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

Lecture 5: Opinion dynamics in social networks

Bertinoro International Spring School 2022

Aristides Gionis

`argioni@kth.se`

KTH Royal Institute of Technology

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

overview of this lecture

- models of opinion formation in social networks
  - DeGroot model
  - Friedkin-Johnsen model
  - polarization and disagreement indices
- interventions for moderating opinions
  - maximizing opinions / minimizing polarization and disagreement
  - emergence of echo chambers

models of opinion formation

- individuals' opinions are influenced by their peers
- how to model the opinion-formation process in a social network?
- so far we saw influence in a network as a cascade
  - a discrete entity (action, meme, virus) propagates in a network
  - cascade is modeled using the independent-cascade model
- social balance can also be seen as an opinion-formation process
  - opinions are on pairwise relations, which can be positive or negative

opinion formation by weighted averaging

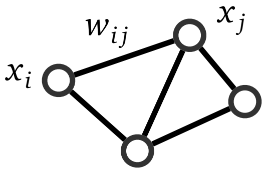
- opinion of each individual is represented by a real number
  - for example, a number in the interval $[0, 1]$
- at each time step, each individual updates its opinion as a weighted average of the opinions of its neighbors
- the process continues until convergence

opinion formation by weighted averaging

- a basic model [DeGroot, 1974]
- we consider an undirected and weighted graph modeling a social network
- weight w_{ij} represents influence of node j to i
- node i has opinion x_i
- node i updates its opinion by

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

where $\sum_j w_{ij} = 1$, for all i



extensions on the basic DeGroot model

- weight w_{ii} models tendency to keep existing opinion
- weights may change with time or with opinions, i.e., $w_{ij}(t, x_i^t)$
- nodes interact pairwise (asynchronous) or all-at-once (synchronous)

convergence of the basic DeGroot model

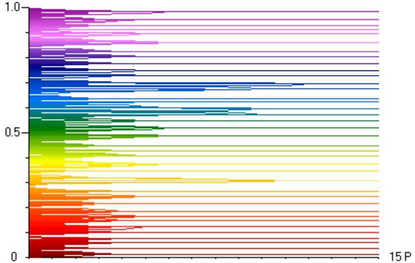
- lemma: in the basic DeGroot model all nodes reach to the same opinion, i.e., the process converges to a consensus
- this result holds with conditions on the structure of the graph
  - connectivity (or strong connectivity in the directed case) and aperiodicity
- corollary: DeGroot's process is not polarizing

the bounded-confidence model

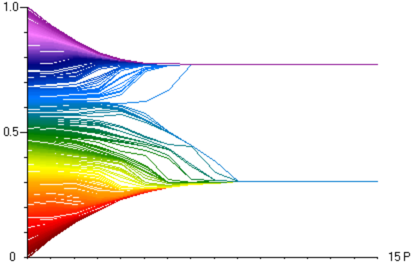
[Deffuant et al., 2000]

- individuals only interact and update their opinions if the difference between their existing opinions is smaller than a threshold ϵ
- this threshold models "openness to discussion"
- larger ϵ produce consensus, while smaller ϵ produce polarized opinions
- the model can be thought as a form of selective exposure
- result: for certain values of ϵ the bounded-confidence model can lead to polarization

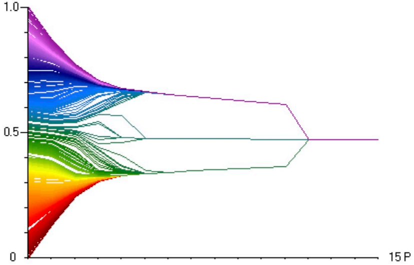
the bounded-confidence model — simulation results



ε=0.02



ε=0.3



ε=0.5

the biased-assimilation model

[Dandekar et al., 2013]

- modify DeGroot's model to explicitly incorporate 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.

