# Opinion formation in social networks: models and computational problems

## Lecture in course “Social networks and online markets”

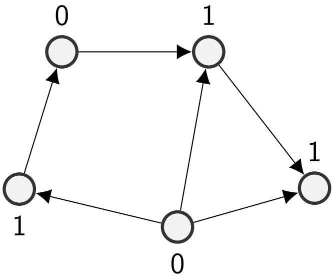
Sapienza, Wednesday, April 3, 2024

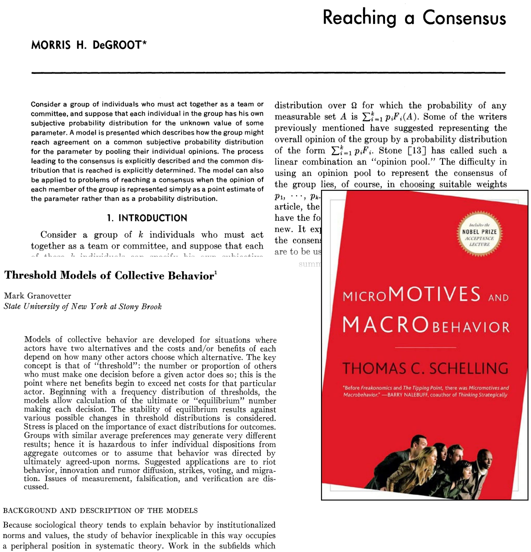
*Aristides Gionis, KTH Royal Institute of Technology, Sweden*

argioni@kth.se

# opinion formation in the scientific literature

- classic works in social science, economics...
- simplistic models of agent interactions



Reaching a Consensus

MORRIS H. DeGROOT*

Consider a group of individuals who must act together as a team or committee, and suppose that each individual in the group has his own subjective probability distribution for the unknown value of some parameter. A model is presented which describes how the group might reach agreement on a common subjective probability distribution for the parameter by pooling their individual opinions. The process leading to the consensus is explicitly described and the common distribution that is reached is explicitly determined. The model can also be applied to problems of reaching a consensus when the opinion of each member of the group is represented simply as a point estimate of the parameter rather than as a probability distribution.

1. INTRODUCTION

Consider a group of $k$ individuals who must act together as a team or committee, and suppose that each

distribution over $\Omega$ for which the probability of any measurable set $A$ is $\sum_{i=1}^{k} p_i F_i(A)$. Some of the writers previously mentioned have suggested representing the overall opinion of the group by a probability distribution of the form $\sum_{i=1}^{k} p_i F_i$. Stone [13] has called such a linear combination an "opinion pool." The difficulty in using an opinion pool to represent the consensus of the group lies, of course, in choosing suitable weights $p_1, \cdots, p_k$ ... article, the ... have the fo... new. It ex... the consens... are to be us...

Threshold Models of Collective Behavior[1]

Mark Granovetter
*State University of New York at Stony Brook*

Models of collective behavior are developed for situations where actors have two alternatives and the costs and/or benefits of each depend on how many other actors choose which alternative. The key concept is that of "threshold": the number or proportion of others who must make one decision before a given actor does so; this is the point where net benefits begin to exceed net costs for that particular actor. Beginning with a frequency distribution of thresholds, the models allow calculation of the ultimate or "equilibrium" number making each decision. The stability of equilibrium results against various possible changes in threshold distributions is considered. Stress is placed on the importance of exact distributions for outcomes. Groups with similar average preferences may generate very different results; hence it is hazardous to infer individual dispositions from aggregate outcomes or to assume that behavior was directed by ultimately agreed-upon norms. Suggested applications are to riot behavior, innovation and rumor diffusion, strikes, voting, and migration. Issues of measurement, falsification, and verification are discussed.

BACKGROUND AND DESCRIPTION OF THE MODELS

Because sociological theory tends to explain behavior by institutionalized norms and values, the study of behavior inexplicable in this way occupies a peripheral position in systematic theory. Work in the subfields which

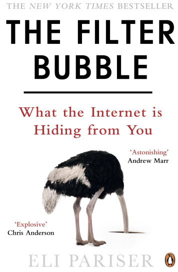
# opinion formation research today

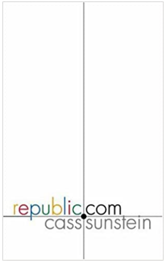
- renewed interest in opinion formation in various scientific domains including computer science

why?

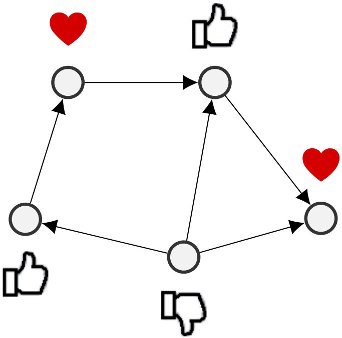
- availability of large-scale social network data
- applications:
  - recommender systems,
  - viral marketing,
  - political campaigning...
- emerging social concerns:
  - political polarization,
  - teenage mental health...

facebook twitter



# social interactions today



# overview

- the DeGroot and Friedkin–Johnsen (FJ) models (consensus)
  - definition of the DeGroot and Friedkin-Johnsen (FJ) models
  - properties
- other opinion formation models (disagreement & polarization)
  - biased assimilation and bounded confidence
  - geometric models
- algorithmic interventions for moderating opinions
  - polarization and disagreement indices
  - efficiently estimating user opinions and indices
  - maximizing opinions / minimizing polarization and disagreement
  - emergence of echo chambers

# the DeGroot and Friedkin–Johnsen (FJ) models

# models of opinion formation

- individuals' opinions are influenced by their peers
- how to model the opinion-formation process in a social network?
- one way is to model influence as information cascades
  - a discrete entity (action, meme, virus) propagates in a network
  - cascade is modeled using the independent-cascade model
- opinion-formation models follow a continuous weighted-averaging process

# opinion formation by weighted averaging

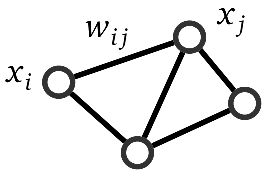
- at each time step, each individual updates their opinion as a weighted average of the opinions of their neighbors
- the process continues until convergence

# models of opinion formation

a basic model [DeGroot, 1974].

- we consider a weighted graph modeling a social network
- weight $w_{ij}$ represents influence of node $j$ on $i$ ($i$ trusts $j$)
- at time $t$, node $i$ has opinion $x_i^{(t)}$, initally $x_i^{(0)} \in [0, 1]$
- node $i$ updates their opinion by

$$x_i^{(t+1)} = \frac{\sum_{j \mid (i,j) \in E} w_{ij} x_j^{(t)}}{\sum_{j \mid (i,j) \in E} w_{ij}}$$

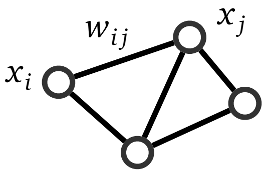


# models of opinion formation

a basic model [DeGroot, 1974].

- we consider a weighted graph modeling a social network
- weight $w_{ij}$ represents influence of node $j$ on $i$ ($i$ trusts $j$)
- at time $t$, node $i$ has opinion $x_i^{(t)}$, initally $x_i^{(0)} \in [0,1]$
- node $i$ updates their opinion by

$$x_i^{(t+1)} = \frac{\sum_{j \mid (i,j) \in E} w_{ij} x_j^{(t)}}{\sum_{j \mid (i,j) \in E} w_{ij}}$$



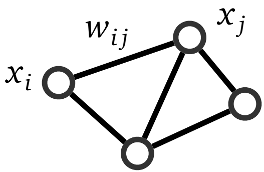

what do you expect to happen?

# models of opinion formation

a basic model [DeGroot, 1974].

- we consider a weighted graph modeling a social network
- weight $w_{ij}$ represents influence of node $j$ on $i$ ($i$ trusts $j$)
- at time $t$, node $i$ has opinion $x_i^{(t)}$, initally $x_i^{(0)} \in [0, 1]$
- node $i$ updates their opinion by

$$x_i^{(t+1)} = \frac{\sum_{j \,|\, (i,j) \in E} w_{ij} x_j^{(t)}}{\sum_{j \,|\, (i,j) \in E} w_{ij}}$$



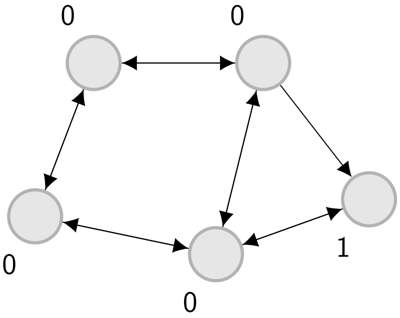

what do you expect to happen?

