Opinions and conflict in social networks: models, computational problems and algorithms

Lecture 1: Introduction

Bertinoro International Spring School 2022

Aristides Gionis

argioni@kth.se

KTH Royal Institute of Technology

your course instructor

since 2020	WASP professor in KTH, Sweden
2013 – 2019	Aalto University, Finland
2018	visiting fellow in ISI, Turin
2016	visiting professor in Sapienza, Rome
2006 – 2012	group leader in Yahoo! Research, Barcelona
2003 – 2006	postdoc in University of Helsinki
2003	PhD, Stanford University, USA

research interests of course instructor

- theoretical and application-driven research in data mining
- most recent work on graph mining
  - e.g., mining labeled, temporal, signed networks
- structure discovery or optimization problems
  - finding communities, events, dense subgraphs, frequent motifs, role mining, network inference, network alignment, team formation
- study complex dynamic phenomena
  - tracking important nodes and network evolution, information diffusion, reconstruction of cascades and epidemics, opinion formation
- develop novel applications
- design efficient algorithms

course organization

- five lectures
- lecture slides will be available
- pointers to various lecture notes and research papers will be given in the slides
- to pass the course you need to turn in
  1. homework with mathematical / algorithmic questions
  2. small programming project

  details will be given later

nature of the course and required background

- course is broadly on social-network analysis and related phenomena
  - especially focusing on polarization, information spread, opinion dynamics
- course topics are in the intersection of computational social science, algorithms, data analysis, machine learning
- emphasis not only on the particular applications, but also on relevant methodology and computational methods
- technical background: graph theory, linear algebra, (approximation) algorithms, optimization
- aim is to make the course self-contained, but some familiarity with these topics will be useful

learning objectives

- learn key concepts on the topic of modeling conflict, polarization, and opinion dynamics in social networks
- learn how research questions in this application domain can be formulated as mathematical and computational problems
- learn about typical algorithmic approaches and methodologies for solving these computational problems
- enrich your analytical toolkit and learn about methods that could be used to solve problems in other application domains
- obtain hands-on experience in this domain with the homework and programming project

course overview

- lecture 1: introduction
  - polarization in social media; methods for detecting polarization
- lecture 2: mathematical background
  - submodular maximization; spectral graph theory
- lecture 3: methods for mitigating polarization
  - maximizing diversity, balancing information exposure
- lecture 4: signed networks; theory and applications
- lecture 5: opinion dynamics in social networks

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

the early days of euphoria

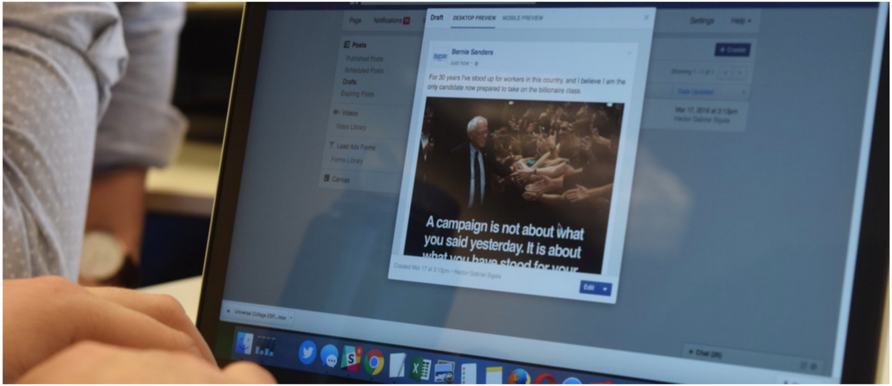
- no information barriers
- democratization of content
- citizen journalism
- social connectivity
- personalization

a different picture has emerged

- polarization, bias, and echo chambers
- radicalization
- misinformation and disinformation
- conversion of popularity into legitimacy
- manipulation and harassment

Global Agenda | Future of Government

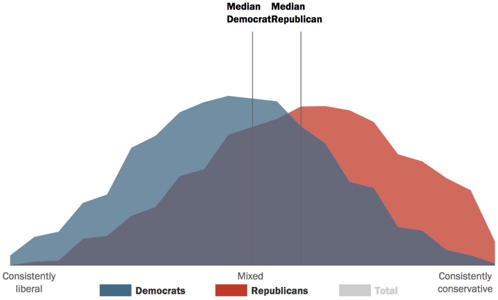
The biggest threat to democracy? Your social media feed



