The Theory of Opinion Formation in Social Networks

Stefan Steinheber

May 7, 2024

The Theory of Opinion Formation in Social Networks

Stefan Steinheber
Department of Computer Science
Technical University Vienna
Vienna, 1040
e12022506@student.tuwien.ac.at

Abstract

With social networks being an integral part to our lives they impact many parts of our lives, from the way we consume news to the things we purchase to our core beliefs that shape our understanding of the world. In recent years the observation of more and more polarizing and more extremist opinions expressed in social networks could be made. The expression of these polarizing and extremist opinions pave the way to the formation of more extreme opinions.

The studies of the theory of opinion formation in social networks tries to understand the processes behind different phenomena observable in real world applications of social networks, like polarization. In this paper a compilation of different approaches and methods for planners, agents with a top-down power over the network, to reduce polarization. It was found that it is possible to reduce polarization in a social network by introducing new interactions between individuals that follow quite simple heuristics like the Disagreement Seeking (DS) heuristic presented in this paper. You can also that employing these strategies, only small changes in the network have to be made to drastically reduce polarization.

1 Introduction

Social Media and Online Social Networks (OSN) have become an integral part of many people's lives. With the transition of media consumption from traditional media, like newspapers, radio & TV, to digital media with platforms like YouTube, X (formerly Twitter), Facebook, Instagram etc. the consumption of news is also shifting away from traditional to digital. [Pew Research Center, 2023]

In 2023 56% of adults in the U.S. reported they "often" get shown news content on social media, whereas the traditional media (Television, Radio & Print) only cumulatively reach 57%. Pairing this with the U.S.-american average of 421 minutes spent on

This notion of increased social media usage and increased news consumption using these media presents itself problematic as the operators of the platform have full power over deciding what a user is shown on the platform. With the incentive of platform-operators to keep users as long as possible on the page to increase their revenue through ads, specific algorithms are in place deciding what a user will want to see and will want them to continue consuming content on the platform. The algorithms different platforms use have been alleged to (intentionally) increase the polarization and radicalization of users. The recommendation-algorithms are often criticised for exploiting the human tendency to more likely accept and engage with content that reflects their preexisting opinions and beliefs than content that contradicts it (confirmation bias) [Pohl, 2012, p.79ff]. This exploitation in turn leads to so called *echo chambers* or *filter bubbles*, in which like-minded people are mostly exposed to similar opinions (due to the way the algorithms function) which in turn disables them from experiencing greatly differing opinions, leading to the escalation and fundamentalizing, i.e. the opinion being more and more engrained in an individuals beliefs, of this *echoed* opinion.

Although there is no consensus on how far the impact of these algorithms actually goes, the general problematic of substantially influencing the opinion of individuals and thus leading to polarization is widely accepted. [Shaw, 2023, Törnberg, 2022]

To study this researchers have developed different models that try to formalize the mentioned and other behaviours in order to run simulations of social networks. In these models the social network is understood as a graph, in which each node represents an individual and an edge between two nodes means that in some way these two individuals can influence each others opinions, for example by seeing a post or receiving a message of the other. The models also define different sets of rules for running the simulation: Some models, like the Friedkin-Johnsen model described in 2.2, define initial opinion variables. If no initial opinion variable is set it is assumed that the opinions either are deferred from real-life data or assigned randomly. Every model needs to define a simulation step, which describes how each individual's opinion changes from one time-frame to the next. Through these an inductive simulation can be executed.

While it is highly discussed what effects the status quo of social networks has on opinion formation, an often overlooked to highlight what administrators or operators of social network platforms can do to mitigate the negative effects.

For this reason this paper will first discuss the most prevalent models in this field of research and then try to give concrete proposals for social network operators for how they can reduce the problematic impact their platforms have.