- in particular, modify weighted average to be non-linear
  - neighbors with similar opinions are weighted more
  - so, opinions of individuals are reinforced by like-minded neighbors
- result: under certain conditions the biased-assimilation model can lead to polarization
  - opinion of moderate individuals can go to extremes (0 or 1)
  - network disagreement index $\mathcal{D}(\mathbf{x}) = \sum_{(i,j)\in E} w_{ij}(x_i - x_j)^2$ can increase with time t

the biased-assimilation model

- update opinion $x_i \in [0, 1]$ of node i after interacting with node j

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

  where, s_i is the average opinion of the neighbors, and β is a bias parameter

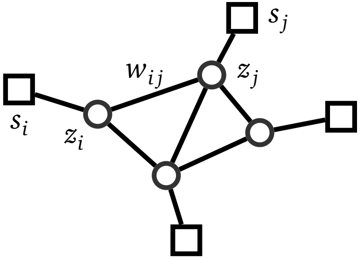
- model becomes equivalent to the DeGroot model for $\beta = 0$
- result: for $\beta > 1$, the biased-assimilation model is polarizing

the Friedkin-Johnsen model

[Friedkin and Johnsen, 1990]

- model introduces innate opinions
- node i has innate opinion s_i and expressed opinion z_i
- node i updates its expressed opinion by

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



- can be seen as node i selecting z_i so as to minimize cost

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

given all other z_j

- model converges to an equilibrium, but not necessarily consensus opinion

deriving the expressed opinions in the Friedkin-Johnsen model

- model can be seen as node i minimizing the cost function

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

- setting $c_i'(z_i) = 0$ gives $(z_i - s_i) + \sum_{j \mid (i,j) \in E} w_{ij}(z_i - z_j) = 0$ for all i
- consider the Laplacian matrix L with $L_{ij} = -w_{ij}$ and $L_{ii} = \sum_{j \mid (i,j) \in E} w_{ij}$
- the equilibrium equation can be written as $(I + L)\,\mathbf{z} = \mathbf{s}$
- thus, we can solve for the expressed opinions by $\mathbf{z} = (I + L)^{-1}\,\mathbf{s}$

the price of anarchy in opinion formation

[Bindel et al., 2015]

- how bad is forming your own opinion?
- in the FJ model, each node is independently minimizing its own cost

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

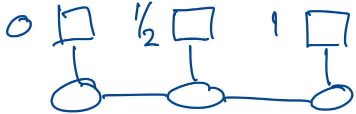
  this results to a Nash equilibrium

- what instead if we ask to optimize the social cost

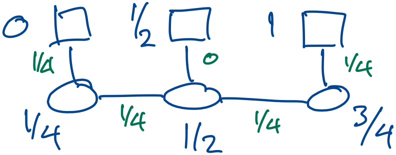
$$c(\mathbf{y}) = \sum_{i \in V} c_i(y_i)$$

- theorem ([Bindel et al., 2015]) ratio of costs is at most $9/8$ for any graph G

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

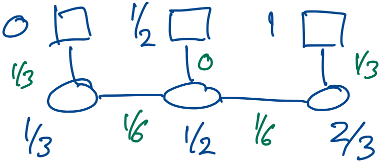


Nash equilibrium



$$\text{Nash cost} = 3 \times 2 \times \left(\frac{1}{4}\right)^2 = 3/8$$

social optimal



$$\text{opt} = 2 \times \left(\left(\frac{1}{3}\right)^2 + \left(\frac{1}{6}\right)^2\right) + 2\left(\frac{1}{6}\right)^2 = 1/3$$

$$\text{price of anarchy} = \frac{\text{Nash}}{\text{opt}} = \frac{3/8}{1/3} = \frac{9}{8}$$

quantities of interest in the Friedkin-Johnsen model

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

interventions

- 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]
- what to optimize
  - minimize price of anarchy [Bindel et al., 2015]
  - maximize sum of opinions [Gionis et al., 2013]
  - reduce polarization and disagreement [Matakos et al., 2017, Musco et al., 2018]
  - increase disagreement [Chen and Racz, 2020, Gaitonde et al., 2020]

opinion maximization in social networks

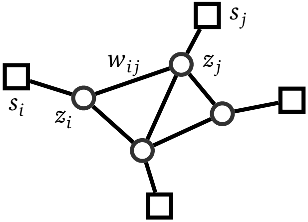
[Gionis et al., 2013]