- under certain conditions all nodes converge to having the same opinion

# the properties of the DeGroot model

DeGroot example

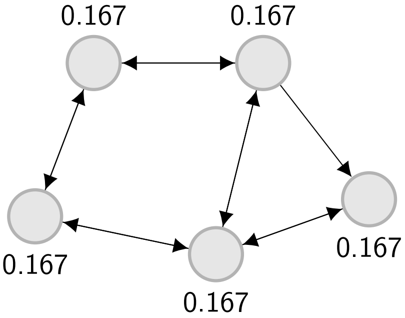
$$x_i^{(t+1)} = \frac{\sum_{j\,|\,(i,j)\in E} w_{ij} x_j^{(t)}}{\sum_{j\,|\,(i,j)\in E} w_{ij}}$$



# the properties of the DeGroot model

DeGroot example

$$x_i^{(t+1)} = \frac{\sum_{j\,|\,(i,j)\in E} w_{ij} x_j^{(t)}}{\sum_{j\,|\,(i,j)\in E} w_{ij}}$$



# the properties of the DeGroot model

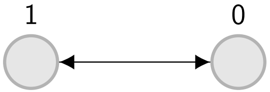
do the opinions of all nodes contribute to the final (common) opinion?

# the properties of the DeGroot model

do the opinions of all nodes contribute to the final (common) opinion?

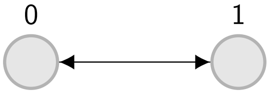
# the properties of the DeGroot model

furthermore, convergence is not guaranteed!



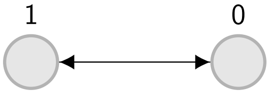
# the properties of the DeGroot model

furthermore, convergence is not guaranteed!



# the properties of the DeGroot model

furthermore, convergence is not guaranteed!




some intuition on convergence:

recall: node $i$ updates their opinion by $x_i^{(t+1)} = \sum_{j \mid (i,j) \in E} w_{ij} x_j^{(t)}$ where $\sum_j w_{ij} = 1$

define matrix $W$ so that $W_{ij} = w_{ij}$; then

- $\mathbf{x}^{(t+1)} = W\mathbf{x}^{(t)}$
- $W$ is row stochastic

# the properties of the DeGroot model

## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

a graph is aperiodic if the maximum common divisor of the length of its cycles is 1

# the properties of the DeGroot model

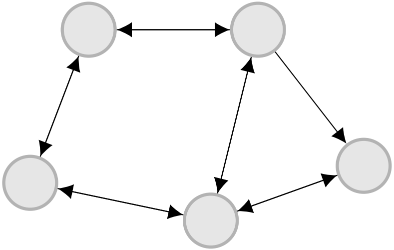
## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

a graph is aperiodic if the maximum common divisor of the length of its cycles is 1

our graph from before:



# the properties of the DeGroot model

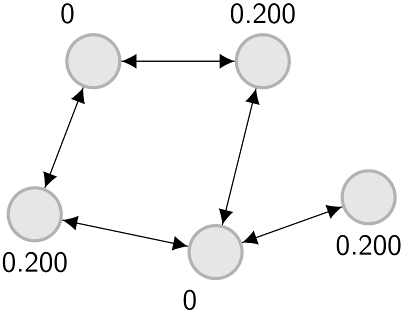
## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

a graph is aperiodic if the maximum common divisor of the length of its cycles is 1

our graph from before:



# the properties of the DeGroot model

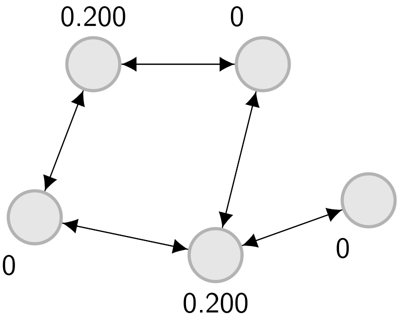
## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

a graph is aperiodic if the maximum common divisor of the length of its cycles is 1

our graph from before:



# the properties of the DeGroot model

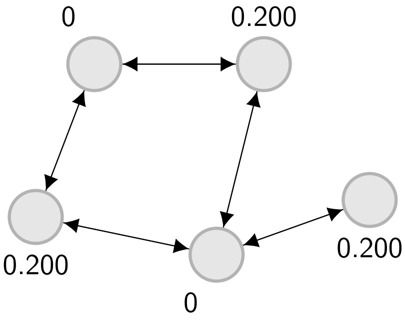
## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

a graph is aperiodic if the maximum common divisor of the length of its cycles is 1

our graph from before:



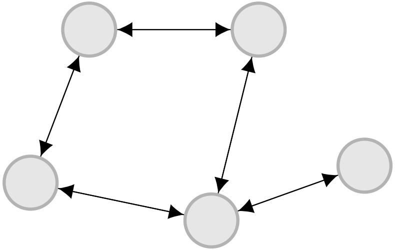
# the properties of the DeGroot model

## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

it is easy to fix oscillations: a loop will do



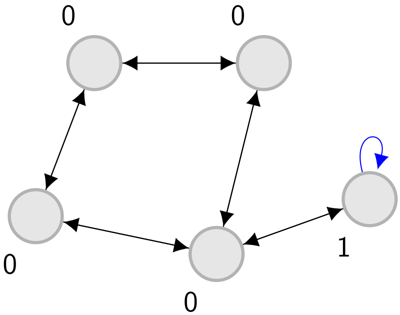
# the properties of the DeGroot model

## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

it is easy to fix oscillations: a loop will do



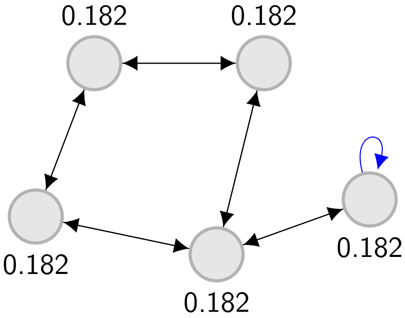
# the properties of the DeGroot model

## convergence

**lemma**

let $G$ be strongly connected; then the DeGroot process converges if and only if $G$ is aperiodic

it is easy to fix oscillations: a loop will do



# the properties of the DeGroot model

## convergence

let $G$ be convergent; what is the consensus value? [Golub and Jackson, 2010]

suppose there is a vector $\mathbf{v}$ of agent influence, i.e.,

$$\left(\lim_{t\to\infty} W^t \mathbf{x}^{(0)}\right)_j = \mathbf{v}^T \mathbf{x}^{(0)} \quad \text{for all } j$$

since $\lim_{t\to\infty} W^t \mathbf{x}^{(0)} = \lim_{t\to\infty} W^t (W\mathbf{x}^{(0)})$,

then $\mathbf{v}^T W \mathbf{x}^{(0)} = \mathbf{v}^T \mathbf{x}^{(0)}$ and so $\mathbf{v}^T W = \mathbf{v}^T$ (under mild assumptions)

in other words, the consensus opinion is $\mathbf{v}^T \mathbf{x}^{(0)}$,

where $\mathbf{v}$ is a left-eigenvector of $W$ with eigenvalue $1$

# the Friedkin-Johnsen model

[Friedkin and Johnsen, 1990]

general model of opinion formation

- $\mathbf{z}^{(1)} = X^{(1)}\mathbf{s}^{(1)}$
- $\mathbf{z}^{(t+1)} = \alpha^{(t)} W\mathbf{z}^{(t)} + \beta^{(t)} X^{(t)}\mathbf{s}^{(t)}$

common setting:

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

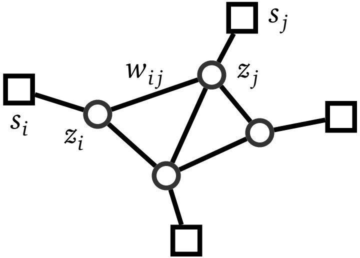
  but now $\|W\| < 1$

thus, node $i$ updates its expressed opinion by

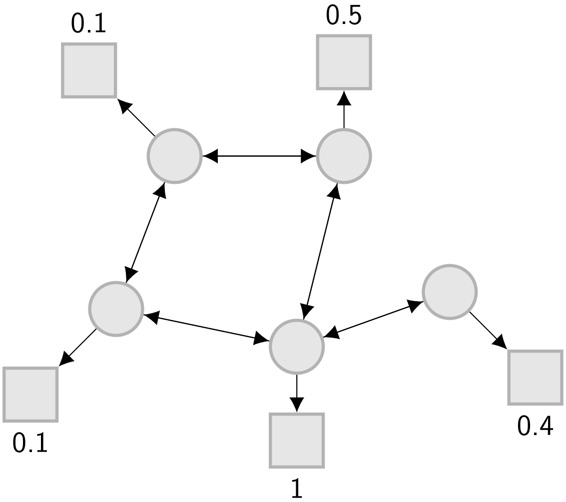
$$z_i^{(t+1)} = \frac{s_i + \sum_{j\,|\,(i,j)\in E} w_{ij} z_j^{(t)}}{1 + \sum_{j\,|\,(i,j)\in E} w_{ij}}$$