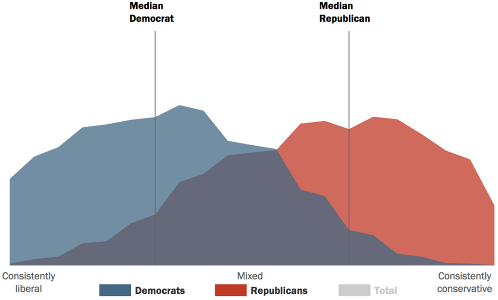
The internet was meant to spread democracy. Could it be having the opposite effect? Image: REUTERS/Melissa Fares

polarization in US politics

1994



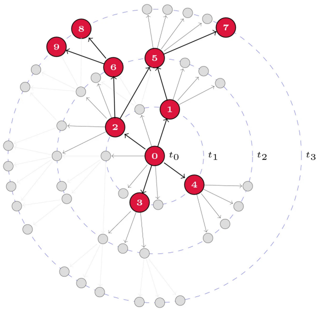
2014



PEW RESEARCH CENTER

we focus on

Baños et al, 2013, EPJ Data Science



- information cascades

  how information propagates in a social network via peer influence

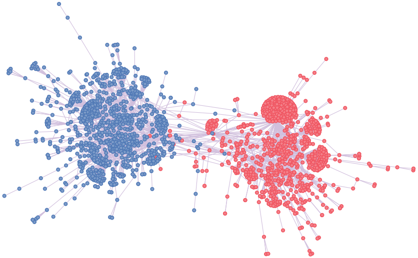
- opinion dynamics

  how people form their opinions via interaction with their peers

we focus on

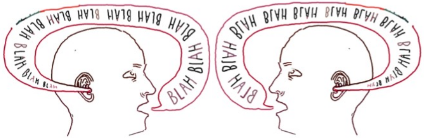
- polarization

  division into sharply contrasting groups or sets of opinions or beliefs



- echo chambers

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



what are the causes? are they necessarily harmful phenomena?

- disagreement is necessary for deliberation and healthy functioning of the society
- individual biases
  - homophily, confirmation bias, cognitive dissonance, selective exposure
- group biases
  - social identity, in-group favoritism
- algorithmic filtering, personalization, "attention economy"
  - users more likely to see what their like-minded share
  - less likely to encounter opposing views
  - system is opaque

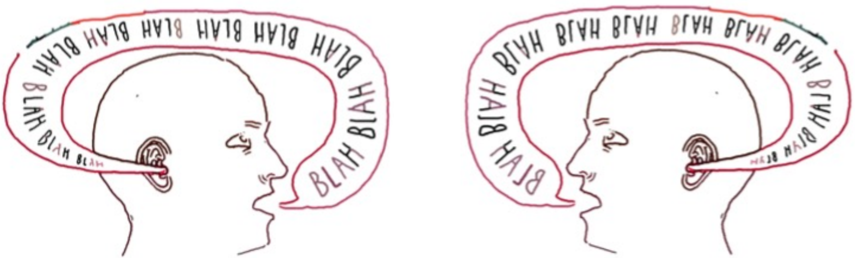
research themes we can study

- DISCOVERY
  - can we detect patterns of bias, polarization, conflict, and lack of information flow in online media?
- EXPLORATION
  - can we help users understand the global information landscape, users gain control on their news diet, and explore alternative view points?
- RECOMMENDATION
  - can we provide recommendations that increase diversity and balance information exposure?

what may cause echo chambers?

echo chambers

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



individual biases

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

individual biases (continued)

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

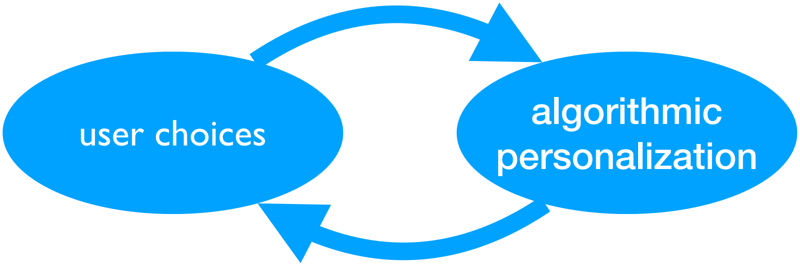
group biases

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

system biases

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

the echo-chamber cycle



case study

do echo chambers exist?

what is the interplay between content and network?

who are the key players?

