UNDERSTANDING FILTER BUBBLES AND POLARIZATION IN SOCIAL NETWORKS

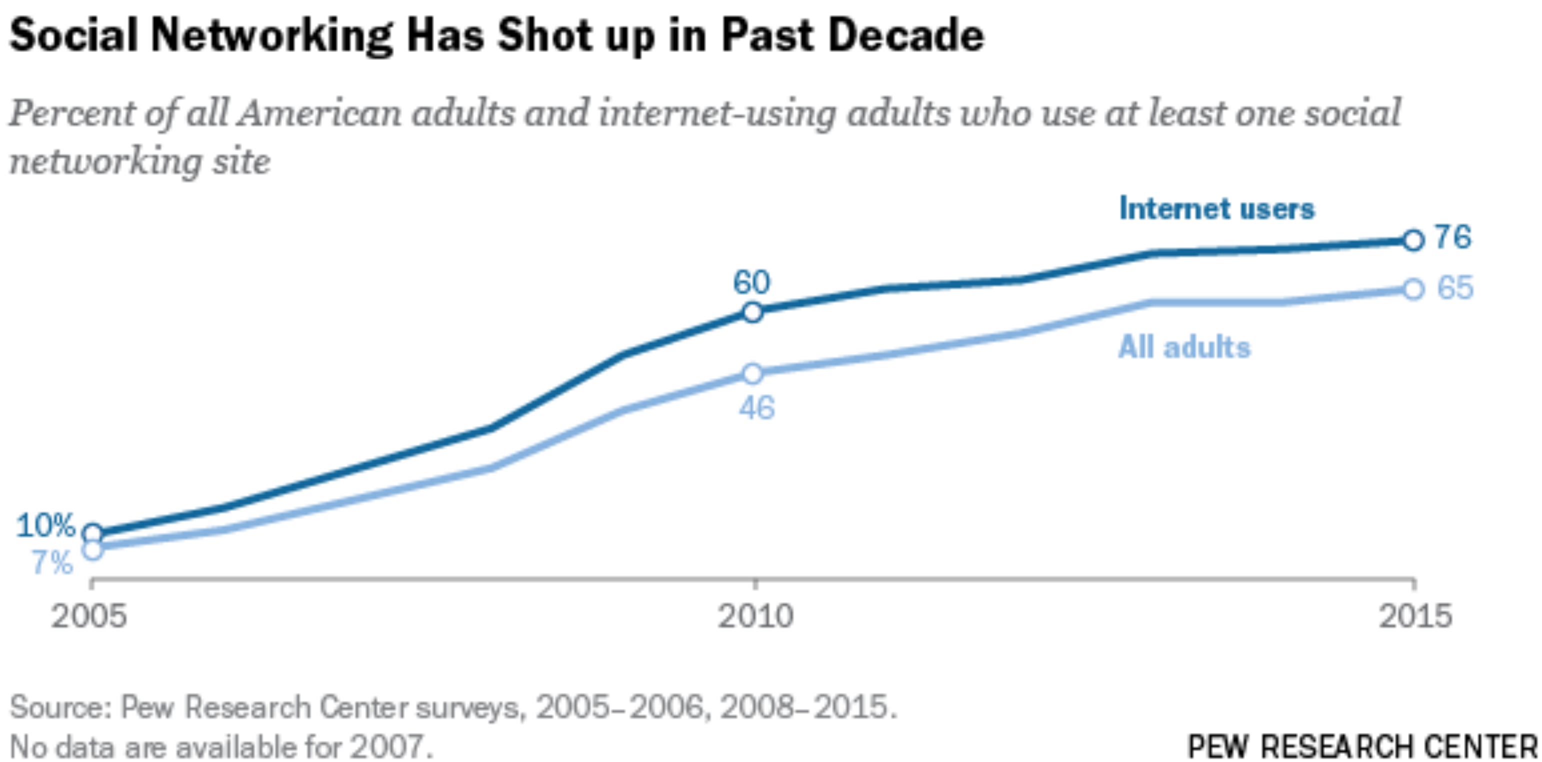
Uthsav Chitra and Christopher Musco
Princeton University

The Impact of Social Networks

In past 10 years, social media usage has skyrocketed

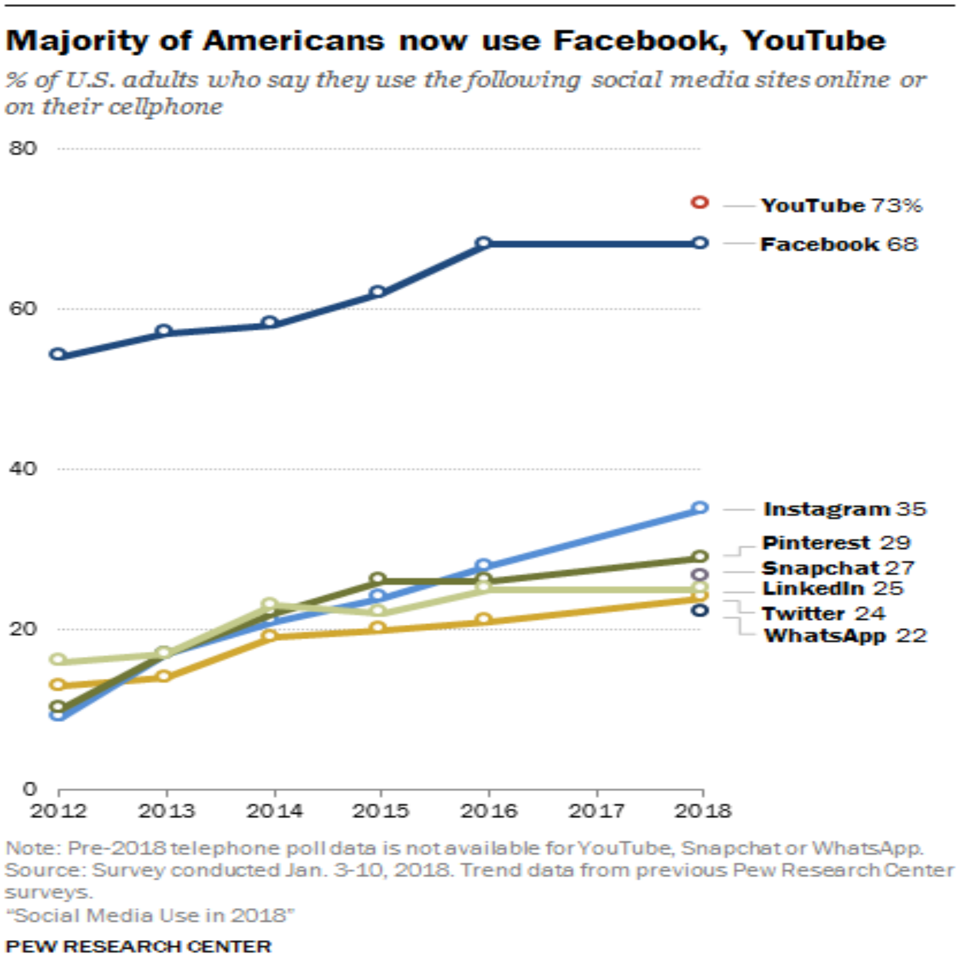
The Impact of Social Networks

In past 10 years, social media usage has skyrocketed



The Impact of Social Networks

In past 10 years, social media usage has skyrocketed



The Impact of Social Networks

Well-known that social media has made world more connected

- easier to get information than ever before

The Impact of Social Networks

Yet surprisingly, social networks are also linked to **increased polarization** across society.

The Impact of Social Networks

Yet surprisingly, social networks are also linked to **increased polarization** across society.

Social media has been blamed for polarization and the spread of misinformation:

- In 2016 election and Brexit [1]
- In protests against immigration in Europe [2]
- And even in measles outbreaks in 2014, 2015 [3]

References:
[1] "Eli Pariser: activist whose filter bubble warnings presaged Trump and Brexit…", Jackson. *The Guardian,* 2017
[2] "The triple-filter bubble…" Geschke, Lorenz, Holtz. *British Journal of Social Psychology* 2019
[3] "The filter bubble and its effect on online personal health information", Holone. *Croatian Medical Journal* 2019.

The Puzzle of Polarization

Two seemingly contradictory facts

1. Social networks make world **more open and connected**
2. Social networks have resulted in **increased polarization** in society

The Puzzle of Polarization

Two seemingly contradictory facts

1. Social networks make world **more open and connected**
2. Social networks have resulted in **increased polarization** in society

Why?

The Puzzle of Polarization

Problem has been studied in psychology

The Puzzle of Polarization

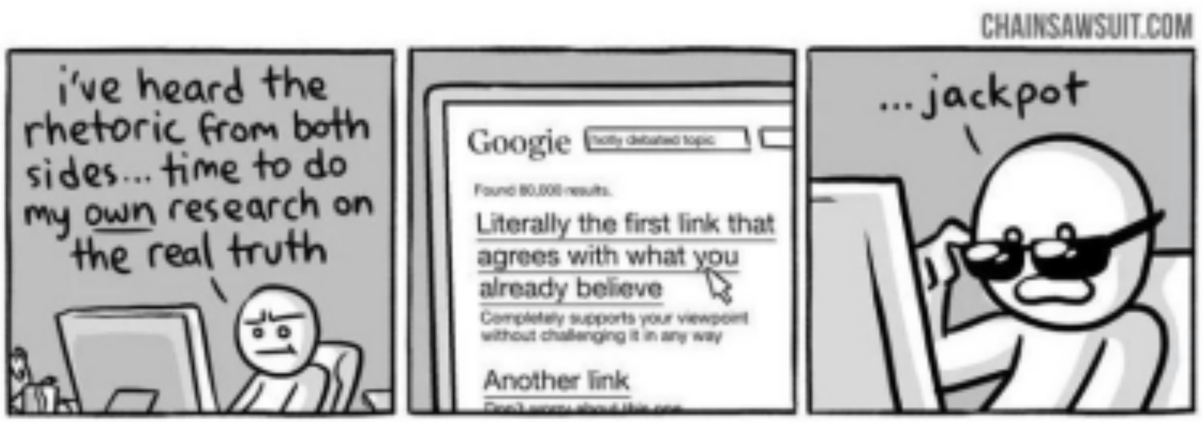
Problem has been studied in psychology

- Prevailing theory: individuals are more likely to trust/share information that already aligns with their views

The Puzzle of Polarization

Problem has been studied in psychology

- Prevailing theory: individuals are more likely to trust/share information that already aligns with their views



- Known as "biased assimilation"

Biased Assimilation in the Internet Era

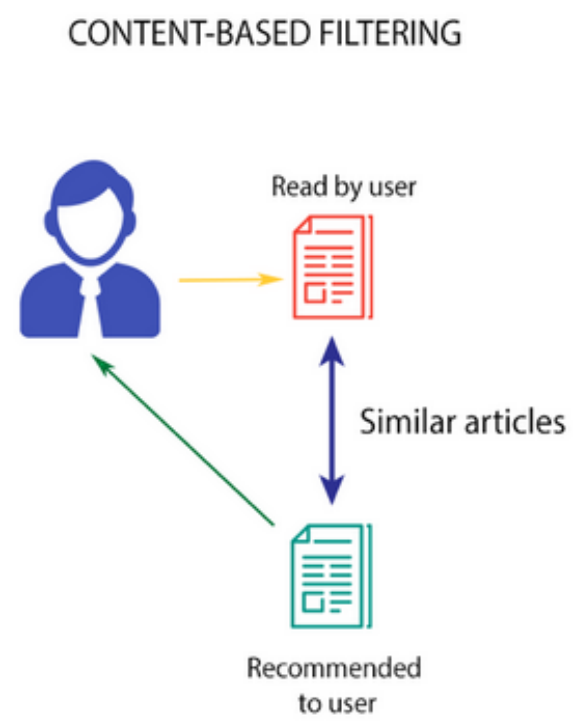
Social media companies *explicitly encourage* users to consume content that aligns with their views

Biased Assimilation in the Internet Era

Social media companies *explicitly encourage* users to consume content that aligns with their views

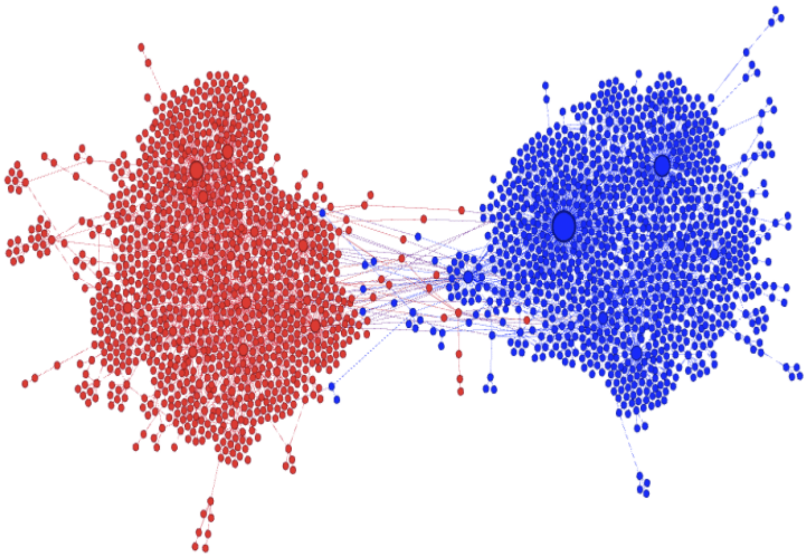
Examples:

- Twitter - follow suggestions
- Facebook - personalized news feed
- Youtube - curated playlists



Filter Bubbles

"Filter bubble" theory (Pariser, 2011): Through recommender systems and content filtering, social media companies create echo chambers of like-minded individuals



Filter Bubbles

However, magnitude of the filter bubble effect has been disputed (e.g. [4])

[4] “Exploring the filter bubble: the effect of using recommender systems on content diversity” Nguyen, Hui, Harper, Terveen, Konstan. WWW 2014.