$\mathbf{s}$ stays fixed: innate opinions

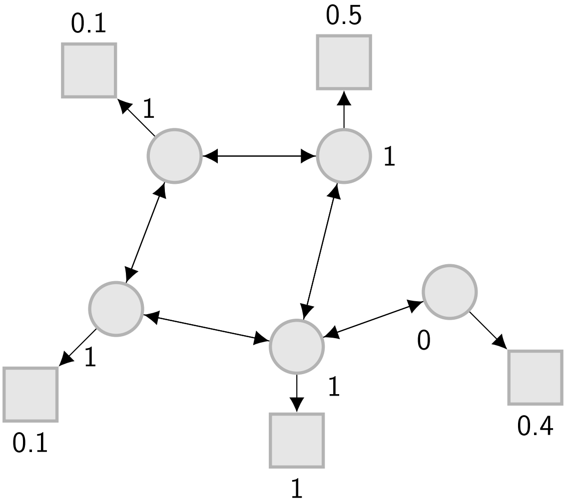
$\mathbf{z}$ changes: expressed opinion



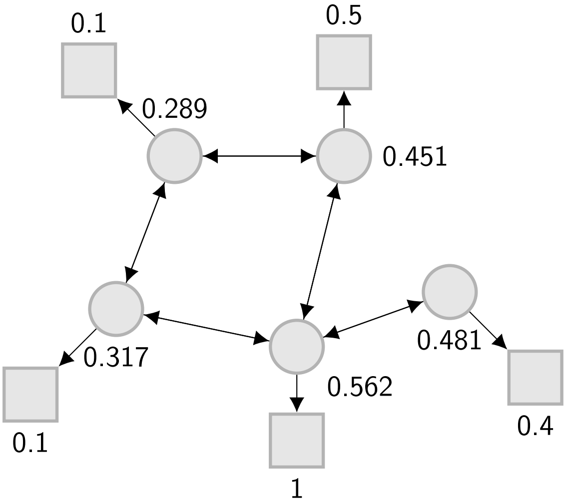
# the Friedkin-Johnsen model



# the Friedkin-Johnsen model



# the Friedkin-Johnsen model



# the Friedkin-Johnsen model

what does this variant of FJ converge to? [Friedkin and Johnsen, 1990]

recall

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

# the Friedkin-Johnsen model

what does this variant of FJ converge to? [Friedkin and Johnsen, 1990]

recall

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

so

- $\mathbf{z}^{(1)} = W\mathbf{z}^{(0)} + \mathbf{s}$

# the Friedkin-Johnsen model

what does this variant of FJ converge to? [Friedkin and Johnsen, 1990]

recall

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

so

- $\mathbf{z}^{(1)} = W\mathbf{z}^{(0)} + \mathbf{s}$
- $\mathbf{z}^{(2)} = W\mathbf{z}^{(2)} + \mathbf{s} = W(W\mathbf{z}^{(0)} + \mathbf{s}) + \mathbf{s} = W^2\mathbf{z}^{(0)} + W\mathbf{s} + \mathbf{s} = W^2\mathbf{z}^{(0)} + (W + I)\mathbf{s}$

# the Friedkin-Johnsen model

what does this variant of FJ converge to? [Friedkin and Johnsen, 1990]

recall

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

so

- $\mathbf{z}^{(1)} = W\mathbf{z}^{(0)} + \mathbf{s}$
- $\mathbf{z}^{(2)} = W\mathbf{z}^{(2)} + \mathbf{s} = W(W\mathbf{z}^{(0)} + \mathbf{s}) + \mathbf{s} = W^2\mathbf{z}^{(0)} + W\mathbf{s} + \mathbf{s} = W^2\mathbf{z}^{(0)} + (W + I)\mathbf{s}$
- $\mathbf{z}^{(3)} = W^3\mathbf{z}^{(0)} + (W^2 + W + I)\mathbf{s}$

# the Friedkin-Johnsen model

what does this variant of FJ converge to? [Friedkin and Johnsen, 1990]

recall

- $\mathbf{z}^{(t+1)} = W\mathbf{z}^{(t)} + \mathbf{s}$

so

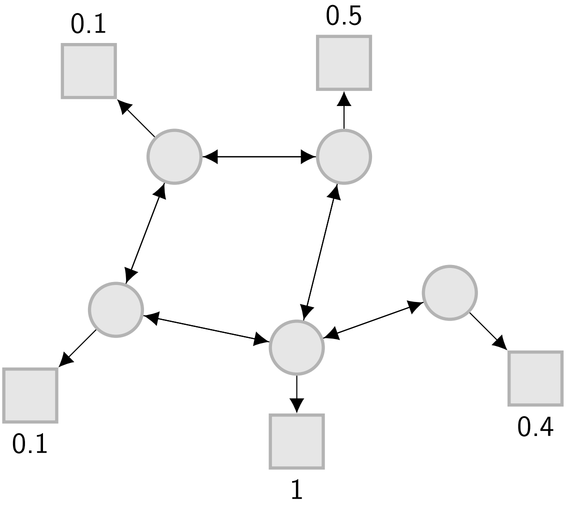
- $\mathbf{z}^{(1)} = W\mathbf{z}^{(0)} + \mathbf{s}$
- $\mathbf{z}^{(2)} = W\mathbf{z}^{(2)} + \mathbf{s} = W(W\mathbf{z}^{(0)} + \mathbf{s}) + \mathbf{s} = W^2\mathbf{z}^{(0)} + W\mathbf{s} + \mathbf{s} = W^2\mathbf{z}^{(0)} + (W + I)\mathbf{s}$
- $\mathbf{z}^{(3)} = W^3\mathbf{z}^{(0)} + (W^2 + W + I)\mathbf{s}$

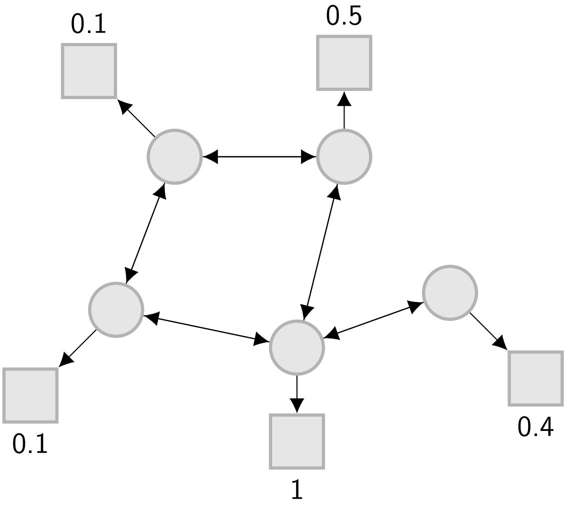
therefore,

$$\mathbf{z}^{(t+1)} = W^t\mathbf{z} + (W^{t-1} + W^{t-2} + \cdots + W^2 + W + I)\mathbf{s}$$

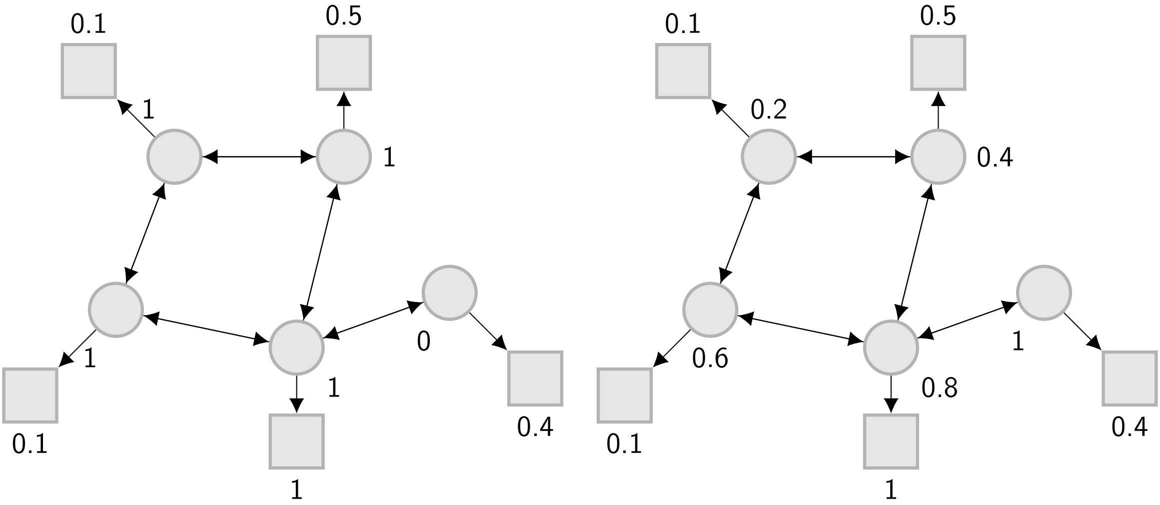
since $\|W\| < 1$, $\mathbf{z}^t \overset{t\to\infty}{\longrightarrow} (I - W)^{-1}\mathbf{s}$