1.1 Types of social media networks

To better illustrate the findings in 3, I briefly want to categorize different types of social networks into 3 categories. They mainly differentiate themselves with how the content is posted not with what type of media is prevalent on the platform:

- **Forums**: In these social networks users can only post content in the context of a topic (sub-forums). Users can usually subscribe to sub-forums to stay updated about posts on these topics. Usually the topics have selected moderators which oversee the posted content. Users may have a *feed* which suggests posts on subscribed sub-forums but may also get recommended posts from sub-forums they are not subscribed to. Popular examples are Reddit and Telegram.
- **Feeds**: In this type of social media the content is mainly posted to the platform itself and doe not have to be linked to a page or a topic. This type itself has two sub-types:
  - **User-centric**: In user-centric social media mostly users are creators themselves, meaning there is a balance between creating and consuming content. Popular examples are TikTok, Instagram & Twitter.
  - **Creator-centric**: For this type the balance has shifted from every user posting and creating to some users creating and many more users consuming content. Popular example: YouTube.

It has to be noted that there is no hard cut between the different types of social media, as the transition for example it is also possible on YouTube to create posts for a specific topic. Even so a user can switch from being a consuming user to a creating user in Feed-style social media quite quickly and effortlessly. This differentiation solely caters to roughly classifying the dynamics prevalent in the proposed types.

1.2 Notation

In this paper a uniform notation to express equations that define models will be used. In this notation

- I will denote the individuals-set, which includes every individual in the network
- i will denote an individual of I, where $i \in I$
- $O_{i,t}$ will denote the opinion of a currently inspected node i at time t,
- $O_{j,t}$ will denote the opinion of a peer of the currently inspected node (with $j \in I$),
- $w_{i,j}$ will denote the weight between an individual i and j

- δ_i will denote the external factors on the opinion of the individual i
- ρ_i will denote the weight of external factors
- $e \in_R E$ will denote the random sampling of an element e in the set E
- $N(i) = \{j \in I : (i, j) \in E$ will denote the neighborhood of a vertex i

2 Related work

2.1 DeGroot Model

In 1974 the first model for representing the formation of opinions was developed by DeGroot. This model is calculating the opinions of individuals on a rudimentary level by assigning each individual an initial opinion: [DeGroot, 1974]

$$O_{i,0} = o \qquad \text{where } o \in \mathbb{R} \tag{1}$$

In each subsequent time step the opinion of an individual is computed as the arithmetic mean of all other opinions in the network and each opinion is weighted with a real-numbered weight $w_{i,j}$: [DeGroot, 1974]

$$O_{i,t} = \frac{\sum_{j \in I} O_{j,t-1} \times w_{i,j}}{|I|} \tag{2}$$

2.2 Friedkin-Johnsen Model

One of the most popular approaches and the approach many recent ones derive from was developed in 1990 by Friedkin and Johnsen. In this model the formation of opinions occurs iteratively, where each individual in the network holds an initial innate opinion s_i. This innate opinion is fixed and will not be shared with others, but contributes solely to the calculation of the opinion $O_{i,t}$.

Each individual i also has an expressed opinion $z_{i,t}$, which is calculated and updated with each iteration of a simulation. The expressed opinion is calculated as follows: [Neumann et al., 2024]

$$z_{i,t} = \frac{s_i + \sum_{j \in I} w_{j,i} \times z_{u,t-1}}{1 + \sum_{j \in I} w_{j,i}} \tag{3}$$

2.3 Hegelsmann-Krause Model

The importance of the effect of differing opinions in opinion formation was introduced into the model by Hegselmann and Krause. In this model the set of influenced opinions is limited by a confidence level ε_i for each individual. This confidence level represents how sure an individual is of their opinion and how willing they are to change it. The set of affected opinions or influence set S is calculated by: [Volkova et al., 2019, Hegselmann and Krause, 2002]

$$S_{i,t} = \forall j \in I, i \neq j : |O_{i,t-1} - O_{j,t-1}| \leqslant \varepsilon \tag{4}$$

Each iteration of the simulation will then calculate the opinion of an individual equally to DeGroot, shown in 2 with the alteration that the only the average opinions of the influence set are taken into account:

$$O_{i,t} = \frac{\sum_{j \in S_{i,t}} O_{j,t-1} \times w_{i,j}}{|S_{i,t}|} \tag{5}$$