Filter Bubbles

However, magnitude of the filter bubble effect has been disputed (e.g. [4])

Our goal: develop a mathematical framework to better justify and understand the filter bubble theory

[4] "Exploring the filter bubble: the effect of using recommender systems on content diversity" Nguyen, Hui, Harper, Terveen, Konstan. WWW 2014.

Outline

1. Friedkin-Johnsen model for opinion dynamics
2. Introducing the Network Administrator
3. Results
    1. Experiments on Reddit and Twitter networks
    2. Theoretical arguments
4. A simple remedy to reduce polarization
5. Conclusion

Mathematical Framework

The Friedkin-Johnsen dynamics model the flow of an information in a social network.

Because of its simplicity, the Friedkin-Johnsen model is well-studied -- often used to study social/economic networks, e.g. [5, 6, 7, 8]

[5] “Modeling opinion dynamics in social networks”, Das, Gollapudi, Munagala, WSDM 2014.
[6] “How Bad is Forming Your Own Opinion”, Bindel, Kleinberg, Oren, FOCS 2011.
[7] "Measuring and Moderating Opinion Polarization in Social Networks.”, Matakos, Terzi, Tsaparas, Data Min. Knowl. Discov. 2017
[8] “Opinion dynamics with varying susceptibility to persuasion”, Abebe, Kleinberg, Parkes, Tsourakakes, KDD 2018.

Mathematical Framework

The Friedkin-Johnsen dynamics model the propagation of an opinion during a series of discrete time steps, $t = 0, 1, 2, \ldots$

Mathematical Framework

The Friedkin-Johnsen dynamics model the propagation of an opinion during a series of discrete time steps, t = 0, 1, 2, …

The opinion can be anything, specific or broad.

- Should we remove the carried interest loophole?
- Are your views more conservative or liberal?

Friedkin-Johnsen Model

Each node in the social network has:

Friedkin-Johnsen Model

Each node in the social network has:

1. s, its innate opinion
   - Reflects internal beliefs; does not change over time

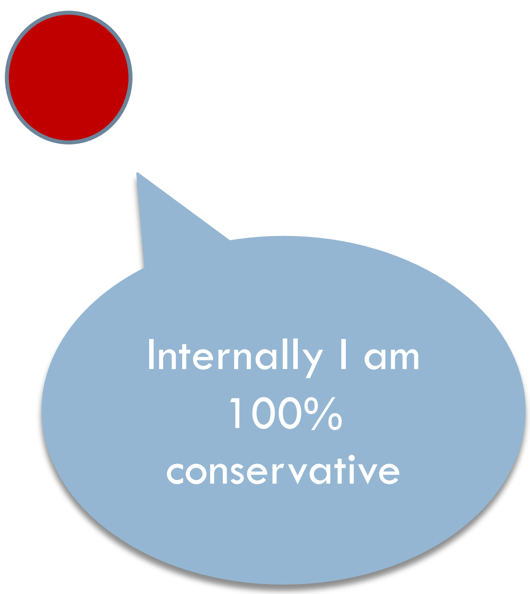
Friedkin-Johnsen Model

Each node in the social network has:

1. *s*, its innate opinion
   - Reflects internal beliefs; does not change over time
2. *z*, its expressed opinion
   - Others only see expressed opinions

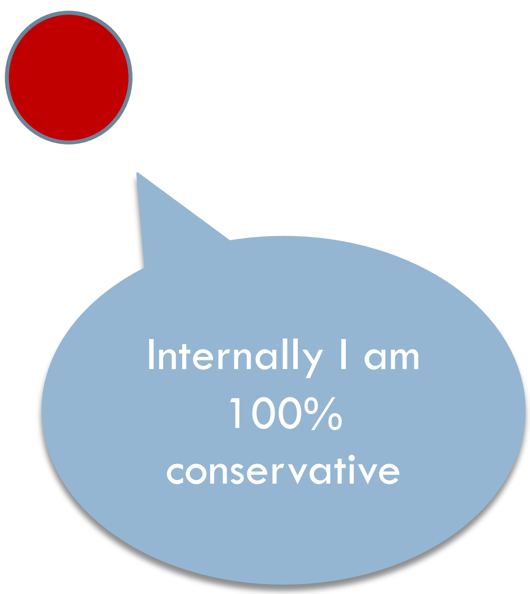
Friedkin-Johnsen Model

innate opinion s

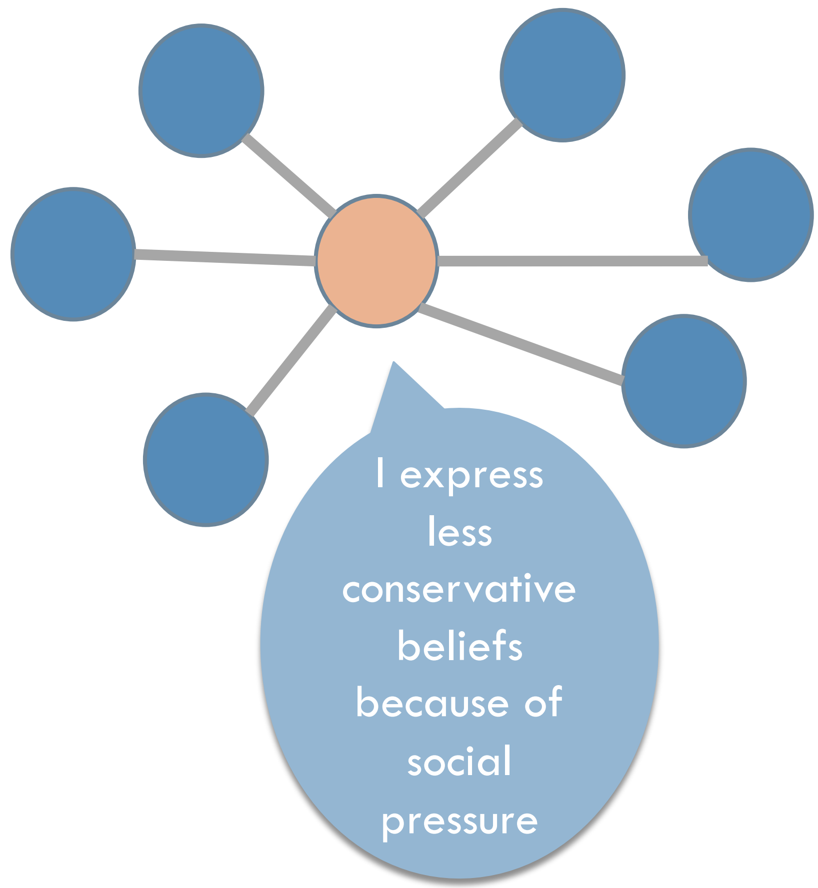


Friedkin-Johnsen Model

innate opinion *s*



expressed opinion *z*



Friedkin-Johnsen Model

Formally, let G be a graph, with:

- nodes $v_1, \ldots, v_n$, edge weights w_{ij}
- innate opinions $s_i \in [-1, 1]$
- expressed opinions $z_i^{(t)} \in [-1, 1]$

Friedkin-Johnsen Model

Formally, let G be a graph, with:

- nodes $v_1, \ldots, v_n$, edge weights w_{ij}
- innate opinions $s_i \in [-1, 1]$
- expressed opinions $z_i^{(t)} \in [-1, 1]$

At time t, expressed opinions are average of innate opinion and neighbors' expressed opinions:

Friedkin-Johnsen Model

Formally, let G be a graph, with:

- nodes $v_1, \ldots, v_n$, edge weights w_{ij}
- innate opinions $s_i \in [-1, 1]$
- expressed opinions $z_i^{(t)} \in [-1, 1]$

At time t, expressed opinions are average of innate opinion and neighbors' expressed opinions:

$$z_i^{(t)} = \frac{s_i + \sum_{j \neq i} w_{ij} z_j^{(t-1)}}{1 + \sum_{j \neq i} w_{ij}}$$

Friedkin-Johnsen Model

Can be shown that opinions converge to an **equilibrium:** $\lim_{t \to \infty} z_i^{(t)} \to z^*$

Friedkin-Johnsen Model

Can be shown that opinions converge to an **equilibrium:** $\lim_{t \to \infty} z_i^{(t)} \to z^*$

Equilibrium opinions $z^* = (L + I)^{-1} s$, where L is graph Laplacian

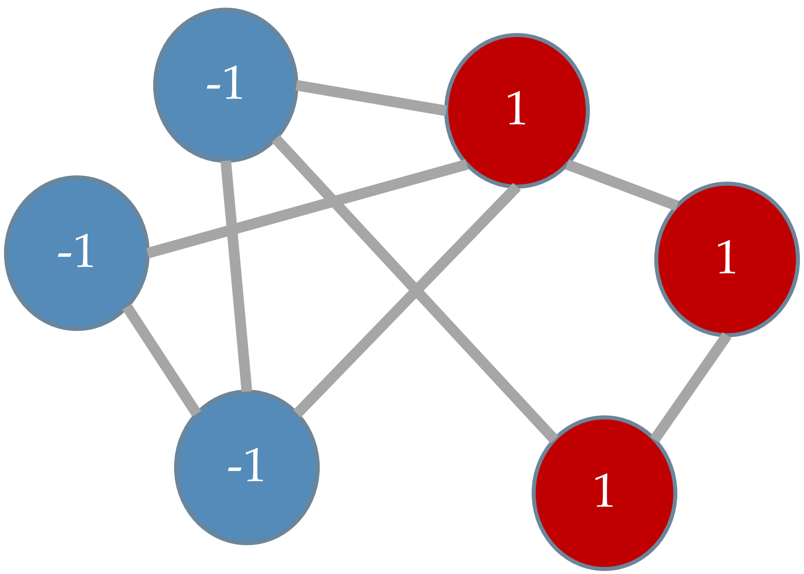
Note: Equilibrium opinions not necessarily all equal (i.e. no consensus)

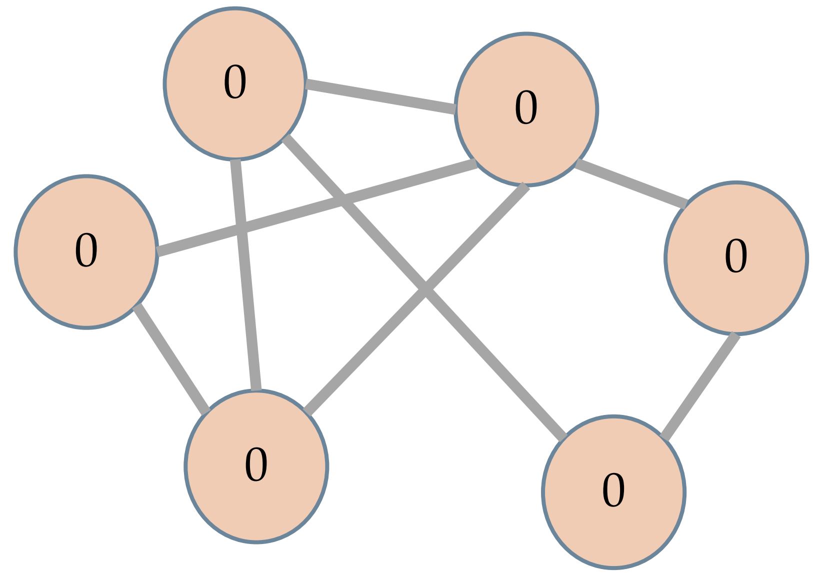
Polarization

One natural definition of polarization is the variance of (equilibrium) expressed opinions.

Polarization

One natural definition of polarization is the variance of (equilibrium) expressed opinions.