# the Friedkin-Johnsen model

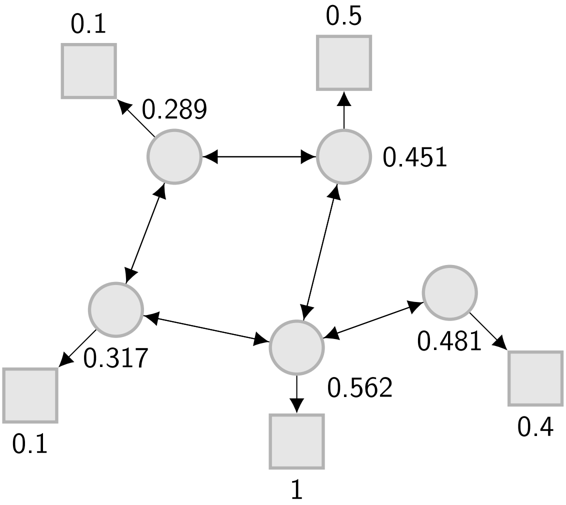


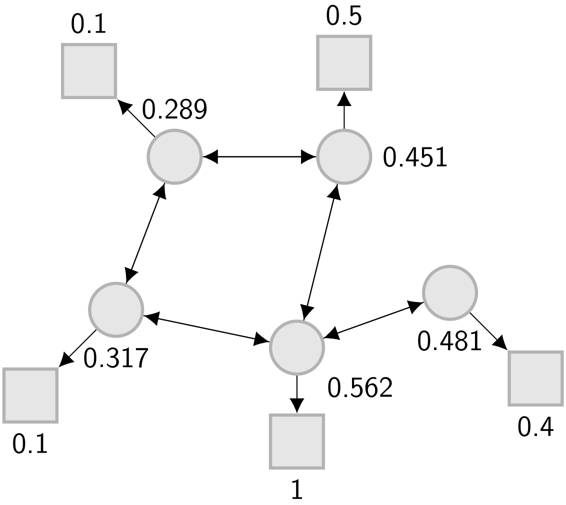



# the Friedkin-Johnsen model



# the Friedkin-Johnsen model



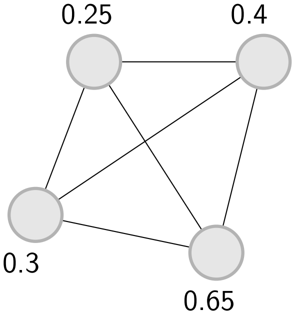
# other opinion formation models

# the bounded-confidence model

[Deffuant et al., 2000, Krause, 2000]

- individuals only interact and update their opinions if the difference between their existing opinions is smaller than a threshold $\epsilon$
- this threshold models "openness to discussion"
- larger $\epsilon$ produce consensus, while smaller $\epsilon$ produce polarized opinions

- the model can be thought as a form of selective exposure
- result: for certain values of $\epsilon$ the bounded-confidence model can lead to polarization



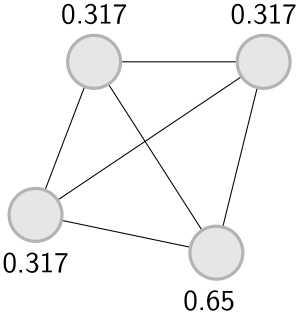

$\epsilon = 0.2$

# the bounded-confidence model

[Deffuant et al., 2000, Krause, 2000]

- individuals only interact and update their opinions if the difference between their existing opinions is smaller than a threshold $\epsilon$
- this threshold models "openness to discussion"
- larger $\epsilon$ produce consensus, while smaller $\epsilon$ produce polarized opinions

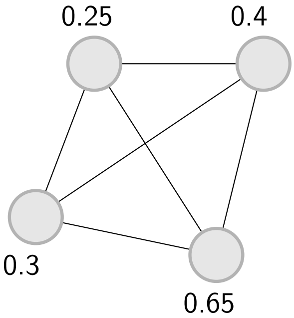
- the model can be thought as a form of selective exposure
- result: for certain values of $\epsilon$ the bounded-confidence model can lead to polarization



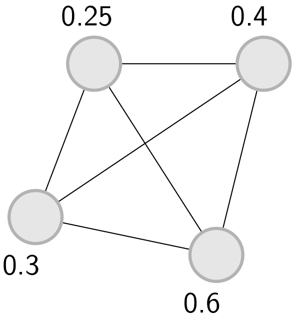

$\epsilon = 0.2$

# the bounded-confidence model

- when does this model reach consensus? [Krause, 2000]
- define $I(i, x) = \{j : |x_i - x_j| \leq \epsilon\}$.
- sufficient cond. for consensus: $I(i, x(t)) \cap I(j, x(t)) \neq \emptyset$ for all $i, j,$ all $t \geq t_0$ for some $t_0$

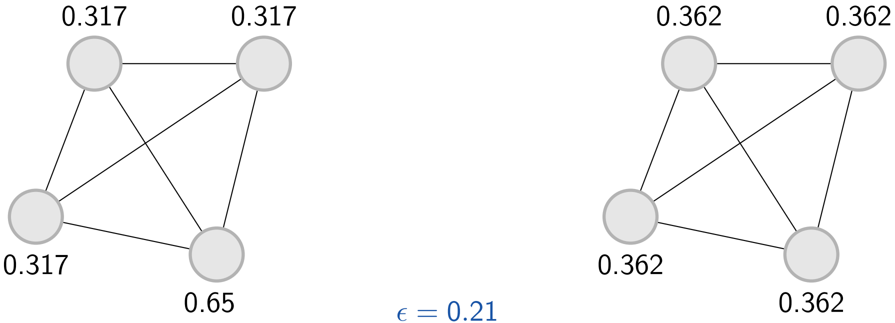



$\epsilon = 0.21$



# the bounded-confidence model

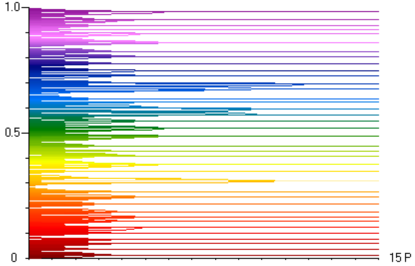
- when does this model reach consensus? [Krause, 2000]
- define $I(i, x) = \{j : |x_i - x_j| \leq \epsilon\}$.
- sufficient cond. for consensus: $I(i, x(t)) \cap I(j, x(t)) \neq \emptyset$ for all $i, j$, all $t \geq t_0$ for some $t_0$



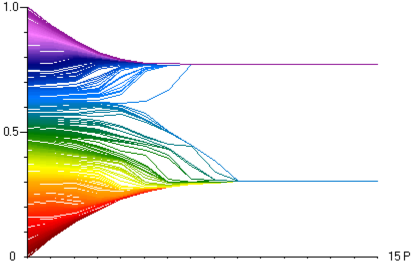

$\epsilon = 0.21$

# the bounded-confidence model — simulation results

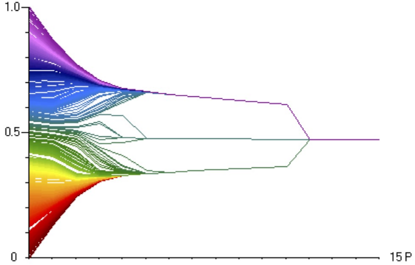
[Hegselmann et al., 2002]




ε=0.02




ε=0.3




ε=0.5

# the biased-assimilation model

[Lord et al., 1979]

biased assimilation:

*people who hold strong opinions on complex social issues are likely to examine relevant empirical evidence in a biased manner. they are apt to accept "confirming" evidence at face value while subjecting "dis-confirming" evidence to critical evaluation, and as a result to draw undue support for their initial positions from mixed or random empirical findings.*

# the biased-assimilation model

[Dandekar et al., 2013]

- modify degroot's model to explicitly incorporate biased assimilation
- homophily not enough for polarization
- update opinion $x_i \in [0, 1]$ of node $i$ after interacting with neighbors

$$x_i \leftarrow \frac{w_{ii} x_i + x_i^{\beta} s_i}{w_{ii} + x_i^{\beta} s_i + (d_i - x_i)^{\beta}(1 - s_i)}$$

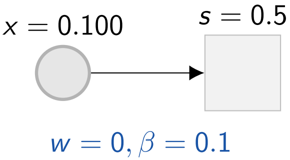
  where, $s_i$ is the average opinion of the neighbors of $i$, $d_i$ is the weighted degree of $i$, and $\beta$ is a bias parameter

- model becomes equivalent to the degroot model for $\beta = 0$
- result: for $\beta > 1$, the biased-assimilation model is polarizing

# the biased-assimilation model

a single agent in a fixed environment:

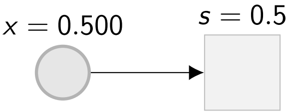
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$