- select k nodes to set their opinion to $z_i = 1$ so as to maximize the sum of opinions

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

  - motivation: lobbying for a cause or campaign

- technical observation: the opinion z_i^* at equilibrium for node i can be interpreted as the expected value at absorption, in an absorbing random walk, with absorbing states the nodes that correspond to the innate opinions



- objective function is monotone and submodular

- GREEDY gives $(1 - 1/e)$ approximation

measuring and moderating opinion polarization in social networks

[Matakos et al., 2017]

- opinions are assumed to be in the interval $[-1, 1]$, furthermore assume $\bar{z} = 0$
- goal is to minimize controversy index $\mathcal{C} = \sum_{i \in V} (z_i^*)^2$
  - equivalent to polarization index $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$, since $\bar{z} = 0$
- select k nodes to set $z_i = 0$, or $s_i = 0$ (i.e., become moderate), to minimize $\mathcal{C}$
- authors show that the problem is **NP**-hard
- they propose a binary orthogonal matching pursuit (BOMP) algorithm
- experimentally compare BOMP with GREEDY, PAGERANK, and other baselines
  - BOMP with GREEDY are the best-performing methods

minimizing polarization and disagreement in social networks

[Musco et al., 2018]

- focus on minimizing the following indices
  - polarization: $\mathcal{P} = \sum_{i \in V} (z_i^* - \bar{z})^2$
  - disagreement: $\mathcal{D} = \sum_{(i,j) \in E} (z_i^* - z_j^*)^2$
  - polarization-disagreement: $\mathcal{I}_{pd} = \mathcal{P} + \mathcal{D}$
- aim to achieve this by finding optimal graph topology with a fixed number of edges
- optimizing $\mathcal{I}_{pd}$ is convex
  - thus, it can be solved with standard-convex optimization methods
- surprisingly, when one of the terms $\mathcal{P}$ or $\mathcal{C}$ is weighted differently, problem is not convex
- optimizing $\mathcal{I}_{pd}$ by changing innate opinions within a given budget, is polynomial

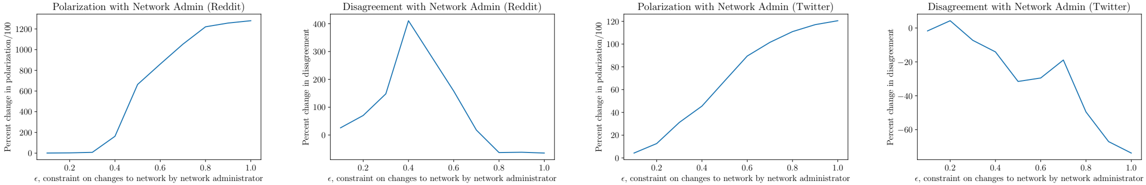
analyzing the impact of filter bubbles on social network polarization

[Chitra and Musco, 2020]

- study the interplay between users and network administrator
- study opinion evolution in the social network, in a game-theoretic setting
- the game proceeds 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
- it is shown experimentally that polarization increases
- authors suggest this as a viable model to explain 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**

Figure 2: Applying network administrator dynamics to real-world social networks. Details in Section 3.

[Chitra and Musco, 2020]

summary

- reviewed most common opinion-formation models in social networks
- discussed how polarization may emerge from these models
- reviewed methods for intervention and mediation

limitations of the methods presented in this course

- no discussion on misinformation and disinformation — need a separate course
- models use mostly network structure
  - language-independent, but incorporating language can help
- simple models
  - two-sided controversies
  - simple influence propagation models
- evaluation is challenging, done on few topics
- most of the analysis limited to twitter

reflections

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 engangemnt and monetization
- are not transparent
- are not offering control to the users
- do not have built-in ethical specifications

thank you!

questions?

argioni@kth.se

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. (2020).
Network disruption: maximizing disagreement and polarization in social networks.
arXiv preprint arXiv:2003.08377.

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.

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.

references II

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.

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.

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.

references III

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.