<u>Large polarization</u>

<u>Small polarization (reached consensus)</u>

Defining Disagreement

Another metric is disagreement

$$\mathcal{D}_{\mathbf{z}} = \sum_{i<j} w_{ij}(z_i - z_j)^2$$

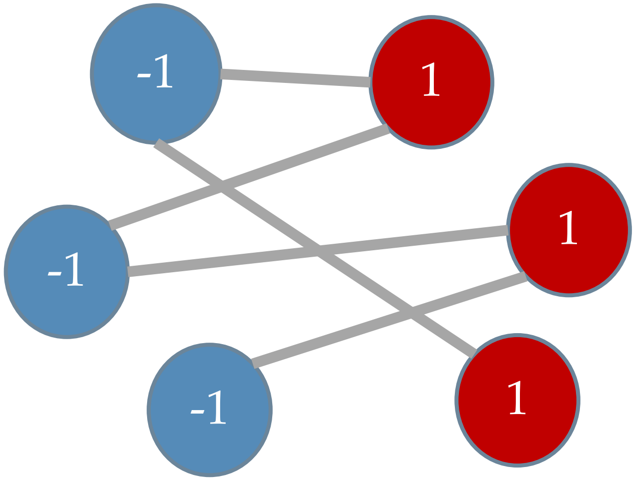
- Measures how much node's opinion differs from neighbors
- Important for studying algorithmic content filtering

Defining Disagreement

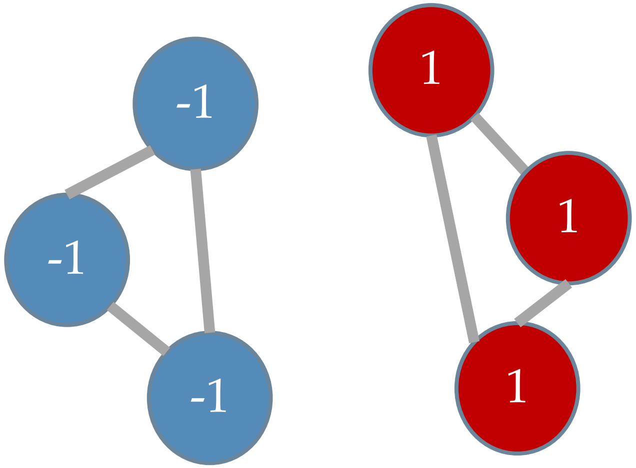
Another metric is disagreement

$$\mathcal{D}_{\mathbf{z}} = \sum_{i<j} w_{ij}(z_i - z_j)^2$$

- Measures how much node's opinion differs from neighbors
- Important for studying algorithmic content filtering



Large disagreement

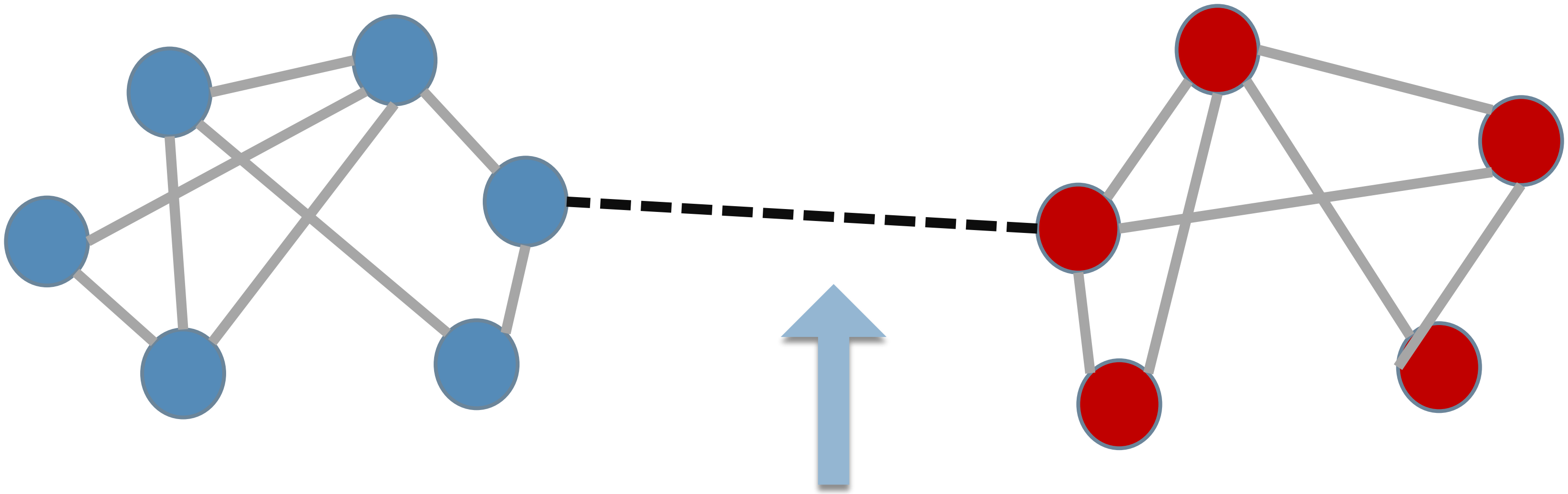


Small disagreement

Previous Literature

Previous work studies polarization in Friedkin-Johnsen model, e.g. polarization minimization is studied in

- Musco, Musco, Tsourakakis, WWW 2018
- Chen, Lijffijt, De Bie, KDD 2018

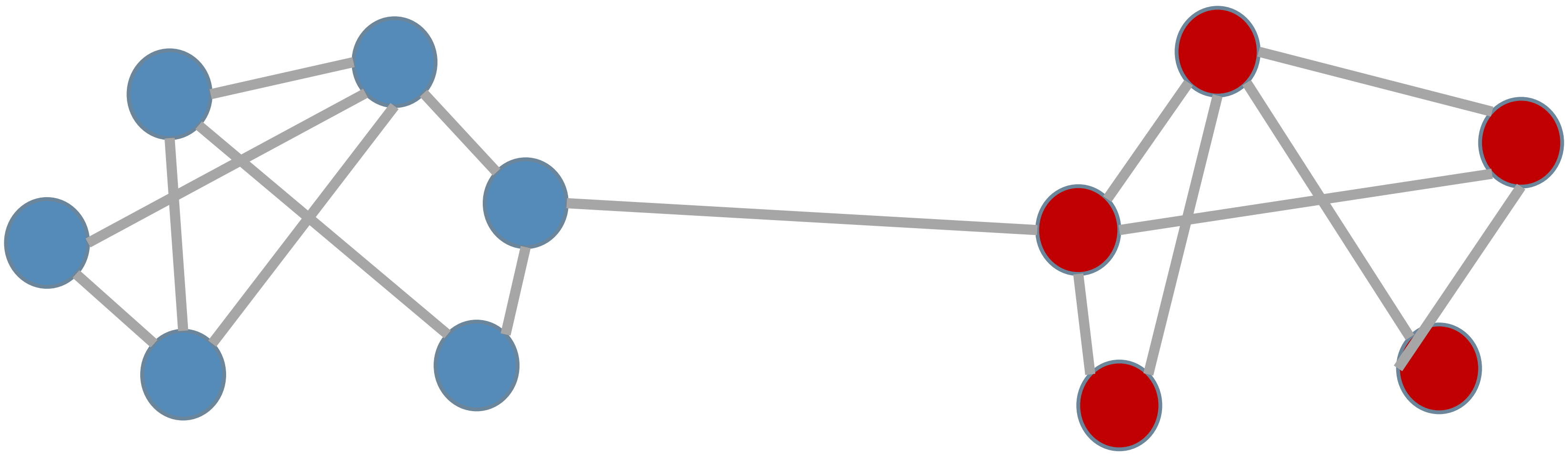


Adding this edge will reduce polarization

Previous Literature

Our work: study polarization **formation** in social networks

i.e. "How did the network become so polarized?"



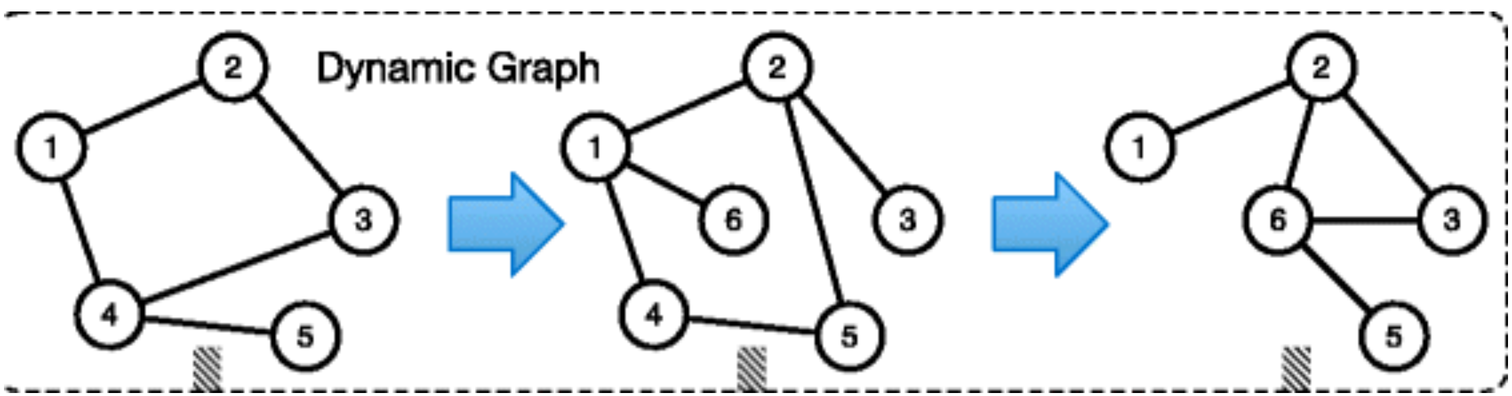
Outline

1. Friedkin-Johnsen model for opinion dynamics
2. Introducing the Network Administrator
3. Results
   1. Experiments on Reddit and Twitter networks
   2. Theoretical arguments
4. A simple remedy to reduce polarization
5. Conclusion

Motivation

One deficiency of Friedkin-Johnsen model: cannot account for dynamic graphs

- Because of algorithmic content filtering, social networks change over time



Network Administrator

Our solution: Introduce a network administrator to Friedkin-Johnsen model

- Make small changes to the network over time
- Models content filtering in social networks

Network Administrator

How would a network administrator change the network?

Network Administrator

How would a network administrator change the network?

- A network administrator models recommender systems, which maximize metrics like engagement or ad revenue

Network Administrator

How would a network administrator change the network?

- A network administrator models recommender systems, which maximize metrics like engagement or ad revenue
- In the Friedkin-Johnsen model, a proxy is minimizing **disagreement**

$$\mathcal{D}_{\mathbf{z}} = \sum_{i<j} w_{ij}(z_i - z_j)^2$$

Network Administrator

Informally, network administrator solves following **optimization** problem

$$\min_{\text{graph } G} \mathcal{D}_{\mathbf{z}}$$

- Where the network administrator can only pick graphs G that are "close" to the original social network

Network Administrator

Example:

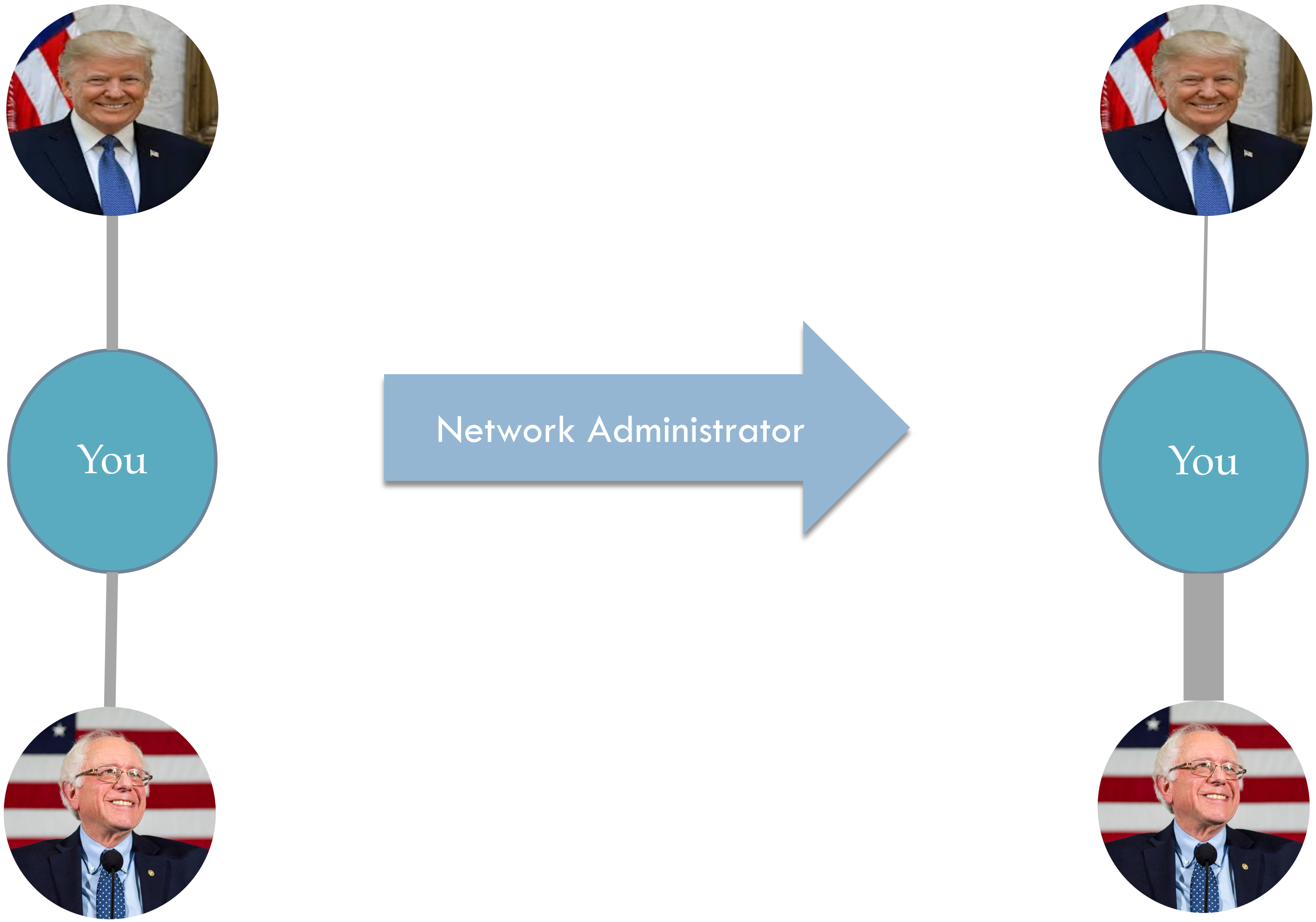
- Edge weights w_{ij} = how often person i sees person j in news feed
- Network administrator = news feed algorithm

Network Administrator

You are friends with Donald Trump and Bernie Sanders on Facebook.

You have a slight liberal lean.

Network Administrator



Network Administrator Dynamics

Model algorithmic filtering via an alternating game:

1. Fixing expressed opinions, network administrator changes graph, to minimize disagreement

Network Administrator Dynamics

Model algorithmic filtering via an alternating game:

1. Fixing expressed opinions, network administrator changes graph, to minimize disagreement

(Note: Network administrator can only make small changes to graph.)

Network Administrator Dynamics

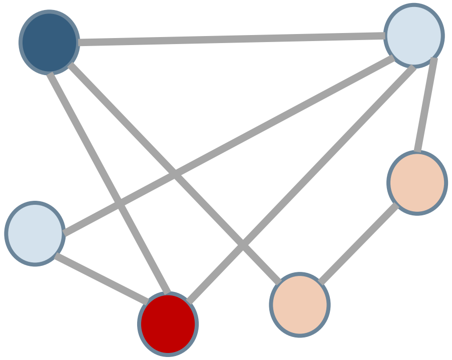
Model algorithmic filtering via an alternating game:

1. Fixing expressed opinions, network administrator changes graph, to minimize disagreement

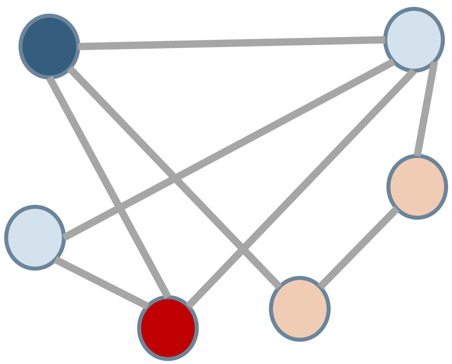
   (Note: Network administrator can only make small changes to graph.)