# the biased-assimilation model

a single agent in a fixed environment:

$$x \leftarrow \frac{wx + x^\beta s}{w + x^\beta s + (1-x)^\beta (1-s)}.$$

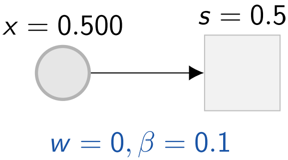


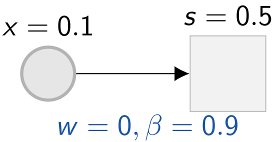

$w = 0, \beta = 0.1$

# the biased-assimilation model

a single agent in a fixed environment:

$$x \leftarrow \frac{wx + x^\beta s}{w + x^\beta s + (1-x)^\beta (1-s)}.$$

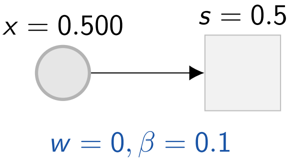


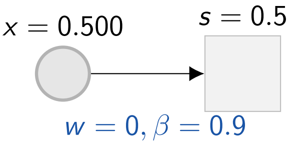



# the biased-assimilation model

a single agent in a fixed environment:

$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

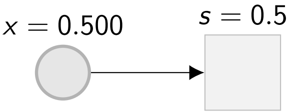


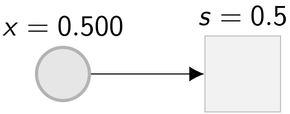
# the biased-assimilation model

a single agent in a fixed environment:

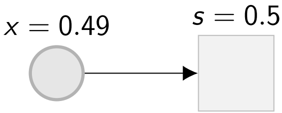
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$




$w = 0, \beta = 0.1$




$w = 0, \beta = 0.9$

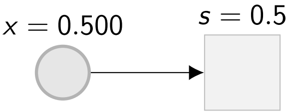



$w = 0, \beta = 5$

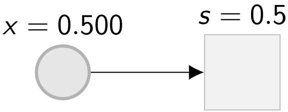
# the biased-assimilation model

a single agent in a fixed environment:

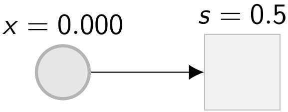
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$




$w = 0, \beta = 0.1$




$w = 0, \beta = 0.9$



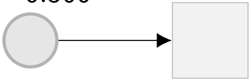

$w = 0, \beta = 5$

# the biased-assimilation model

a single agent in a fixed environment:

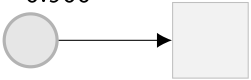
$$x \leftarrow \frac{wx + x^{\beta} s}{w + x^{\beta} s + (1-x)^{\beta}(1-s)}.$$

$x = 0.500$ $s = 0.5$



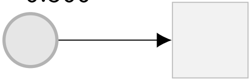
$w = 0, \beta = 0.1$

$x = 0.900$ $s = 0.1$



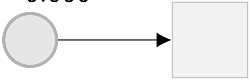
$w = 0, \beta = 2$

$x = 0.500$ $s = 0.5$



$w = 0, \beta = 0.9$

$x = 0.000$ $s = 0.5$



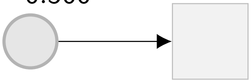
$w = 0, \beta = 5$

# the biased-assimilation model

a single agent in a fixed environment:

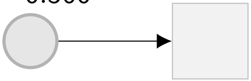
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

$x = 0.500$ $\quad s = 0.5$



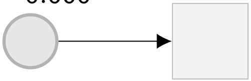
$w = 0, \beta = 0.1$

$x = 0.500$ $\quad s = 0.5$



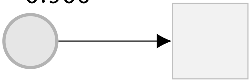
$w = 0, \beta = 0.9$

$x = 0.000$ $\quad s = 0.5$



$w = 0, \beta = 5$

$x = 0.900$ $\quad s = 0.1$



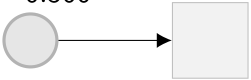
$w = 0, \beta = 2$

# the biased-assimilation model

a single agent in a fixed environment:

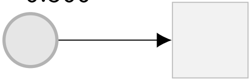
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

$x = 0.500$ $s = 0.5$



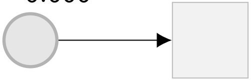
$w = 0, \beta = 0.1$

$x = 0.500$ $s = 0.5$



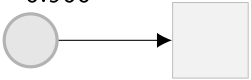
$w = 0, \beta = 0.9$

$x = 0.000$ $s = 0.5$



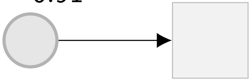
$w = 0, \beta = 5$

$x = 0.900$ $s = 0.1$



$w = 0, \beta = 2$

$x = 0.91$ $s = 0.1$

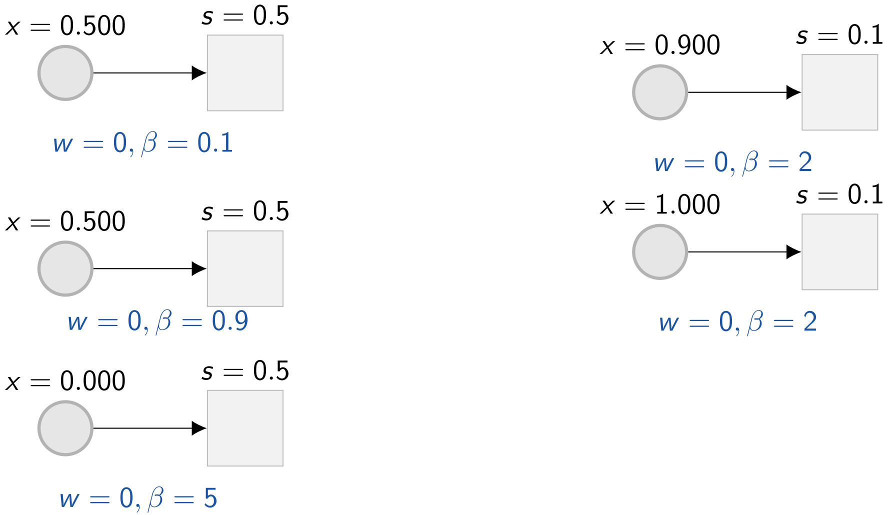


$w = 0, \beta = 2$

# the biased-assimilation model

a single agent in a fixed environment:

$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

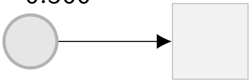


# the biased-assimilation model

a single agent in a fixed environment:

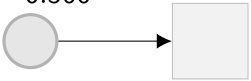
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

$x = 0.500$ $s = 0.5$



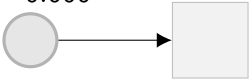
$w = 0, \beta = 0.1$

$x = 0.500$ $s = 0.5$



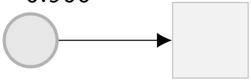
$w = 0, \beta = 0.9$

$x = 0.000$ $s = 0.5$



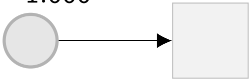
$w = 0, \beta = 5$

$x = 0.900$ $s = 0.1$



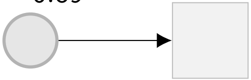
$w = 0, \beta = 2$

$x = 1.000$ $s = 0.1$



$w = 0, \beta = 2$

$x = 0.89$ $s = 0.1$



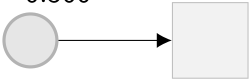
$w = 0, \beta = 2$

# the biased-assimilation model

a single agent in a fixed environment:

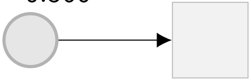
$$x \leftarrow \frac{wx + x^{\beta}s}{w + x^{\beta}s + (1-x)^{\beta}(1-s)}.$$

$x = 0.500$ $s = 0.5$



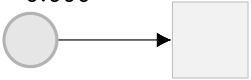
$w = 0, \beta = 0.1$

$x = 0.500$ $s = 0.5$



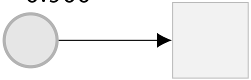
$w = 0, \beta = 0.9$

$x = 0.000$ $s = 0.5$



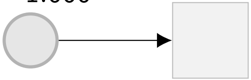
$w = 0, \beta = 5$

$x = 0.900$ $s = 0.1$



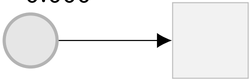
$w = 0, \beta = 2$

$x = 1.000$ $s = 0.1$



$w = 0, \beta = 2$

$x = 0.000$ $s = 0.1$



$w = 0, \beta = 2$

# beyond neighbour averaging

[Hazla et al., 2019]

population of agents with unit-length opinions in $\mathbb{R}^d$.

intervention $\mathbf{v}$ on agent with opinion $\mathbf{u}$:

$$\mathbf{u} := \frac{\mathbf{u} + \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v}}{\| \mathbf{u} + \langle \mathbf{u}, \mathbf{v} \rangle \mathbf{v} \|}.$$

interesting quirk: $\mathbf{v}$ and $-\mathbf{v}$ have the same effect.