2.4 Deffuant-Weißbuch Model

Deffuant et al. introduced a model quite similar to the Hegelsmann-Krause Model described in 2.3, but instead of updating the opinions of all individuals in each time-step, a single individual i is sampled randomly from the set of all individuals in the network I. From this individual i a random neighbour of i, j is sampled, whose opinion will be used for updating $O_{i,t}$ [Deffuant et al., 2001]:

$$j \in_R I \setminus \{i\} : w_{i,j} > 0 \tag{6}$$

The opinion of i, $O_{i,t}$, is then updated if the opinion of j, $O_{j,t}$ lies within a confidence bound ε Deffuant et al., 2001:

$$O_{i,t} = O_{j,t-1} \times w_{i,j} \qquad \text{where } |O_{i,t-1} - O_{j,t-1}| \overset{!}{<} \varepsilon \tag{7}$$

3 Reducing Polarization

In the following I will compile a list of different approaches on actions planners can take to reduce polarization and filter bubbles in social networks. Planners in this context are parties that have a top-down power over the network, meaning they have the power to unresentedly change the structure of the network. This approach of course is highly theoretical and will most likely not be able to be implemented in real social networks, as these structural changes cannot be forced by administrators but need to be freewillingly accepted by the users.

In pursuit of minimizing polarization we firstly need to define what polarization actually means. Polarization can be understood as the variance of expressed opinions in a network and will be defined as [Racz and Rigobon, 2022]:

$$P(X) := \sum_{i=1}^{n} (x_i - \bar{x})^2 = ||\widetilde{x}||^2 \qquad \text{where } \bar{x} = \frac{\sum_{i=1}^{n} x_i}{n} \tag{8}$$

In this context it another important metric is the disagreement between two individuals i and j, which is the distance of their opinions from consensus [Racz and Rigobon, 2022]:

$$D_{i,j} = (x_i - x_j)^2 \tag{9}$$

3.1 Increasing edge weight

A way for planners to modify the network structure is increasing the weight of existing edges in a given network. As this means, from a theoretical point, increasing the interaction between two selected individuals, this may correspond to recommender algorithms promoting content of two individuals for each other in user-centric feed social networks or show increased amount of posts to a sub-forum in forum social networks (defined in 1.1).

Racz and Rigobon proposed three heuristics for selecting a non existing edge from the complementary edge-matrix E^C, seeking to minimize the present polarization in the network. They gathered their findings using the Friedkin-Johnsen model (2.2) and applied it to both real-world (Twitter, Reddit & Blogs) and artificially constructed networks. [Racz and Rigobon, 2022]