2. Fixing graph, users adopt new (equilibrium) expressed opinions

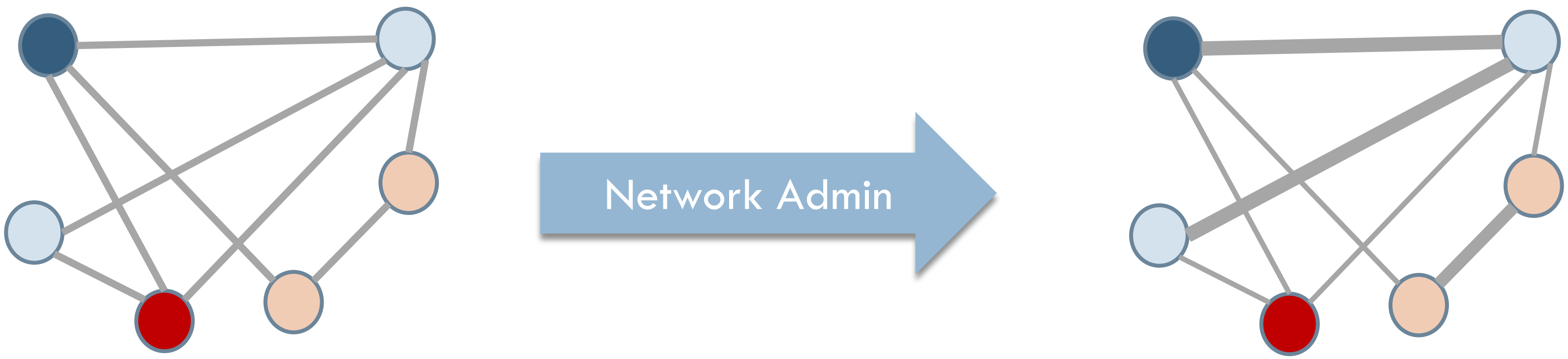
Network Administrator Dynamics



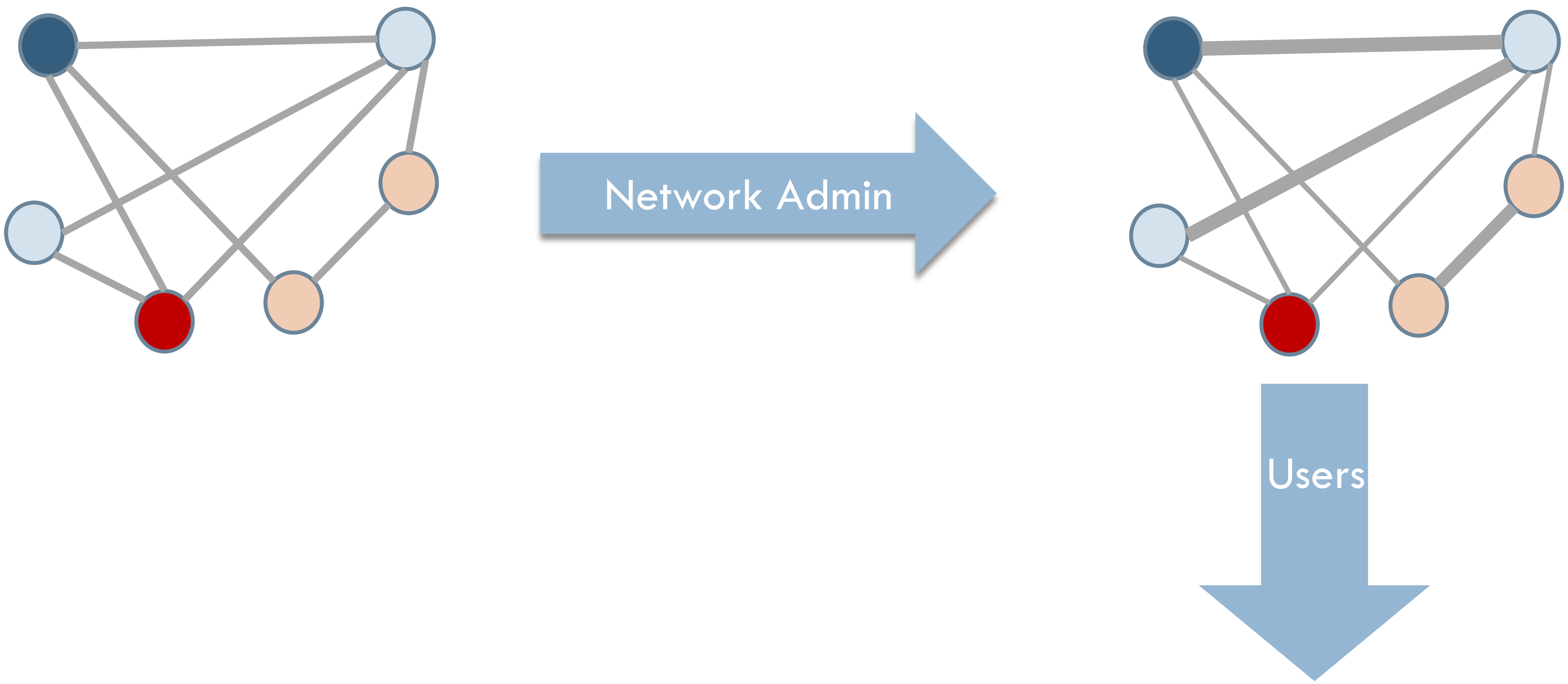
Network Administrator Dynamics



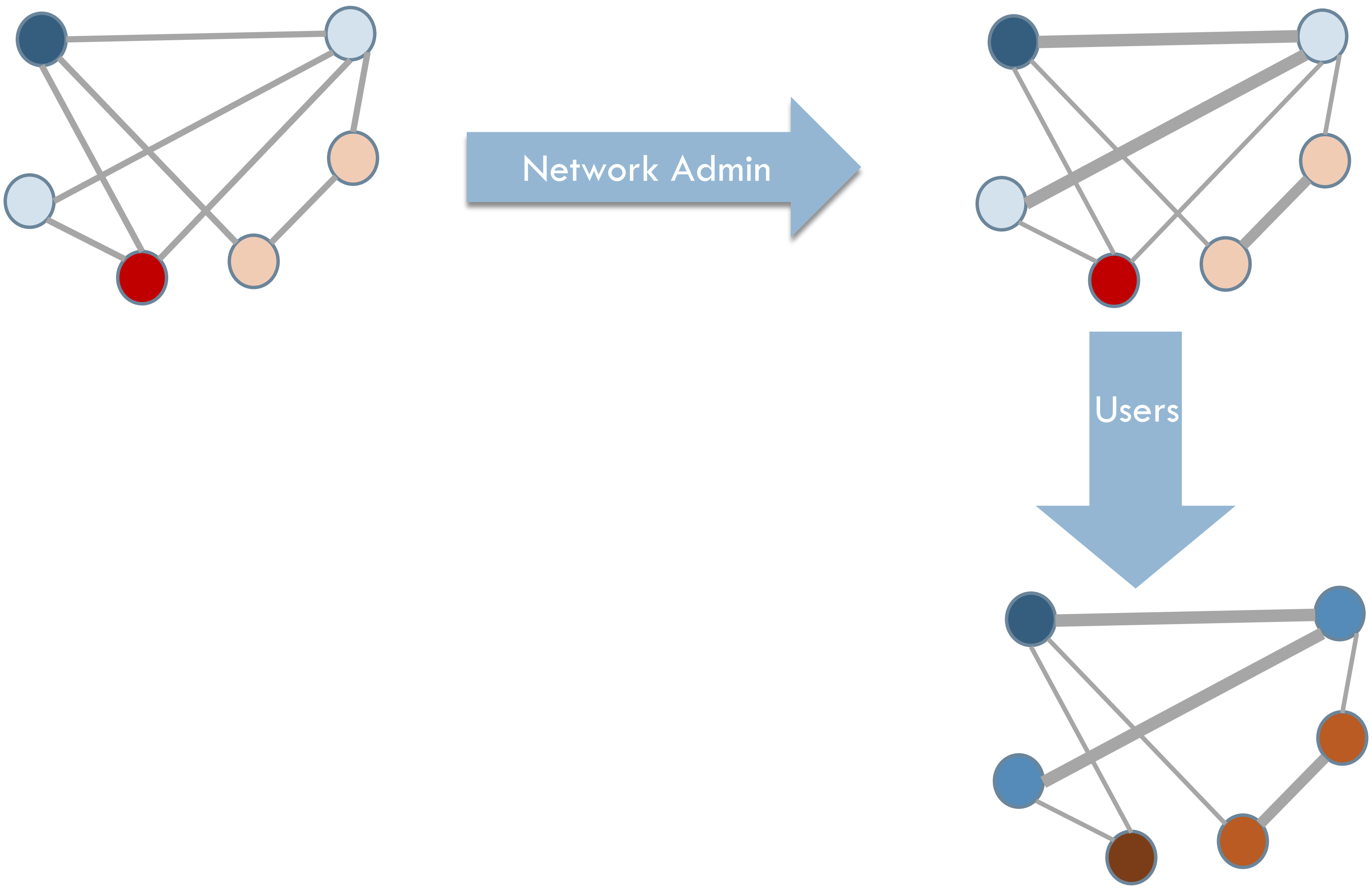
Network Administrator Dynamics



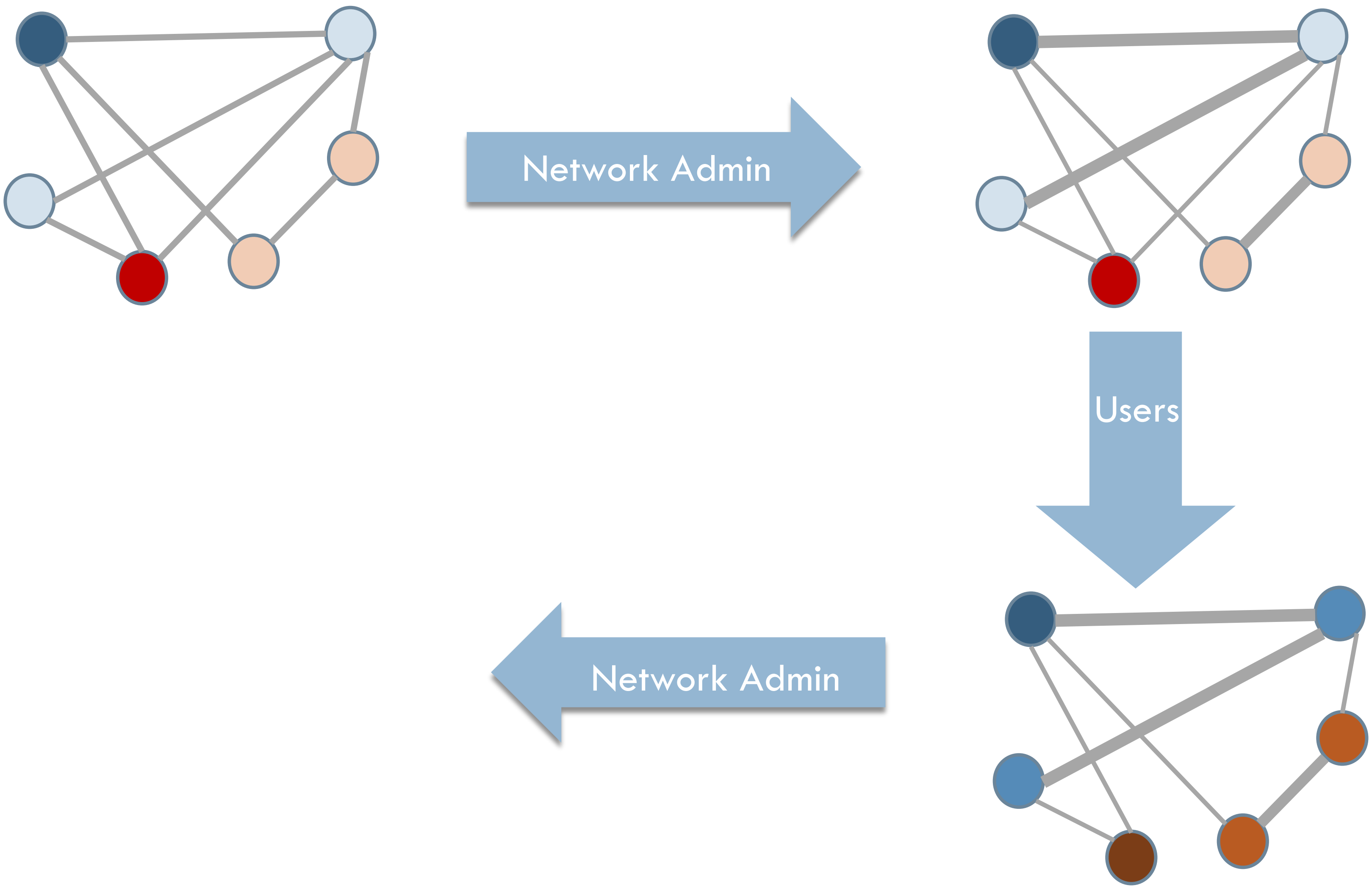
Network Administrator Dynamics



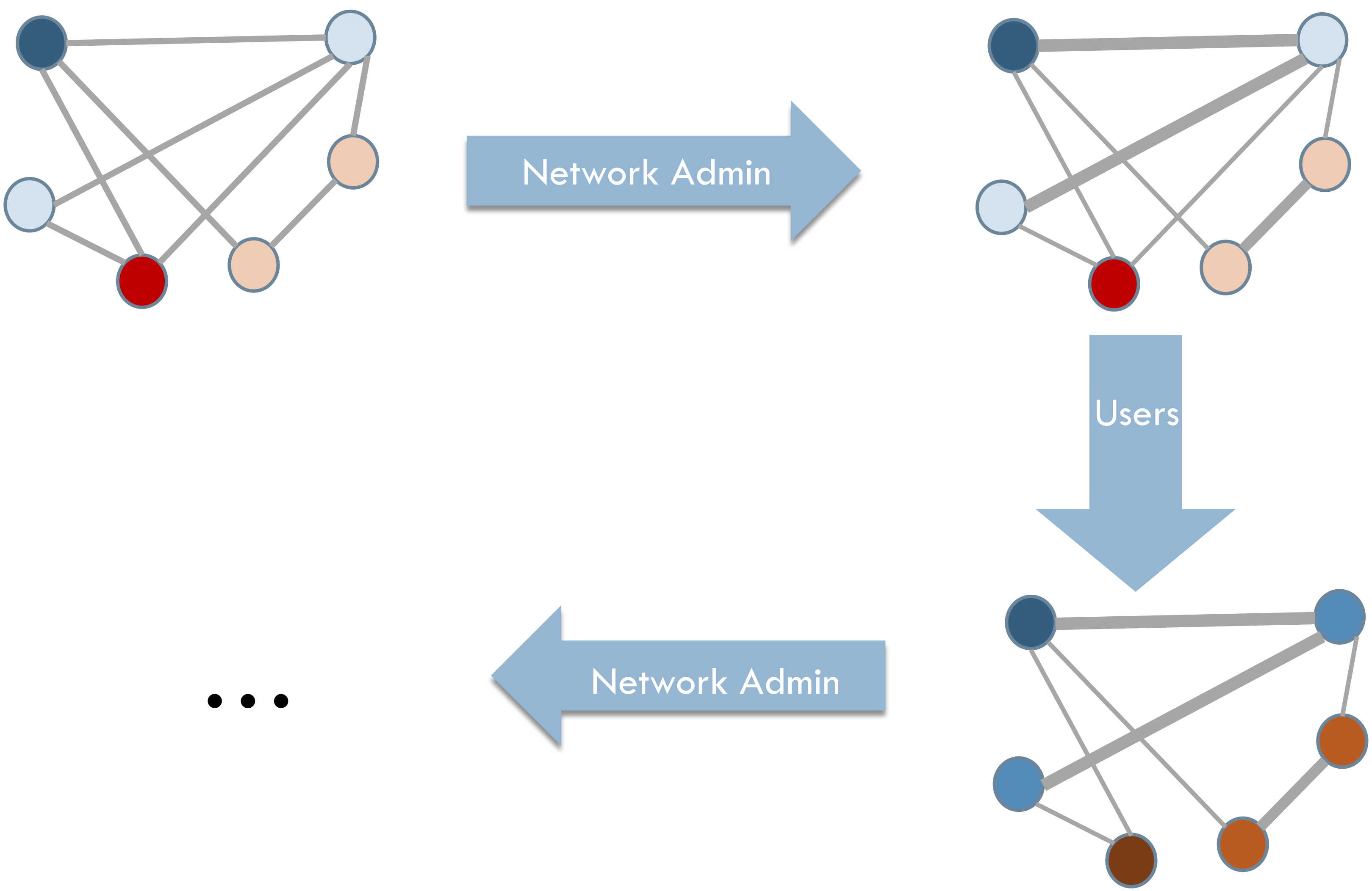
Network Administrator Dynamics



Network Administrator Dynamics



Network Administrator Dynamics



Network Administrator Dynamics

Question: If we model recommender systems in a social network, by introducing the network administrator:

- will polarization increase?
- do echo chambers form?

Outline

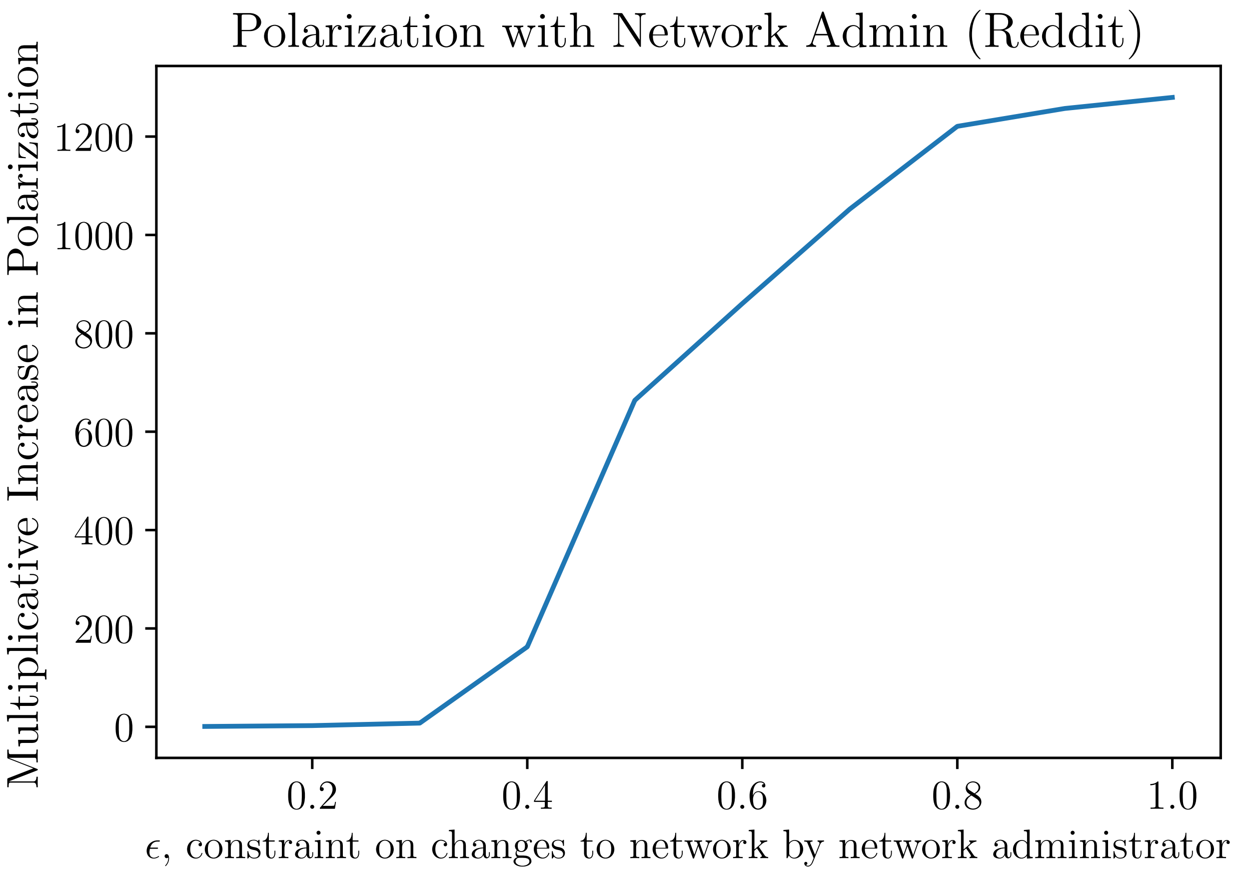
1. Friedkin-Johnsen model for opinion dynamics
2. Introducing the Network Administrator
3. Results
    1. Experiments on Reddit and Twitter networks
    2. Theoretical arguments
4. A simple remedy to reduce polarization
5. Conclusion

Experiments

We use two networks:

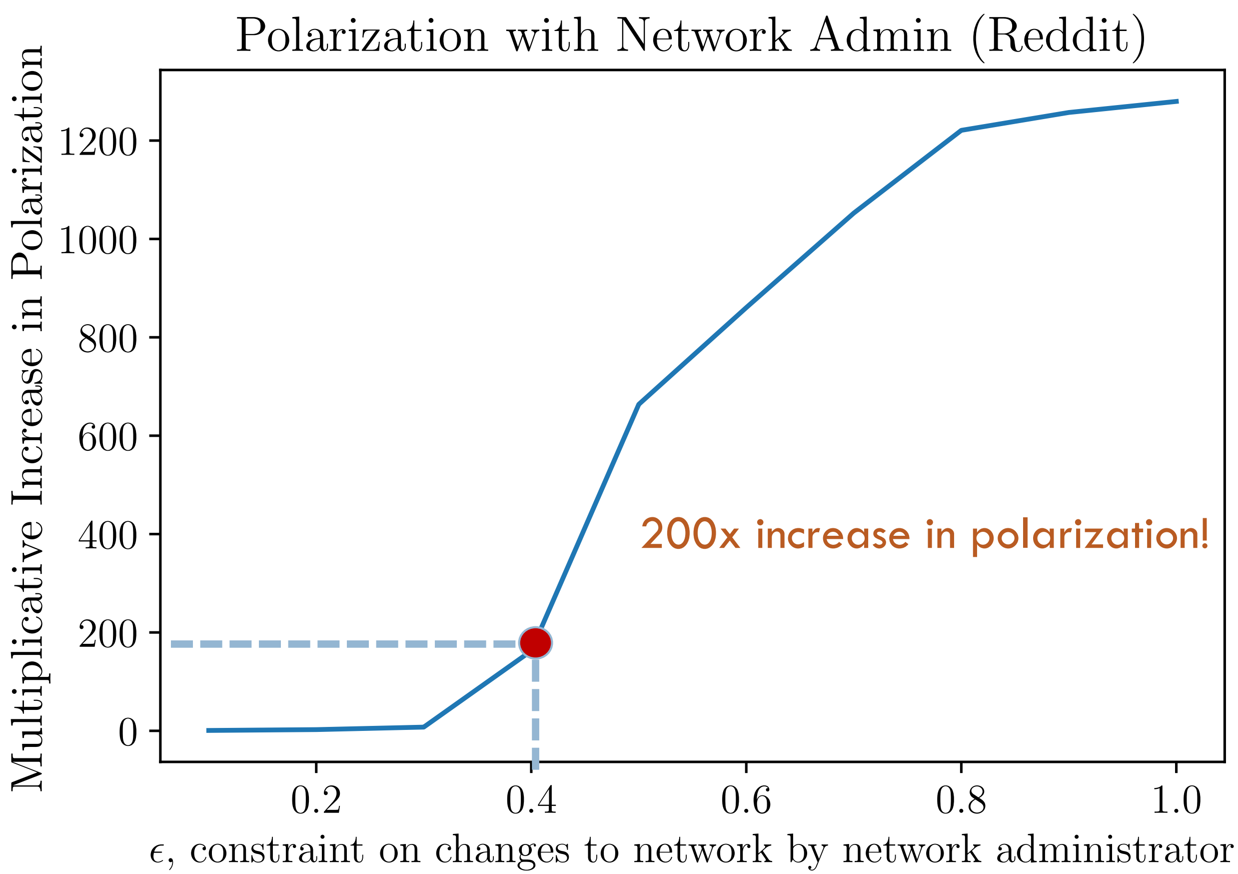
1. Twitter
   1. 548 nodes, 3638 edges
   2. Nodes = users
   3. Edges = user interactions about the Delhi legislative assembly elections of 2013.
2. Reddit
   1. 556 nodes, m = 8969 edges.
   2. Nodes = users posting in r/politics
   3. Edges = users that both posted in the same subreddit

Results

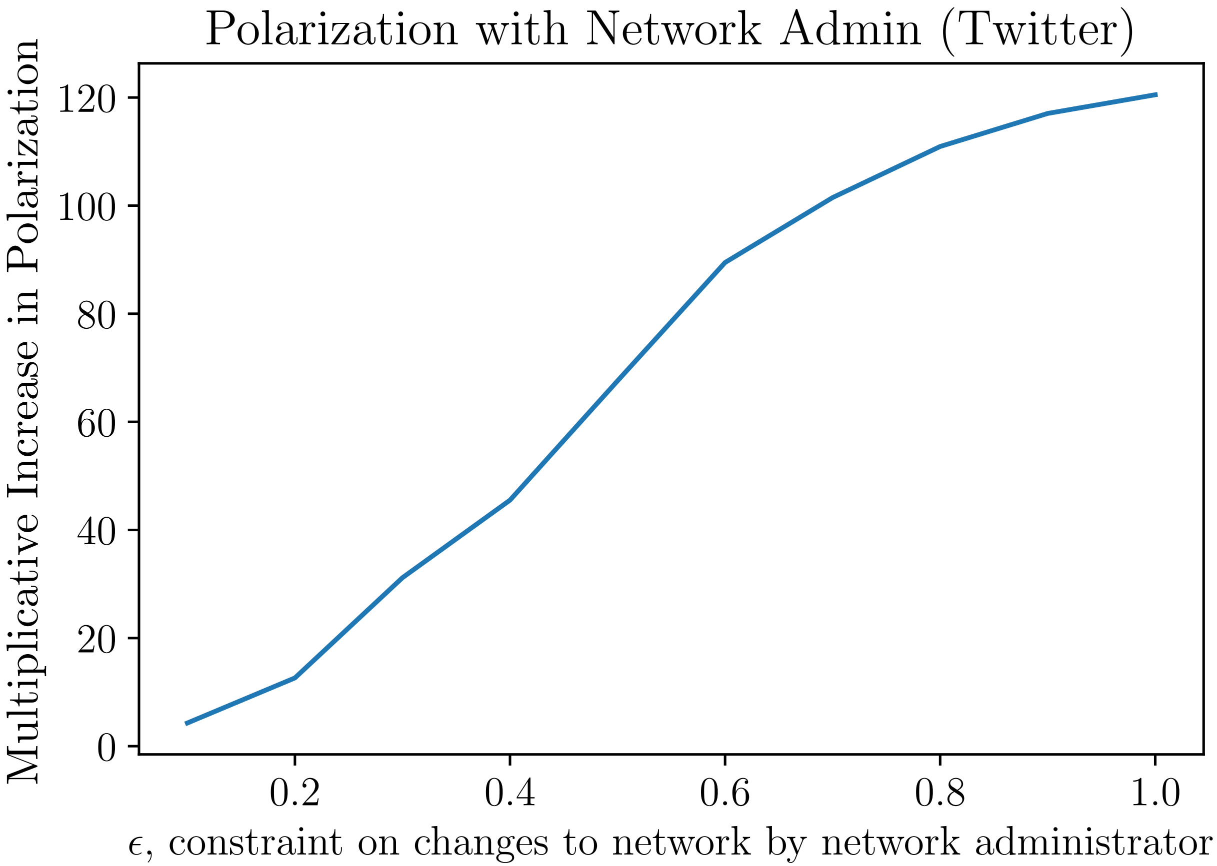


Polarization with Network Admin (Reddit)
Multiplicative Increase in Polarization
0
200
400
600
800
1000
1200
0.2
0.4
0.6
0.8
1.0
ϵ, constraint on changes to network by network administrator

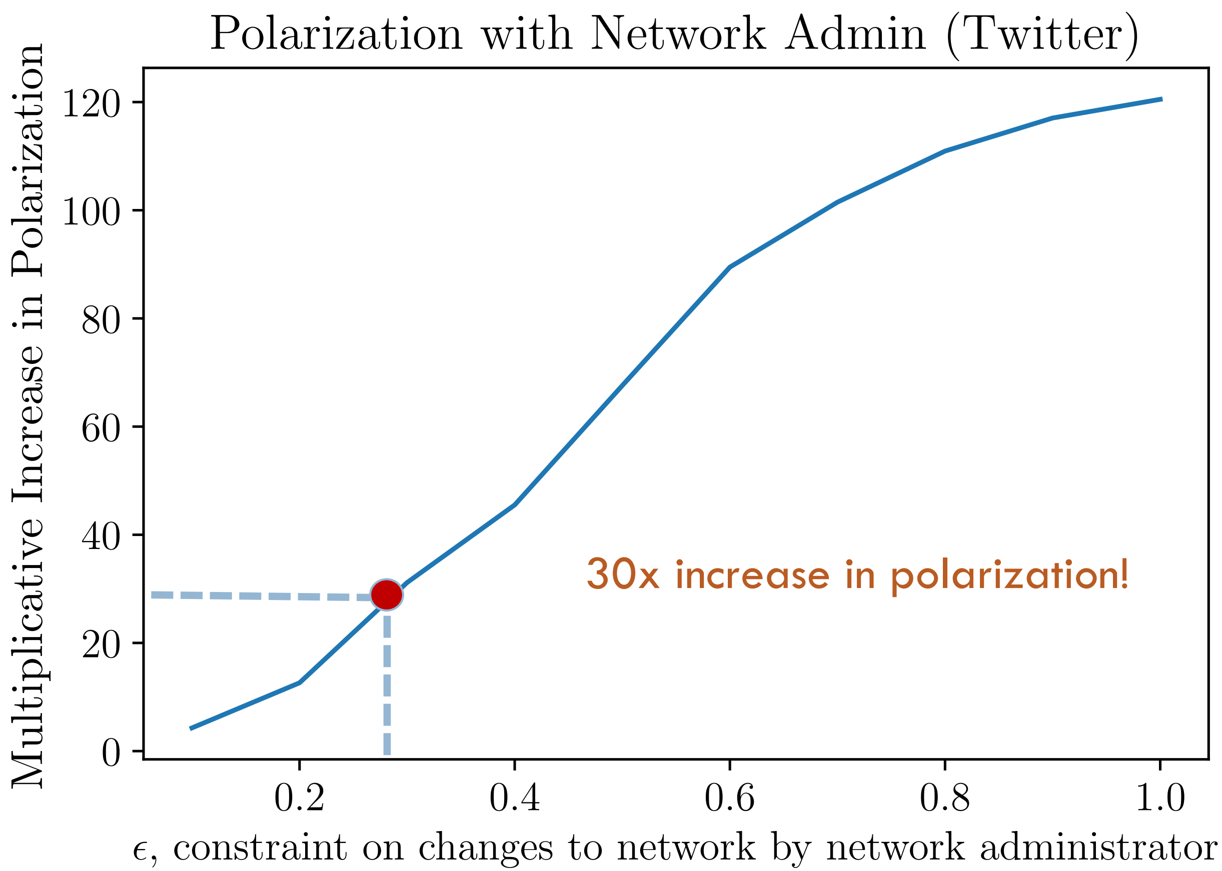
Results


Polarization with Network Admin (Reddit)
Multiplicative Increase in Polarization
0
200
400
600
800
1000
1200
0.2
0.4
0.6
0.8
1.0
ϵ, constraint on changes to network by network administrator
200x increase in polarization!

Results


Polarization with Network Admin (Twitter)
Multiplicative Increase in Polarization
120
100
80
60
40
20
0
0.2
0.4
0.6
0.8
1.0
ϵ, constraint on changes to network by network administrator

Results



Polarization with Network Admin (Twitter)
Multiplicative Increase in Polarization
120
100
80
60
40
20
0
0.2
0.4
0.6
0.8
1.0
ϵ, constraint on changes to network by network administrator
30x increase in polarization!

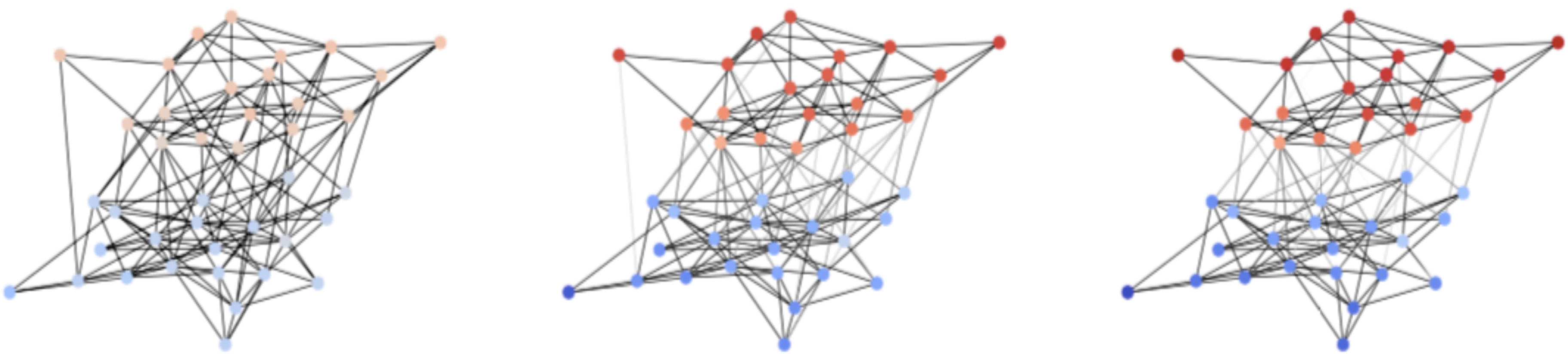
Experiments

Do echo chambers form?