# beyond neighbour averaging

[Hazla et al., 2019]

example: 500 agents $\mathbf{u}_i$ sampled uniformly from the sphere $u_{i,4} = 0$ in $\mathbb{R}^4$, with $\|\mathbf{u}_i\| = 1$

that is, $\mathbf{u}_i^{(1)} = (u_{i,1}, u_{i,2}, u_{i,3}, 0)$

intervention:

$$\mathbf{v} = (\beta, 0, 0, \alpha), \quad \text{where} \quad \alpha = \frac{3}{4}, \ \beta = \sqrt{1 - \alpha^2}$$

$\mathbf{u}_i^{(1)} + \langle \mathbf{u}_i^{(1)}, \mathbf{v} \rangle \mathbf{v} = ((1 + \beta) u_{i,1}, u_{i,2}, u_{i,3}, \alpha\beta u_{i,1})$

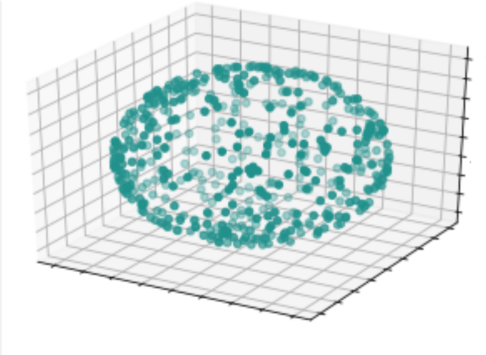
opinion on new product is represented by 4th coordinate, which is initially 0 for all agents

agents form opinion with respect to 4th coordinate

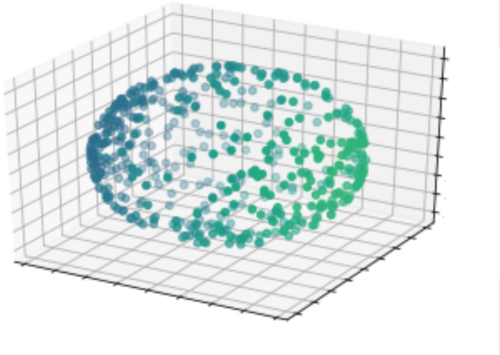
in addition, agents get polarized with respect to the 3 first coordinates

# beyond neighbour averaging

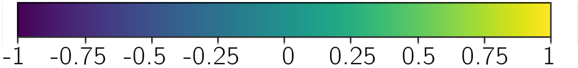
[Hazla et al., 2019]



$t = 1$

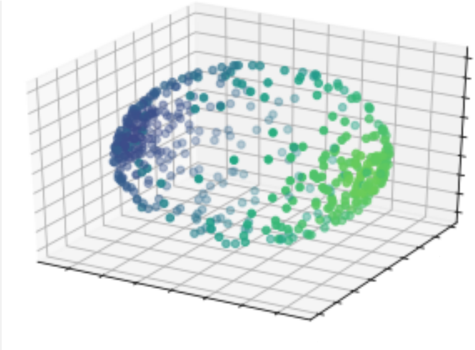


$t = 2$

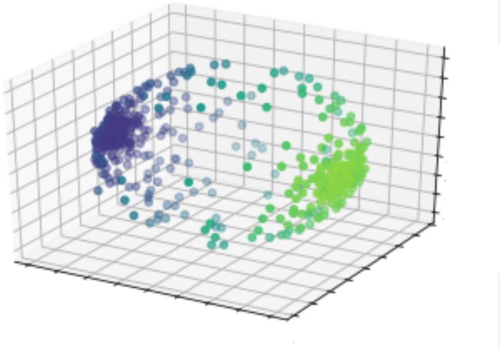


# beyond neighbour averaging

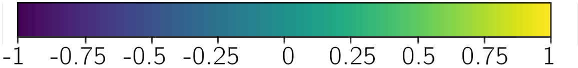
[Hazla et al., 2019]



$t = 3$

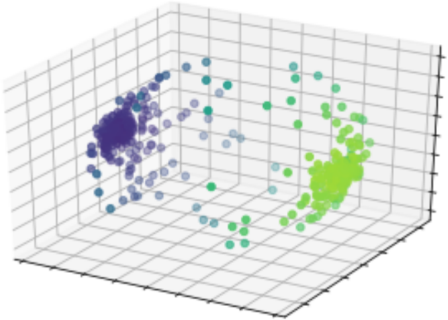


$t = 4$

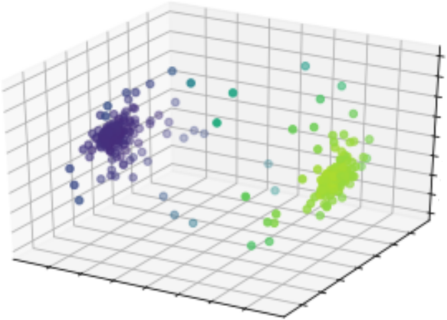


# beyond neighbour averaging

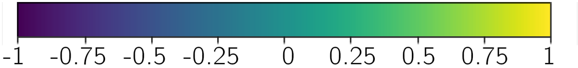
[Hazla et al., 2019]



$t = 5$



$t = 6$



# recap

- the DeGroot and Friedkin–Johnsen (FJ) models
  - convergence
  - consensus
- other opinion formation models
  - bounded confidence ($\epsilon$-threshold for updates)
  - biased assimilation (DeGroot with reweighted updates, favoring similar opinions)
  - geometric model (opinions in $\mathbb{R}^n$)

# properties of the Friedkin–Johnsen (FJ) model

# Friedkin–Johnsen (FJ) model: brief reminder

- network given by a graph $G = (V, E, w)$
- each user $i$ has an innate opinion $s_i$ and an expressed opinion $z_i$
- model proceeds in rounds, with the following updating rule for expressed opinions:

$$z_i^{(t+1)} = \frac{s_i + \sum_{j \mid (i,j) \in E} w_{ij} z_j^{(t)}}{1 + \sum_{j \mid (i,j) \in E} w_{ij}}$$

- equilibrium expressed opinions are given by $\mathbf{z}^* = \lim_{t \to \infty} \mathbf{z}^{(t)} = (I + L)^{-1} \mathbf{s}$

  where $L$ is a Laplacian matrix associated with the social network

# property of the expressed opinions

- other justifications for the update rule of expressed opinions?

$$z_i^{(t+1)} = \frac{s_i + \sum_{j \mid (i,j) \in E} w_{ij} z_j^{(t)}}{1 + \sum_{j \mid (i,j) \in E} w_{ij}}$$

# property of the expressed opinions

- other justifications for the update rule of expressed opinions?

$$z_i^{(t+1)} = \frac{s_i + \sum_{j \mid (i,j) \in E} w_{ij} z_j^{(t)}}{1 + \sum_{j \mid (i,j) \in E} w_{ij}}$$

- for user $i$, consider the cost function

$$(z_i^{(t)} - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij} (z_i^{(t)} - z_j^{(t)})^2$$

  - first term corresponds to conflict between internal and expressed opinion
  - second term corresponds to $i$'s conflict with their neighbors

# property of the expressed opinions

- other justifications for the update rule of expressed opinions?

$$z_i^{(t+1)} = \frac{s_i + \sum_{j \mid (i,j) \in E} w_{ij} z_j^{(t)}}{1 + \sum_{j \mid (i,j) \in E} w_{ij}}$$

- for user $i$, consider the cost function

$$(z_i^{(t)} - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij} (z_i^{(t)} - z_j^{(t)})^2$$

  - first term corresponds to conflict between internal and expressed opinion
  - second term corresponds to $i$'s conflict with their neighbors

- if user $i$ sets $z_i^{(t+1)}$ to minimize this cost function,
  the choice of $z_i^{(t+1)}$ is the same as in the update rule above

# the price of anarchy in opinion formation

- how bad is forming your own opinion? [Bindel et al., 2015]

# the price of anarchy in opinion formation

- how bad is forming your own opinion? [Bindel et al., 2015]
- in the FJ model, each node is independently minimizing their own cost

$$c_i(z_i) = (z_i - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij}(z_i - z_j)^2$$

this results to a Nash equilibrium

# the price of anarchy in opinion formation

- how bad is forming your own opinion? [Bindel et al., 2015]