[Garimella et al., 2018a]

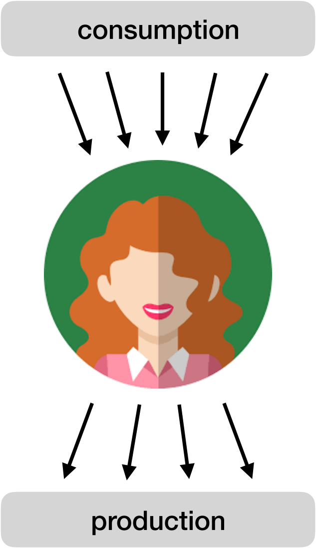
studying echo chambers

- working definition

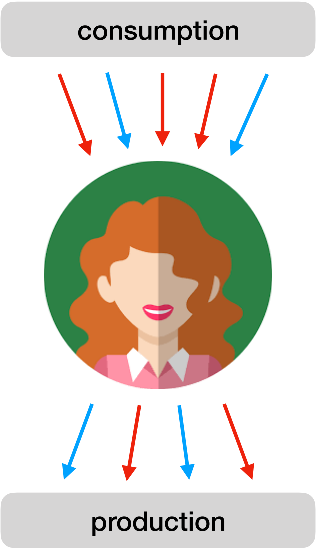
  the political leaning of the content that users receive from the network agrees with that of the content they share

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

methodology



methodology



datasets

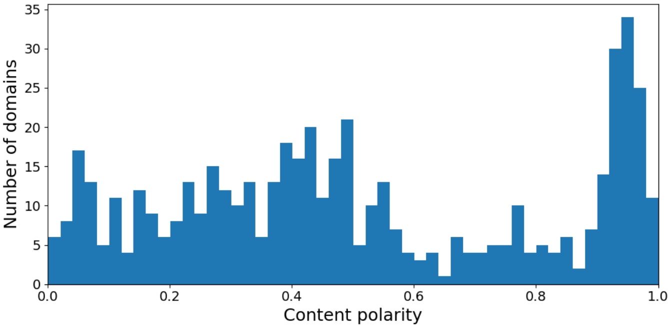
Topic	#Tweets	#Users	Event
guncontrol	19M	7506	Democrat filibuster for gun-control reforms (June 12–18, 2016)[6]
obamacare	39M	8773	Obamacare subsidies preserved in us supreme court ruling (June 22–29, 2015)[7]
abortion	34M	3995	Supreme court strikes down Texas abortion restrictions (June 27–July 3, 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	
#gameofthrones	5M	2159	
#love	3M	2940	filtering for these hashtags
#tbt	28M	12 778	
#foodporn	8M	3904	

content

- focus on news outlets e.g., NYT, BBC, CNN, etc.
- assign content polarity score at each outlet

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

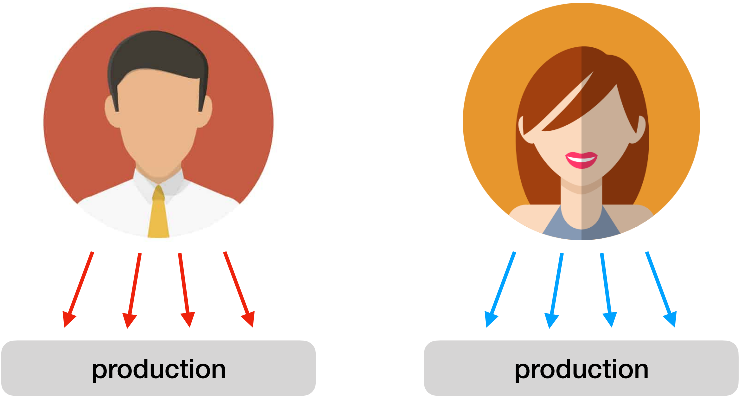
[Bakshy et al., 2015]