Experiments

Do echo chambers form?

Apply network administrator to synthetic graph (for better visualization)



(a) Example synthetic social network graph.

(b) Graph after network administrator changes just 20% of edge weight.

(c) Graph after network administrator changes just 30% of edge weight.

Summarizing our experiments

Thus, when the network administrator filters content:

1. Polarization increases
2. Echo chambers form

Summarizing our experiments

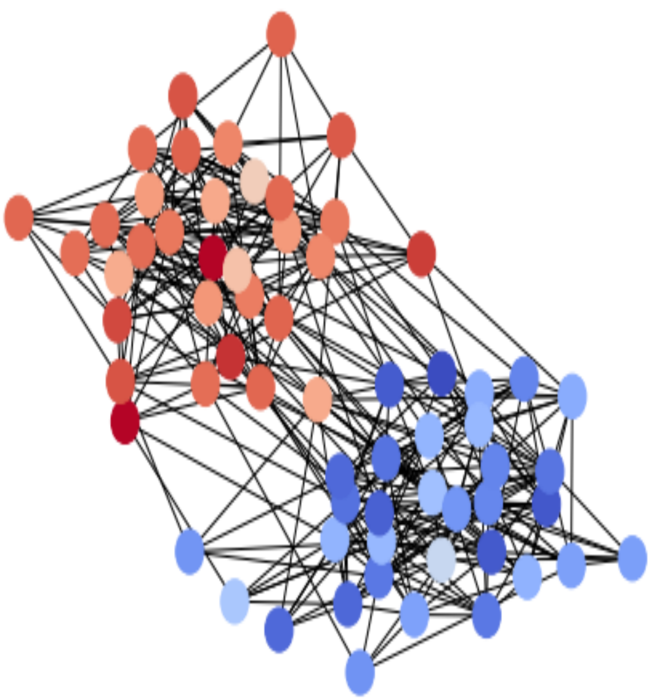
Thus, when the network administrator filters content:

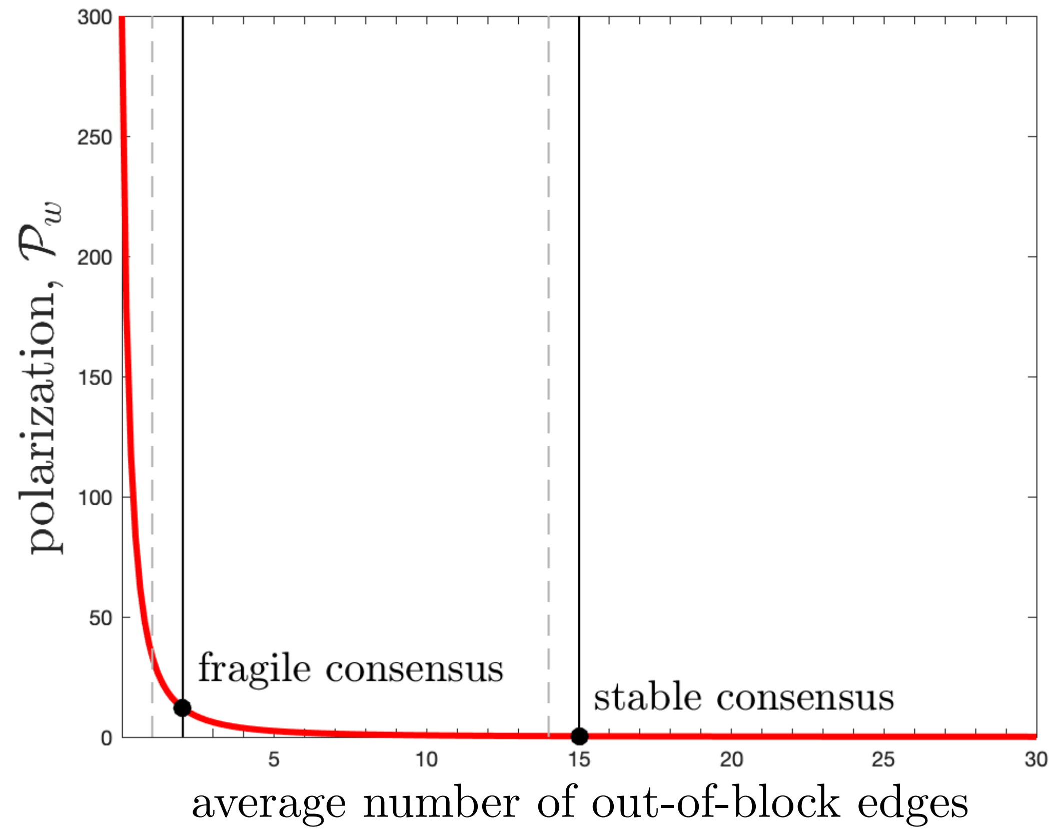
1. Polarization increases
2. Echo chambers form

Our model confirms the filter bubble phenomenon!

Theoretical Results

Theorem (informal): With 99% probability, social networks generated from stochastic block model is in a state of fragile consensus.





Outline

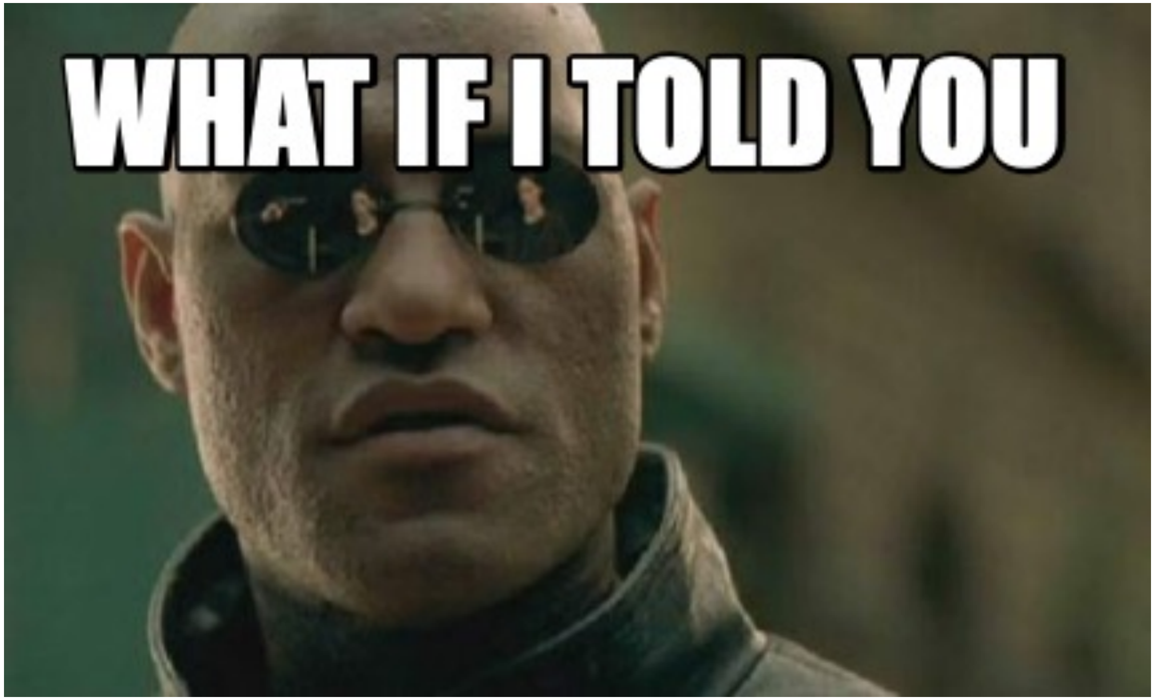
1. Friedkin-Johnsen model for opinion dynamics
2. Introducing the Network Administrator
3. Results
   1. Experiments on Reddit and Twitter networks
   2. Theoretical arguments
4. A simple remedy to reduce polarization
5. Conclusion

Current State of Affairs

Up until now, our results have been negative.

1. Algorithmic content filtering can dramatically increase polarization and form echo chambers
2. Social networks are often in a state of fragile consensus

A Simple Fix



… with one small fix, the filter bubble effect can be mitigated

A Simple Fix

Network administrator adds a regularization term to their objective

A Simple Fix

Network administrator adds a regularization term to their objective

Before

$$\min_{\text{graph } G} \mathcal{D}_{\mathbf{z}}$$

After

$$\min_{\text{graph } G} \mathcal{D}_{\mathbf{z}} + \lambda \sum_{i<j} w_{ij}^2$$

A Simple Fix

Network administrator adds a regularization term to their objective

Before

$$\min_{\text{graph } G} \mathcal{D}_{\mathbf{z}}$$

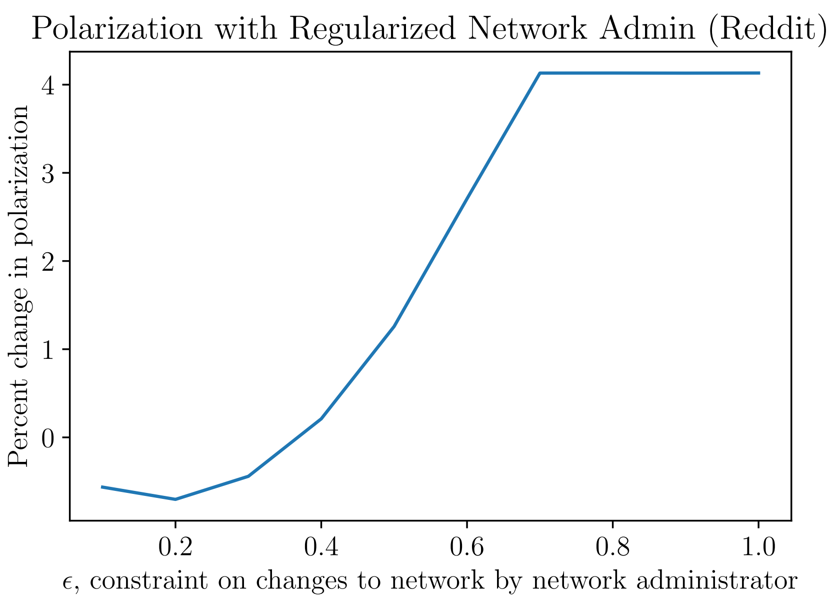
After

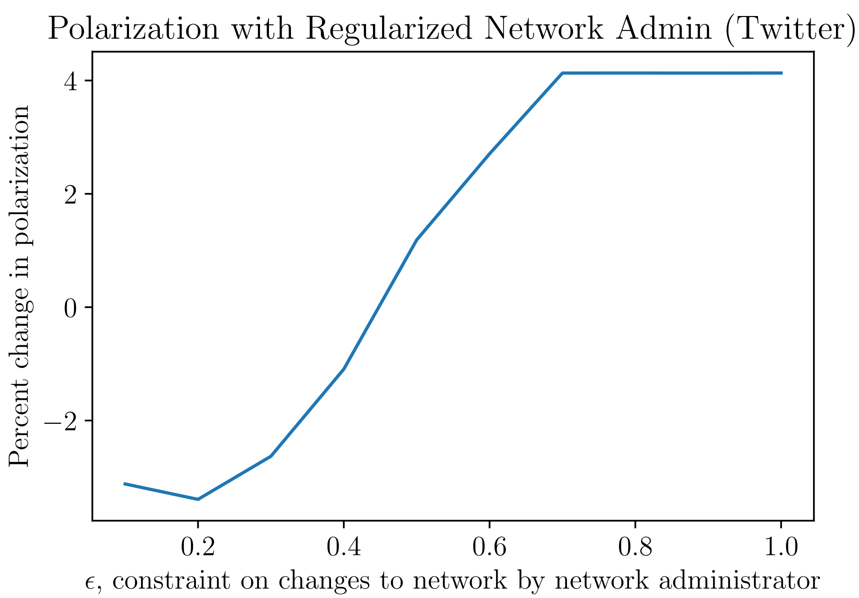
$$\min_{\text{graph } G} \mathcal{D}_{\mathbf{z}} + \lambda \sum_{i<j} w_{ij}^2$$

Intuition: Similar to FB news feed showing you a random story from a random friend