- in the FJ model, each node is independently minimizing their own cost

$$c_i(z_i) = (z_i - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij}(z_i - z_j)^2$$

  this results to a Nash equilibrium

- what instead if we ask to optimize the social cost

$$c(\mathbf{y}) = \sum_{i \in V} c_i(y_i)$$

# the price of anarchy in opinion formation

- how bad is forming your own opinion? [Bindel et al., 2015]
- in the FJ model, each node is independently minimizing their own cost

$$c_i(z_i) = (z_i - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij}(z_i - z_j)^2$$

  this results to a Nash equilibrium
- what instead if we ask to optimize the social cost

$$c(\mathbf{y}) = \sum_{i \in V} c_i(y_i)$$

- theorem ([Bindel et al., 2015])

  price of anarchy (ratio of costs) is at most $9/8$ for any undirected graph $G$

# the price of anarchy in opinion formation

- how bad is forming your own opinion? [Bindel et al., 2015]
- in the FJ model, each node is independently minimizing their own cost

$$c_i(z_i) = (z_i - s_i)^2 + \sum_{j \mid (i,j) \in E} w_{ij}(z_i - z_j)^2$$

  this results to a Nash equilibrium
- what instead if we ask to optimize the social cost

$$c(\mathbf{y}) = \sum_{i \in V} c_i(y_i)$$

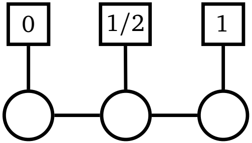
- theorem ([Bindel et al., 2015])

  price of anarchy (ratio of costs) is at most $9/8$ for any undirected graph $G$

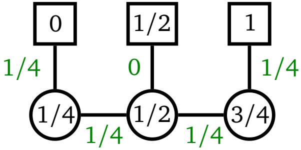
$\Rightarrow$ this result is for undirected networks;

for directed networks the price of anarchy can be much higher

# the price of anarchy in opinion formation — example [Bindel et al., 2015]

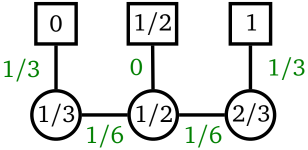



Nash equilibrium




$$\text{Nash cost} = 3 * 2 * (1/4)^2 = 3/8$$

Social optimal




$$\text{Opt cost} = 2 * ((1/3)^2 + (1/6)^2) + 2 * (1/6)^2 = 1/3$$

$$\text{Price of anarchy} = \frac{\text{Nash cost}}{\text{Opt cost}} = \frac{3/8}{1/3} = \frac{9}{8}$$

# quantities of interest in the Friedkin-Johnsen model

- given the equilibrium expressed opinions $\mathbf{z}^*$ and innate opinions $\mathbf{s}$, we can study more complex phenomena in the network
- we can quantify polarization, disagreement, etc.

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |

sum of opinions: sums all user opinions — relevant for marketing campaigns

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |

polarization: the variance of the opinions, where $\bar{z} = \frac{1}{|V|} \sum_{i \in V} z_i^*$ is average opinion

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |

controversy: measures extremity of opinions, can also be viewed as radicalization

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |

if $z^*$ is mean-centered, i.e., $\bar{z} = \sum_i z_i^* = 0$, controversy $\mathcal{C}$ and polarization $\mathcal{P}$ are identical

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ |

internal conflict: measures tension between users' innate and expressed opinions

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$ |

disagreement: measures the tension between neighbors in the network;
sometimes called *external conflict*

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$ |
| polarization-disagreement index | $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$ |

polarization-disagreement: combination of polarization and disagreement, useful for analysis

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$ |
| polarization-disagreement index | $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$ |

conservation law of conflict: $\mathcal{I} + 2\mathcal{D} + \mathcal{C} = \mathbf{s}^\intercal \mathbf{s}$ [Chen et al., 2018]

# quantities of interest in the Friedkin-Johnsen model

| | |
|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} (z_i^* - z_j^*)^2$ |

using $\mathbf{z}^* = (I + L)^{-1}\mathbf{s}$, we can express these measures as quadratic forms

# quantities of interest in the Friedkin-Johnsen model

| | | |
|---|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ | $= \mathbf{1}^\top \mathbf{z}^* = \mathbf{1}^\top (I + L)^{-1} \mathbf{s}$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ | $= \mathbf{s}^\top (I + L)^{-1} (I - \frac{\mathbf{1}\mathbf{1}^\top}{n}) (I + L)^{-1} \mathbf{s}$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ | $= \mathbf{s}^\top (I + L)^{-2} \mathbf{s}$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ | $= \mathbf{s}^\top (I + L)^{-1} L^2 (I + L)^{-1} \mathbf{s}$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} (z_i^* - z_j^*)^2$ | $= \mathbf{s}^\top (I + L)^{-1} L (I + L)^{-1} \mathbf{s}$ |

where $\mathbf{1}$ is the all-ones vectors, $I$ is the identity matrix, $L$ is the graph Laplacian, $\mathbf{s}$ is the vector of innate opinions, and $\bar{z} = \frac{1}{|V|} \sum_{i \in V} z_i^*$

# quantities of interest in the Friedkin-Johnsen model

| | | |
|---|---|---|
| sum of opinions | $\mathcal{S} = \sum_{i \in V} z_i^*$ | $= \mathbf{1}^\intercal \mathbf{z}^* = \mathbf{1}^\intercal (I + L)^{-1} \mathbf{s}$ |
| polarization index | $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$ | $= \mathbf{s}^\intercal (I + L)^{-1} (I - \frac{\mathbf{1}\mathbf{1}^\intercal}{n}) (I + L)^{-1} \mathbf{s}$ |
| controversy index | $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$ | $= \mathbf{s}^\intercal (I + L)^{-2} \mathbf{s}$ |
| internal-conflict index | $\mathcal{I} = \sum_{i \in V} (z_i^* - s_i)^2$ | $= \mathbf{s}^\intercal (I + L)^{-1} L^2 (I + L)^{-1} \mathbf{s}$ |
| disagreement index | $\mathcal{D} = \sum_{(i,j) \in E} (z_i^* - z_j^*)^2$ | $= \mathbf{s}^\intercal (I + L)^{-1} L (I + L)^{-1} \mathbf{s}$ |

where $\mathbf{1}$ is the all-ones vectors, $I$ is the identity matrix, $L$ is the graph Laplacian,
$\mathbf{s}$ is the vector of innate opinions, and $\bar{z} = \frac{1}{|V|} \sum_{i \in V} z_i^*$
all these matrices are positive semidefinite

# algorithmic interventions for moderating opinions

# interventions

- examples for interventions: a timeline algorithm changes the network structure, an adversary makes people change their innate opinions, . . .

# interventions

- examples for interventions: a timeline algorithm changes the network structure, an adversary makes people change their innate opinions, . . .

- *formal way to study this:* define an optimization problem, where:
  - the objective function encodes the desired goal
  - the constraints encode the “power” of the intervention

# interventions

- examples for interventions: a timeline algorithm changes the network structure, an adversary makes people change their innate opinions, . . .
- *formal way to study this:* define an optimization problem, where:
  - the objective function encodes the desired goal
  - the constraints encode the "power" of the intervention
- example: a social network provider wants to minimize polarization and disagreement by changing the network structure [Musco et al., 2018, Zhu et al., 2021]
- example: an adversary wants to maximize the disagreement and has the power to change $k$ user opinions [Chen and Racz, 2021, Gaitonde et al., 2020]

# interventions: literature overview

- what to optimize
  - minimize price of anarchy [Bindel et al., 2015]
  - reduce polarization and disagreement [Matakos et al., 2017, Musco et al., 2018]
  - maximize sum of opinions [Gionis et al., 2013, Tu and Neumann, 2022]
  - increase disagreement [Chen and Racz, 2021, Gaitonde et al., 2020]
- what properties to modify
  - innate or expressed opinions [Gionis et al., 2013, Matakos et al., 2017]
  - graph weights [Abebe et al., 2018]
  - graph structure [Bindel et al., 2015, Musco et al., 2018] [Zhu et al., 2021, Rácz and Rigobon, 2022]

# opinion maximization in social networks

- select $k$ nodes to set their expressed opinion to $z_i^* = 1$ so as to maximize the sum of opinions [Gionis et al., 2013]

$$\mathcal{S} = \sum_{i \in V} z_i^*$$

  - motivation: lobbying for a cause or campaign

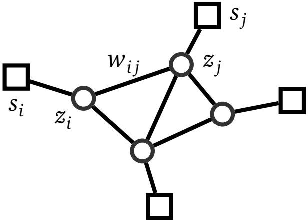
# opinion maximization in social networks

- select $k$ nodes to set their expressed opinion to $z_i^* = 1$ so as to maximize the sum of opinions [Gionis et al., 2013]

$$\mathcal{S} = \sum_{i \in V} z_i^*$$

  - motivation: lobbying for a cause or campaign

- Greedy gives $(1 - 1/e)$ approximation
- objective function is monotone and submodular
- technical observation: consider an absorbing random walk, with absorbing states the nodes that correspond to the innate opinions; then $z_i^*$ is can be interpreted as the expected value at absorption, when starting a random walk in node $i$



# minimizing polarization and disagreement in social networks

- focus on minimizing the following indices: [Musco et al., 2018]
  - polarization: $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$
  - disagreement: $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$
  - polarization-disagreement: $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$

# minimizing polarization and disagreement in social networks

- focus on minimizing the following indices: [Musco et al., 2018]
  - polarization: $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$
  - disagreement: $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$
  - polarization-disagreement: $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$
- constraint: we can decrease the innate opinions within a given budget and $\ell_1$-distances, i.e., $||\mathbf{s} - \mathbf{s}'||_1 \leq B$ and $\mathbf{s}' \leq \mathbf{s}$
- result: optimizing these indices is convex and can be solved in polynomial time