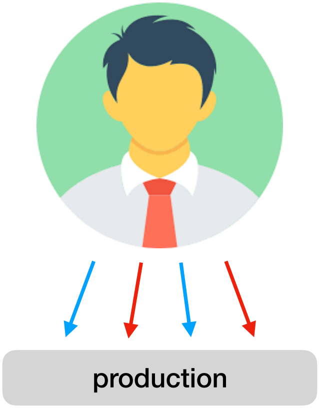
characterize users based on

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

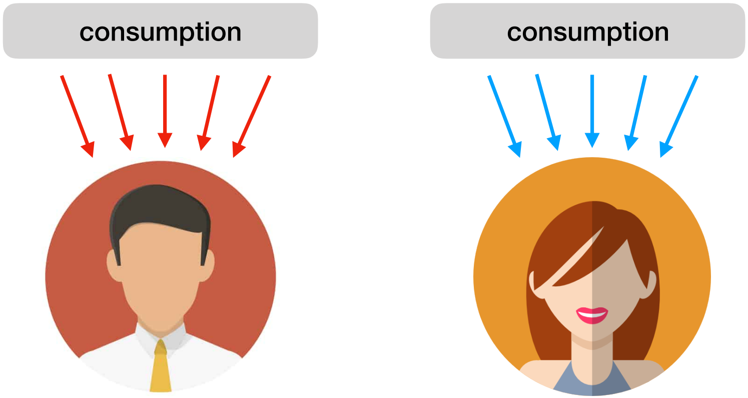
user roles : partisan



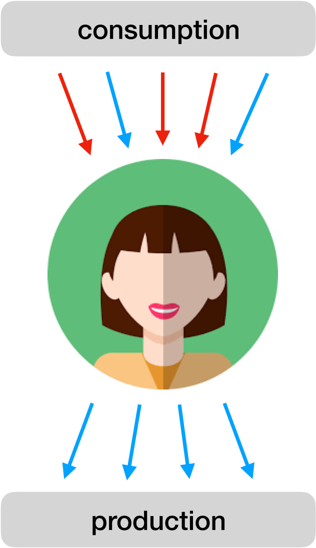
user roles : bi-partisan



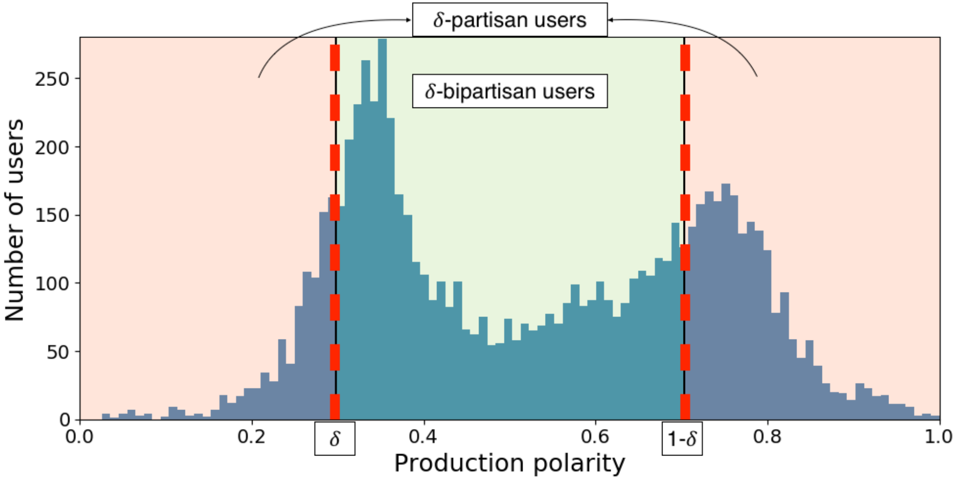
user roles : consumer



user roles : gatekeeper



users — production-polarity distribution



network features

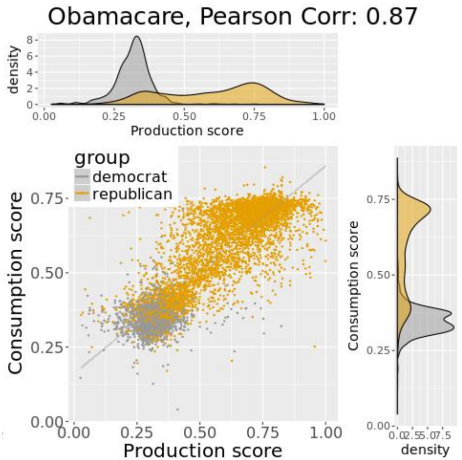
- user polarity (democrat vs. republican) [Barberá, 2015]
- network centrality : PageRank, in-degree
- clustering coefficient
- retweet ratio
- retweet volume

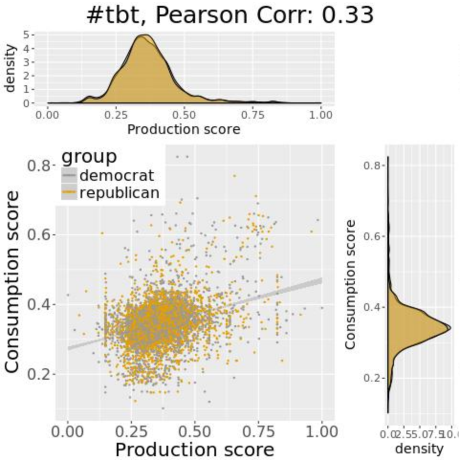
questions

- 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

Table 2: Comparison of various features for partisans & bi-partisans and gatekeepers & non-gatekeepers. A ✓ indicates that the corresponding feature is significantly higher for the group of the column ($p < 0.001$) for at least 4 of the 6 thresholds δ used, for most datasets. A minus next to the checkmark (-) indicates that the feature is significantly lower.

