platforms intervene to…
  - reduce polarization?
  - nudge users outside echo chambers?

Ethics of Bubble Bursting in Search

- Search engine → Filter bubble → Confirmation bias → Echo chambers
- How can search engines present results for polarized topics better?
- Challenges
  - What is the responsibility of the medium (search engine)?
  - Show both sides? Even if one side can be harmful? (vaccines-autism)
- No clear answer
  - Some studies suggest that exposing users to opposing opinions increases their interest in seeking diverse opinions
    Dori-Hacohen, S., Yom-Tov, E., & Allan, J. "Navigating Controversy as a Complex Search Task." (2015)
  - Others show adverse effects.
    Bail CA, Argyle L, Brown T, Bumpus J, Chen H, Hunzaker MF, Lee J, Mann M, Merhout F, Volfovsky A. Exposure to Opposing Views can Increase Political Polarization: Evidence from a Large-Scale Field Experiment on Social Media. (2018)

Thank you