# minimizing polarization and disagreement in social networks

- focus on minimizing the following indices: [Musco et al., 2018]
  - polarization: $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$
  - disagreement: $\mathcal{D} = \sum_{(i,j) \in E} w_{ij} (z_i^* - z_j^*)^2$
  - polarization-disagreement: $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$
- constraint: we can decrease the innate opinions within a given budget and $\ell_1$-distances, i.e., $||\mathbf{s} - \mathbf{s}'||_1 \leq B$ and $\mathbf{s}' \leq \mathbf{s}$
- result: optimizing these indices is convex and can be solved in polynomial time
- what if we can change the graph topology with a fixed number of edges?
  - minimizing $\mathcal{I}_{pd}$ is convex
  - thus, it can be solved with standard-convex optimization methods
  - when one of the terms $\mathcal{P}$ or $\mathcal{C}$ is weighted differently, problem is not convex

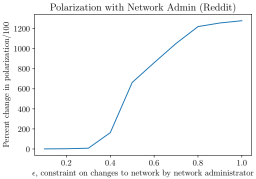
# analyzing the impact of filter bubbles on social network polarization

- study the interplay between users and a network administrator [Chitra and Musco, 2020]
- the dynamics proceed in iterations — in each iteration
  - the users adjust their expressed opinions according to the FJ model
  - the network administrator slightly adjusts the network to minimize disagreement $\mathcal{D}$ until convergence
- intuition: network administrators want less disagreement, as this implies "happier" users

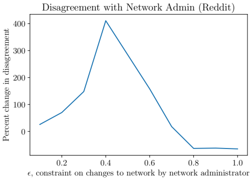
# analyzing the impact of filter bubbles on social network polarization

- study the interplay between users and a network administrator [Chitra and Musco, 2020]
- the dynamics proceed in iterations — in each iteration
  - the users adjust their expressed opinions according to the FJ model
  - the network administrator slightly adjusts the network to minimize disagreement $\mathcal{D}$ until convergence
- intuition: network administrators want less disagreement, as this implies "happier" users
- it is shown experimentally that polarization increases
- authors suggest this explains why recommender systems increase polarization and introduce filter bubbles

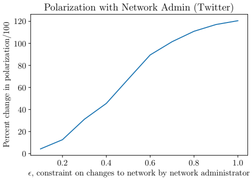
# analyzing the impact of filter bubbles on social network polarization



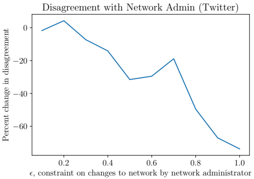

**(a) Change in polarization, Reddit network**




**(b) Change in disagreement, Reddit network**




**(c) Change in polarization, Twitter network**




**(d) Change in disagreement, Twitter network**

[Chitra and Musco, 2020]

# conclusion, limitations, reflections

# summary

- opinion formation in social networks is an active area of research
  - work both in mathematical modeling and computational social science
- we reviewed common opinion-formation models
  - DeGroot and Friedkin-Johnsen models, other opinion formation models
  - discussed properties of the models and measures of interest
- discussed how polarization may emerge from these models
  - e.g., emergence of echo chambers
- reviewed interventions for moderating opinions
- no discussion on misinformation and disinformation

# challenges, limitations

- validation of the mathematical models is very challenging
  - models are often too simplistic, e.g., opinions in [0,1], opinions are updated by a simple weighted-averaging operation
  - lack of complete and unbiased data
    often access data to a single social-media platform, e.g., twitter
    data is biased: representativeness, US politics, impact of bots
  - models involve parameters that are difficult to estimate in practice
- models use mostly network structure, and ignore language analysis
  - this makes them language-independent, but incorporating language can help greatly

# ethical issues on interventions

a common intervention action is to aim to reduce polarization, or increase diversity, by making judicious recommendations

Q: is it ethical to tamper with users' feed?

Q: can such methods facilitate manipulation?

A: UI, user control, and transparency needs to be addressed separately

A: content prioritization and recommendation algorithms are already in place, and they

- are mainly aiming at increasing engangemnnt and monetization
- are not transparent
- are not offering control to the users
- do not have built-in ethical specifications

# references I


Abebe, R., Kleinberg, J., Parkes, D., and Tsourakakis, C. E. (2018).
Opinion dynamics with varying susceptibility to persuasion.
In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*, pages 1089–1098.

Bindel, D., Kleinberg, J., and Oren, S. (2015).
How bad is forming your own opinion?
*Games and Economic Behavior*, 92:248–265.

Chen, M. F. and Racz, M. Z. (2021).
An adversarial model of network disruption: Maximizing disagreement and polarization in social networks.
*IEEE Transactions on Network Science and Engineering*, 9(2):728 – 739.

Chen, X., Lijffijt, J., and Bie, T. D. (2018).
Quantifying and minimizing risk of conflict in social networks.
In *KDD*, pages 1197–1205.

Chen, X., Tsaparas, P., Lijffijt, J., and De Bie, T. (2021).
Opinion dynamics with backfire effect and biased assimilation.
*PloS one*, 16(9):e0256922.

# references II


Chitra, U. and Musco, C. (2020).
Analyzing the impact of filter bubbles on social network polarization.
In *Proceedings of the 13th International Conference on Web Search and Data Mining*, pages 115–123.

Cohen, M. B., Kyng, R., Miller, G. L., Pachocki, J. W., Peng, R., Rao, A. B., and Xu, S. C. (2014).
Solving SDD linear systems in nearly $m\log^{1/2} n$ time.
In *STOC*, pages 343–352.

Dandekar, P., Goel, A., and Lee, D. T. (2013).
Biased assimilation, homophily, and the dynamics of polarization.
*Proceedings of the National Academy of Sciences*, 110(15):5791–5796.

Deffuant, G., Neau, D., Amblard, F., and Weisbuch, G. (2000).
Mixing beliefs among interacting agents.
*Advances in Complex Systems*, 3(01n04):87–98.

DeGroot, M. H. (1974).
Reaching a consensus.
*Journal of the American Statistical Association*, 69(345):118–121.

# references III


Friedkin, N. E. and Johnsen, E. C. (1990).
Social influence and opinions.
*Journal of Mathematical Sociology*, 15(3-4):193–206.

Gaitonde, J., Kleinberg, J., and Tardos, E. (2020).
Adversarial perturbations of opinion dynamics in networks.
In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 471–472.

Gaitonde, J., Kleinberg, J., and Tardos, É. (2021).
Polarization in geometric opinion dynamics.
In *Proceedings of the 22nd ACM Conference on Economics and Computation*, pages 499–519.

Gionis, A., Terzi, E., and Tsaparas, P. (2013).
Opinion maximization in social networks.
In *Proceedings of the 2013 SIAM International Conference on Data Mining*, pages 387–395. SIAM.

Golub, B. and Jackson, M. O. (2010).
Naive learning in social networks and the wisdom of crowds.
*American Economic Journal: Microeconomics*, 2(1):112–49.

# references IV


Granovetter, M. (1978).
Threshold models of collective behavior.
*American journal of sociology*, 83(6):1420–1443.

Hazla, J., Jin, Y., Mossel, E., and Ramnarayan, G. (2019).
A geometric model of opinion polarization.
*arXiv preprint arXiv:1910.05274*.

Hegselmann, R., Krause, U., et al. (2002).
Opinion dynamics and bounded confidence models, analysis, and simulation.
*Journal of artificial societies and social simulation*, 5(3).

Kempe, D., Kleinberg, J., and Tardos, E. (2003).
Maximizing the spread of influence through a social network.
In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM Press.

Krause, U. (2000).
A discrete nonlinear and non-autonomous model of consensus.
In *Communications in Difference Equations: Proceedings of the Fourth International Conference on Difference Equations*, page 227. CRC Press.

# references V


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*, 37(11):2098.

Matakos, A., Terzi, E., and Tsaparas, P. (2017).
Measuring and moderating opinion polarization in social networks.
*Data Mining and Knowledge Discovery*, 31(5):1480–1505.

Musco, C., Musco, C., and Tsourakakis, C. E. (2018).
Minimizing polarization and disagreement in social networks.
In *Proceedings of the 2018 World Wide Web Conference*, pages 369–378.

Rácz, M. Z. and Rigobon, D. E. (2022).
Towards consensus: Reducing polarization by perturbing social networks.
*CoRR*, abs/2206.08996.

Tu, S. and Neumann, S. (2022).
A viral marketing-based model for opinion dynamics.
In *WebConf*.

# references VI


Xu, W., Bao, Q., and Zhang, Z. (2021).
Fast Evaluation for Relevant Quantities of Opinion Dynamics.
In *WWW*, pages 2037–2045.

Zhu, L., Bao, Q., and Zhang, Z. (2021).
Minimizing polarization and disagreement in social networks via link recommendation.
In *NeurIPS*.