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

Table 2: Comparison of various features for partisans & bi-partisans and gatekeepers & non-gatekeepers. A ✓ indicates that the corresponding feature is significantly higher for the group of the column ($p < 0.001$) for at least 4 of the 6 thresholds δ used, for most datasets. A minus next to the check-mark (-) indicates that the feature is significantly lower.

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

predicting partisans (accuracy ≈ 0.81)

is easier than

predicting gatekeepers (accuracy ≈ 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., open-minded citizens or "soldiers" of one side?

identifying controversy and polarization in social media

how can we identify controversy and 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 ?

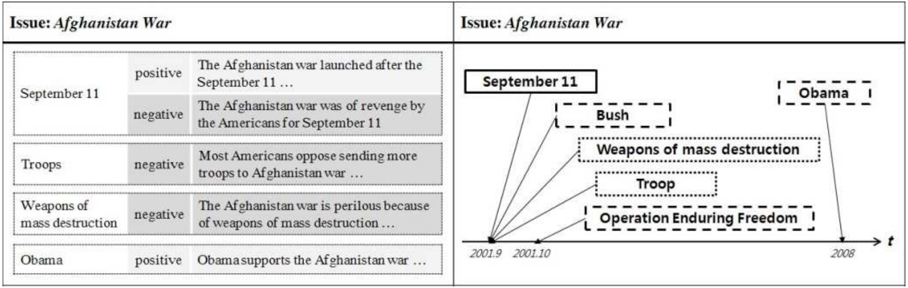
sentiment variance in news

- controversial topic — a concept that invokes conflicting sentiments
- subtopic — an aspect that induces 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
- idea has been explored by different works

sentiment variance

- method: [Choi et al., 2010]
  - identify candidate entities (noun phrases)
  - compute sentiment in sentences involving these entities
  - controversial if

$$\text{positive\_sentiment} + \text{negative\_sentiment} > \delta \text{ and}$$
$$|\text{positive\_sentiment} - \text{negative\_sentiment}| > \gamma$$



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 et al., 2014]

- create a controversy lexicon via crowdsourcing
- use lexicon + sentiment analysis tools to identify controversy
- controversial topics are shown to 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"

Table 1: List of words identified during the crowdsourcing task.

Strongly Controversial (145):	*abuse, administration, afghanistan, aid, america, american, army, attack, attacks, authorities, authority, ban, banks, benefits, bill, bills, border, budget, campaign, candidate, candidates, catholic, china, chinese, church, concerns, congress, conservative, control, country, court, crime, criminal, crisis, cuts, debate, debt, defense, deficit, democrats, disease, dollar, drug, drugs, economy, education, egypt, election, elections, enforcement, fighting, finance, fiscal, force, funding, gas, government, gun, health, immigration, inaccuracies, india, insurance, investigation, investigators, iran, israel, job, jobs, judge, justice, killing, korea, labor, land, law, lawmakers, laws, lawsuit, leadership, legislation, marriage, media, mexico, military, money, murder, nation, nations, news, obama, offensive, officials, oil, parties, peace, police, policies, policy, politics, poll, power, president, prices, primary, prison, progress, race, reform, republican, republicans, restrictions, rule, rules, ruling, russia, russian, school, security, senate, sex, shooting, society, spending, strategy, strike, support, syria, syrian, tax, taxes, threat, trial, unemployment, union, usa, victim, victims, violence, vote, voters, war, washington, weapons, world*
Somewhat Controversial (45):	*account, advantage, amount, attorney, chairman, charge, charges, cities, class, comment, companies, cost, credit, delays, effect, expectations, families, family, february, germany, goal, housing, information, investment, markets, numbers, oklahoma, parents, patients, population, price, projects, raise, rate, reason, sales, schools, sector, shot, source, sources, status, stock, store, worth*
Non-Controversial (272):	*60s, 70s, addition, address, afternoon, agreed, amp, angeles, answer, april, attention, avenue, average, ball, base, bay, beach, beginning, bit, block, blue, bowl, box, boy, boys, brother, building, bus, call, calling, calls, camp, car, cars, central, cents, click, close, cloudy, club, coast, cup, dallas, date, daughter, davis, day, decade, decades, december, def, delivery, door, download, drive, eagles, end, entire, era, evening, face, faces, facility, fall, fans, father, feel, feeling, feet, fell, field, finish, floor, form, fort, francisco, friday, friend, friends, fun, girl, girls, ground, gt, guy, guys, half, hall, hand, hands, hawaii, heart, heat, heavy, hill, hits, hold, hopes, host, hotel, hour, hours, house, houston, hundreds, husband, ice, illinois, index, indiana, innings, island, january, johnson, jones, june, kansas, kind, lack, lake, leave, lee, letter, levels, light, line, lines, lot, lows, lt, main, make, makes, mark, mass, material, matter, medium, men, mid, middle, miles, mind, minneapolis, minutes, moment, monday, month, months, morning, mother, mountain, move, mph, museum, names, natural, net, night, north, note, notes, november, october, opening, park, part, parts, pass, period, person, philadelphia, pick, pitch, plant, play, player, playing, pm, point, post, practice, put, quarter, rain, read, reading, red, rest, restaurant, rise, rock, rose, round, sale, san, saturday, scene, search, season, seasons, seconds, selling, september, series, set, showers, showing, shows, sign, signs, smith, son, sox, special, spot, spring, square, stadium, stage, start, starting, starts, station, stay, step, stores, street, student, summer, sun, sunday, thing, things, thinking, thought, thousands, thunderstorms, thursday, time, title, top, total, transportation, type, unit, valley, vehicle, version, village, visit, wait, walk, wall, watch, water, ways, wednesday, week, weekend, weeks, williams, wind, winds, winner, winter, word, writer, yards, year, years, york*