A Simple Fix

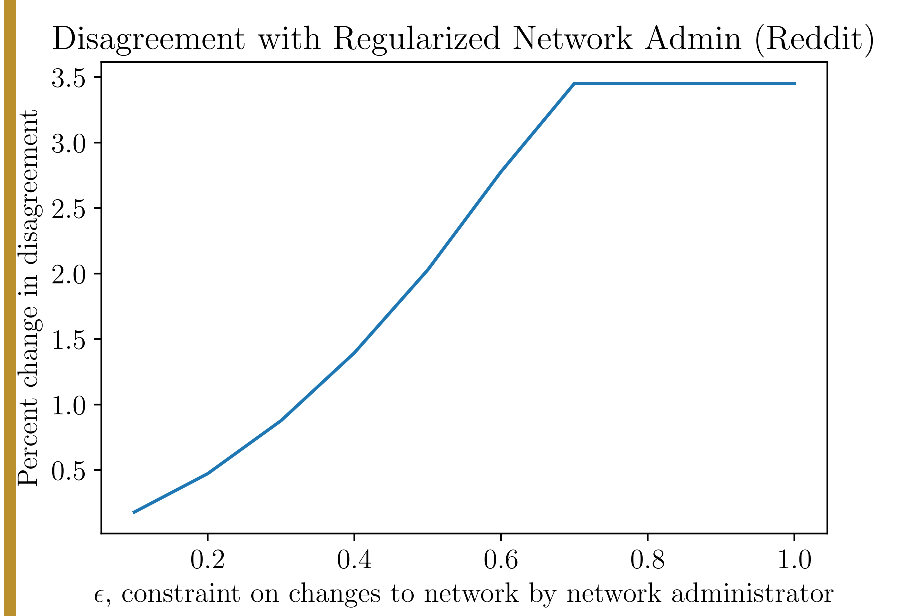
Polarization increases only 2-4% with regularization

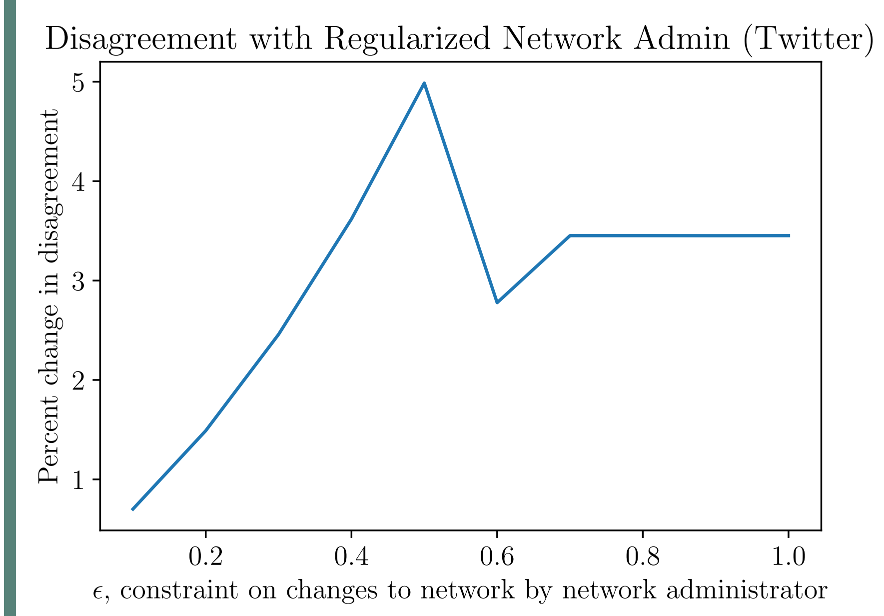




A Simple Fix

Disagreement, the objective of the network administrator, also only increases by 3-5%





The Whole Story

Network administrator maximizes metrics like engagement or ad revenue by changing structure of network

The Whole Story

Network administrator maximizes metrics like engagement or ad revenue by changing structure of network

1. Without regularization (i.e. increasing diversity of stories seen by users):

The Whole Story

Network administrator maximizes metrics like engagement or ad revenue by changing structure of network

1. Without regularization (i.e. increasing diversity of stories seen by users):
   - network administrator dramatically increases polarization,
   - network administrator forms echo chambers

The Whole Story

Network administrator maximizes metrics like engagement or ad revenue by changing structure of network

2. With regularization:

The Whole Story

Network administrator maximizes metrics like engagement or ad revenue by changing structure of network

2. With regularization:
   - Network administrator does not increase polarization
   - Network administrator only loses small % of bottom line (disagreement)

Outline

1. Friedkin-Johnsen model for opinion dynamics

2. Introducing the Network Administrator

3. Results
    1. Experiments on Reddit and Twitter networks
    2. Theoretical arguments

4. A simple remedy to reduce polarization

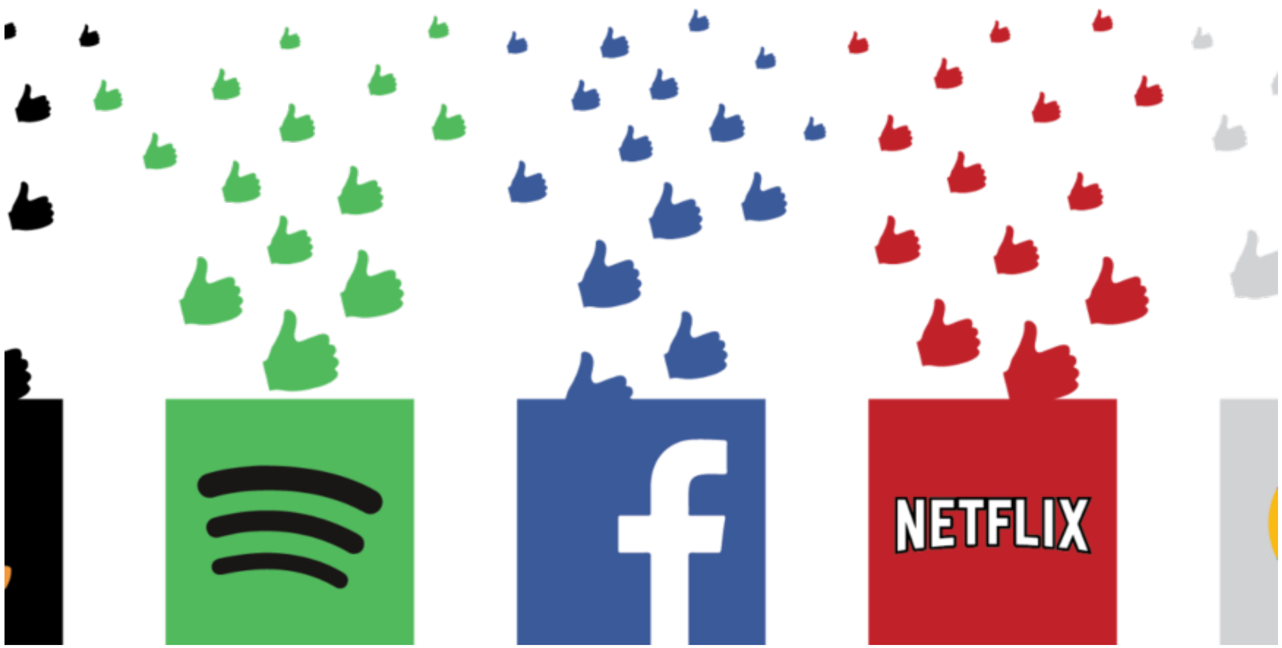
5. Conclusion

Conclusions

Summarizing our work:

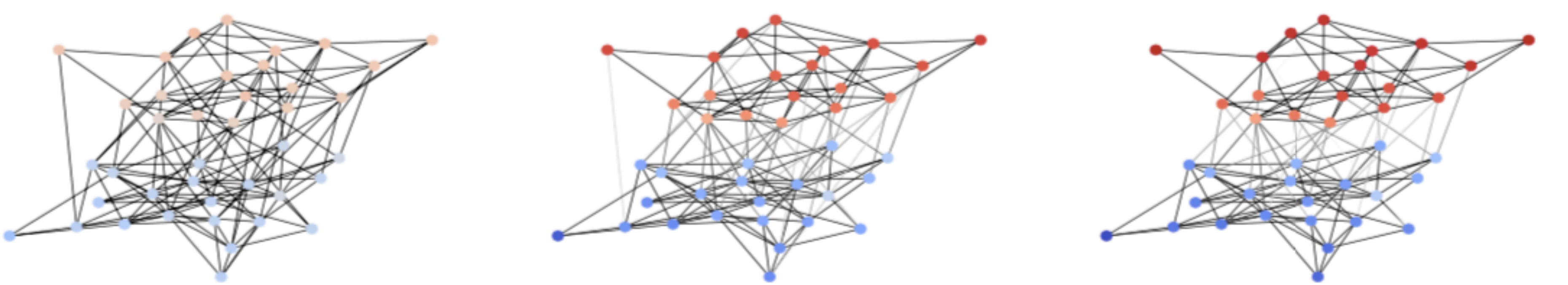
Conclusions

1. Model recommender systems in social networks by introducing a network administrator to the Friedkin-Johnsen model



Conclusions

2. Show that the filter bubble theory holds true in our model, as the network administrator will:
    1. dramatically increase polarization, and
    2. cause echo chambers to form.



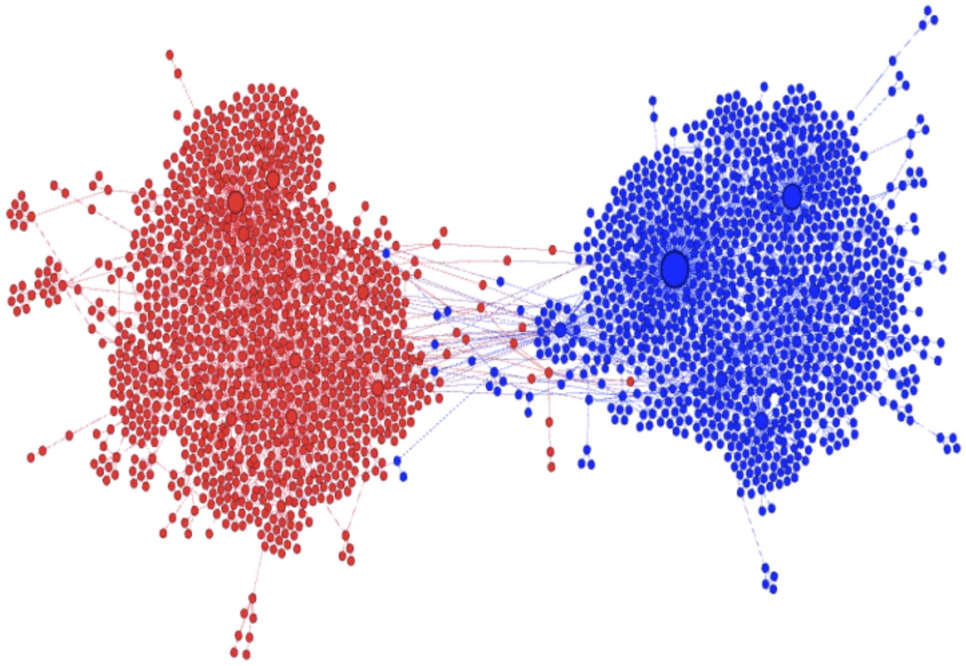
(a) Example synthetic social network graph.

(b) Graph after network administrator changes just 20% of edge weight.

(c) Graph after network administrator changes just 30% of edge weight.

Conclusions

3. When network administrator explicitly optimizes for diversity (via regularization), the filter bubble effect is mitigated



Thank You For Listening!

arXiv.org > cs > arXiv:1906.08772

Search... All fields Search

Help | Advanced Search

Computer Science > Social and Information Networks

Understanding Filter Bubbles and Polarization in Social Networks

Uthsav Chitra, Christopher Musco

(Submitted on 20 Jun 2019)

Recent studies suggest that social media usage -- while linked to an increased diversity of information and perspectives for users -- has exacerbated user polarization on many issues. A popular theory for this phenomenon centers on the concept of "filter bubbles": by automatically recommending content that a user is likely to agree with, social network algorithms create echo chambers of similarly-minded users that would not have arisen otherwise. However, while echo chambers have been observed in real-world networks, the evidence for filter bubbles is largely post-hoc.
In this work, we develop a mathematical framework to study the filter bubble theory. We modify the classic Friedkin-Johnsen opinion dynamics model by introducing another actor, the network administrator, who filters content for users by making small changes to the edge weights of a social network (for example, adjusting a news feed algorithm to change the level of interaction between users).
On real-world networks from Reddit and Twitter, we show that when the network administrator is incentivized to reduce disagreement among users, even relatively small edge changes can result in the formation of echo chambers in the network and increase user polarization. We theoretically support this observed sensitivity of social networks to outside intervention by analyzing synthetic graphs generated from the stochastic block model. Finally, we show that a slight modification to the incentives of the network administrator can mitigate the filter bubble effect while minimally affecting the administrator's target objective, user disagreement.

Download:
- PDF
- Other formats

(license)

Current browse context:
cs.SI
< prev | next >
new | recent | 1906

Change to browse by:
cs
physics
physics.soc-ph

References & Citations
- NASA ADS

DBLP - CS Bibliography
listing | bibtex
Uthsav Chitra
Christopher Musco

Full workshop paper on arXiv: https://arxiv.org/abs/1906.08772

Questions?