detecting controversy on the Web

- problem: finding if a web page discusses a (known) controversial topic

  [Dori-Hacohen and Allan, 2013]
- map topics (named entities) in the 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 research on identifying controversial topics on wikipedia
  - edit wars, hyperlink structure, etc.
- related:
  - in addition, language models can be built to directly detect controversy

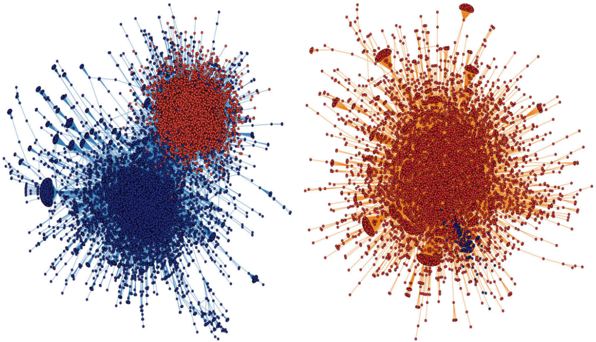
  [Jang et al., 2016]

identifying polarization using network structure

- methods rely on analyzing the structure of a network
  - different networks to consider: social media, hyperlinks, wikipedia edits, etc.
- in particular, a lot of research has been done using twitter data
  - retweet network
  - reply network
  - social network (follow)
- idea: controversial topics exhibit clustered structure in (some) associated network

political polarization on twitter

- retweet network for political hashtags has a bi-clustered structure [Conover et al., 2011]
  - retweet network exhibits a highly modular structure, segregating users into two homogeneous communities corresponding to the political left and right
- users mention/reply to others from their opposing viewpoint, resulting in non-segregated structure

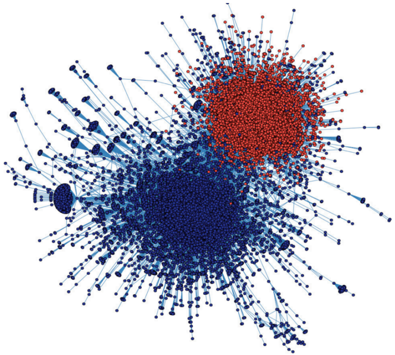


Retweet Reply

quantifying polarization

[Conover et al., 2011]

- quantify polarization using 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

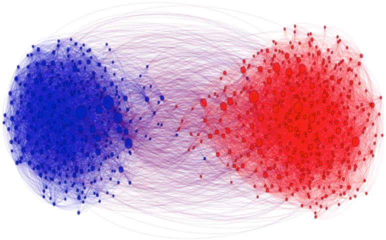


Modularity: 0.48

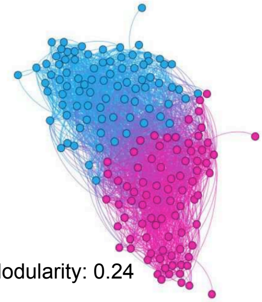
modularity not direct measure of polarization

[Guerra et al., 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

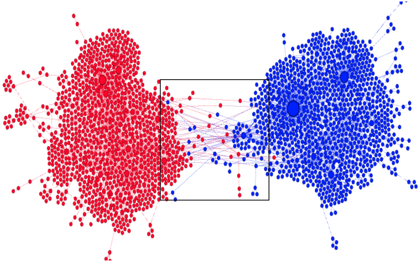


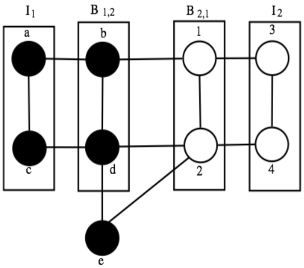
Modularity: 0.24

community boundary

[Guerra et al., 2013]

- boundary nodes
  - 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
- for a boundary node v, define $P(v) = d_{in}(v)/(d_{out}(v) + d_{in}(v)) - \frac{1}{2}$
  - $P(v) > 0$ indicates that node prefers internal connections (antagonism?)
  - $P(v) < 0$ indicates that node prefers connections with members of the other group
- polarization measure: average $P(v)$ value over all boundary nodes





case study

quantifying controversy on social media

[Garimella et al., 2018b]

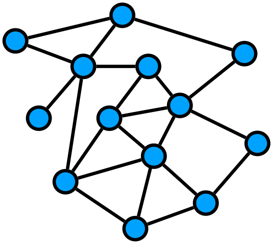
how can we identify polarization ?

ideas

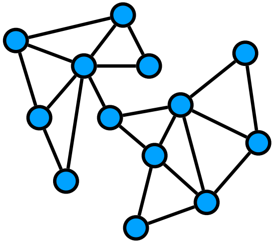
- content
  - do opposing sides say different things ?
- sentiment
  - do polarized topics exhibit wider range of emotions ?
- interactions
  - do people interact more with their own side ?

high-level approach

- build an interaction graph
- is the interaction graph polarized?
- output polarization score



non polarized



polarized
two sides well separated

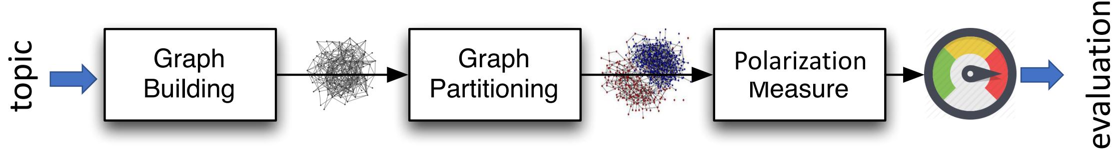
many different options

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?



retweets
replies
connections

any state-of-the-art algorithm

random-walk
edge betweenness
embedding-based

random-walk controversy score (RWC)

- how likely a random user on either side to be exposed to authoritative content from the opposing side
- 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)

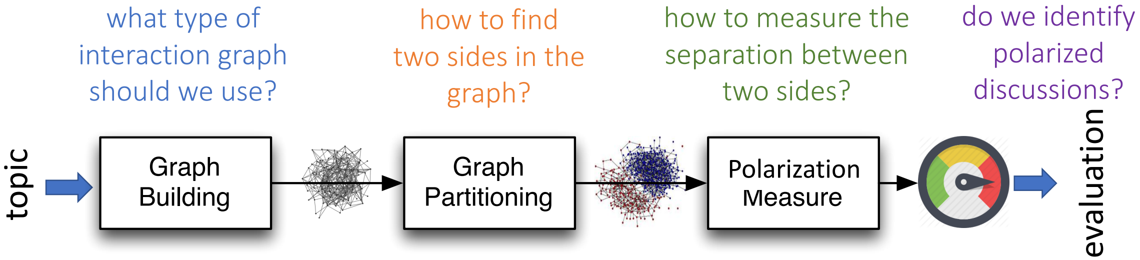
$$\textsc{rwc} = P_{AA}P_{BB} - P_{AB}P_{BA}$$

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, ukraine, baltimore riots
- 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

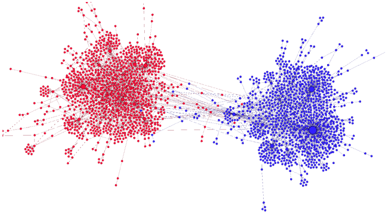
sentiment variance

example of results

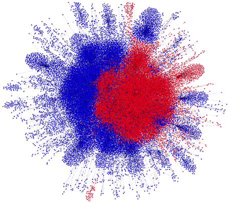
high RWC | low RWC

polarized topics

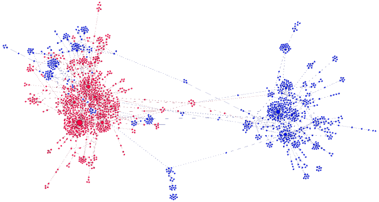
nemtsov protests



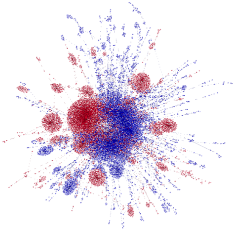
germanwings plane crash



indian beef ban



sxsw conference

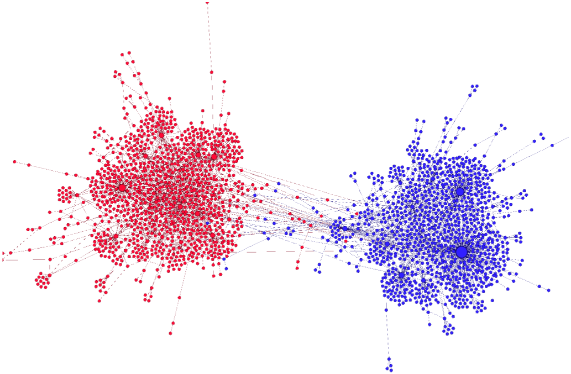


non-polarized topics

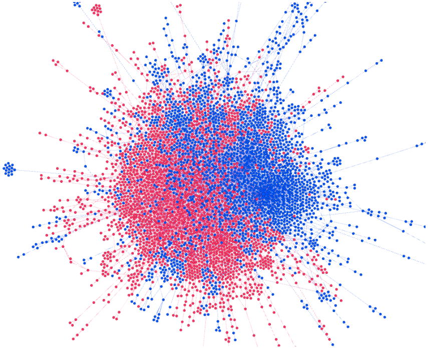
using retweet graph

example of results

nemtsov protests



retweets



replies

in summary

- an overview of polarization and echo chambers in social media
- discussed biases that may lead to these phenomena
- overviewed different methods and use cases for identifying and quantifying polarization
- coming lectures: dive into more technical aspects and mathematical / algorithmic tools

references I

Bakshy, E., Messing, S., and Adamic, L. A. (2015).
Exposure to ideologically diverse news and opinion on facebook.
Science, 348(6239):1130–1132.

Barberá, P. (2015).
Birds of the same feather tweet together: Bayesian ideal point estimation using twitter data.
Political analysis, 23(1):76–91.

Choi, Y., Jung, Y., and Myaeng, S.-H. (2010).
Identifying controversial issues and their sub-topics in news articles.
In *Pacific-Asia Workshop on Intelligence and Security Informatics*, pages 140–153. Springer.

Conover, M., Ratkiewicz, J., Francisco, M., Gonçalves, B., Menczer, F., and Flammini, A. (2011).
Political polarization on twitter.
Proceedings of the International AAAI Conference on Web and Social Media, 5(1):89–96.

Dori-Hacohen, S. and Allan, J. (2013).
Detecting controversy on the web.
In *Proceedings of the 22nd ACM international conference on Information & Knowledge Management*, pages 1845–1848.

references II

Garimella, K., De Francisci Morales, G., Gionis, A., and Mathioudakis, M. (2018a).
Political discourse on social media: Echo chambers, gatekeepers, and the price of bipartisanship.
In *Proceedings of the 2018 world wide web conference*, pages 913–922.

Garimella, K., Morales, G. D. F., Gionis, A., and Mathioudakis, M. (2018b).
Quantifying controversy on social media.
ACM Transactions on Social Computing, 1(1):1–27.

Guerra, P., Meira Jr, W., Cardie, C., and Kleinberg, R. (2013).
A measure of polarization on social media networks based on community boundaries.
Proceedings of the international AAAI conference on web and social media, 7(1):215–224.

Jang, M., Foley, J., Dori-Hacohen, S., and Allan, J. (2016).
Probabilistic approaches to controversy detection.
In *Proceedings of the 25th ACM international on conference on information and knowledge management*, pages 2069–2072.

Mejova, Y., Zhang, A. X., Diakopoulos, N., and Castillo, C. (2014).
Controversy and sentiment in online news.
arXiv preprint arXiv:1409.8152.