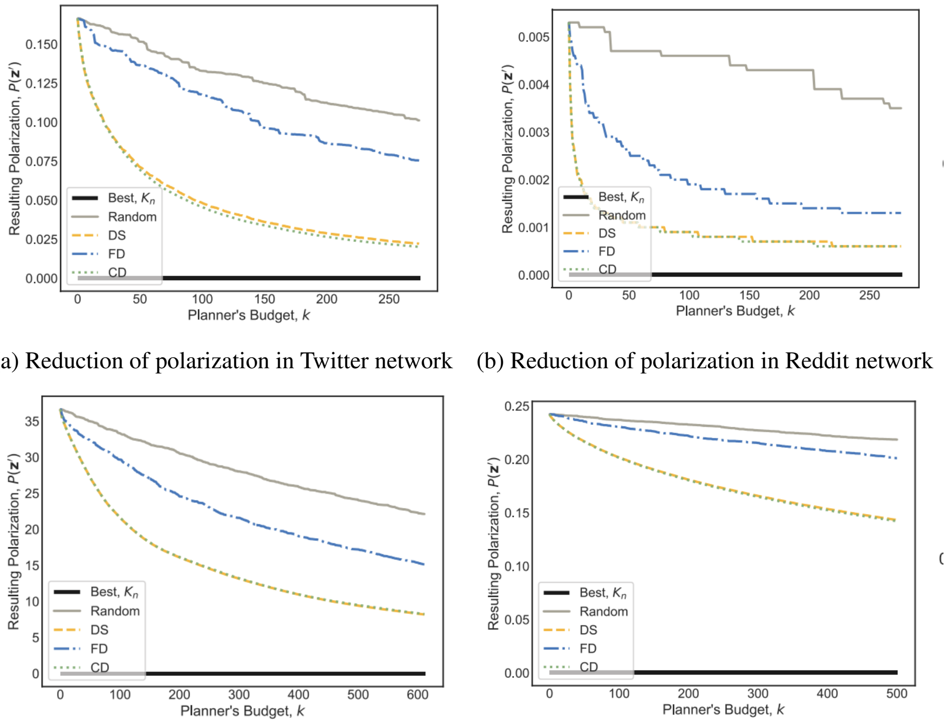
3.1.1 Promoting Same Neighborhoods

Interestingly and against the notion of avoiding filter bubbles, Racz and Rigobon have found that when two selected nodes i and j and the grapgh $\mathcal{G}$ satisfy $N_{\mathcal{G}}(i) = N_{\mathcal{G}}(j)$, i.e. i and j having the same neighborhoods, adding to the weight between these nodes $w_{i,j}$ will decrease polarization. The reason for this is that only the opinions of the two selected nodes i and j are affected by this change and no global effect is produced. As the authors also state, this change has miniscule effect on the polarization of the entire network and will thus not be considered an actual solution to reducing polarization. [Racz and Rigobon, 2022]

3.1.2 Disagreement Seeking (DS)

Another approach porposed by Racz and Rigobon was the method of DS. For this heuristic a planner will search the two vertices with the most Disagreement and the biggest distance from actual weight and maximum weight in the graph:

$$argmax_{(i,j) \in E^C} (\bar{w} - w_{i,j})(z_i - z_j)^2 \tag{10}$$



(a) Reduction of polarization in Twitter network (b) Reduction of polarization in Reddit network

(c) Reduction of polarization in Reddit network (d) Reduction of polarization in Reddit network

Figure 1: Effectiveness of different heuristics [Racz and Rigobon, 2022]

3.1.3 Coordinate Descent (CD)

This heuristic is derived from calculating the derivative of the polarization over a weight between two edges $-\delta\partial w_{i,j}P$. From this can be derived that it is optimal to choose direction of the steepest descent of the polarization, which gives this heuristic its name. The resulting heuristic can be denoted as:

$$argmax_{(i,j)\in E^C} - (\bar{w} - w_{i,j})\partial w_{i,j}P(z) \tag{11}$$

This heuristic suggests that a planner should increase weights between individuals that are closest to the maximum allowed weight between two nodes.

3.1.4 Fiedler Difference (FD)

This heuristic utilizes a value called the Fiedler value or Fiedler vector v, which is defined by fulfilling the equation $\lambda_2 v = Lv$. The intricacies of the Fielder vector and the Laplacian matrix L do not need to be understood to understand this equation. The important thing is the meaning of the magnitude of v, $|v_i - v_j|$. As the Fielder vector describes how interconnectedness of a graph, the magnitude of $v_i - v_j$ describes how separated the nodes i and j are in the graph. This leads to the notion that a planner should increase weights between two nodes that are far apart in the graph or belong to different *parts* of the graph.

$$argmax_{(i,j)\in E^C}(\bar{w} - w_{i,j})|v_i - v_j| \qquad \text{where } \lambda_2 v = Lv \tag{12}$$

In Figure 1 you can see that the DS algorithm in every case yields the best performance. For simulation Racz and Rigobon used a greedy algorithm, i.e. an algorithm that calculates the single best edge at a time and adds that to the graph. The algorithm was also designed to constraining the amount of edges that can be added by k (the x-axis of the plots). The Twitter network was constructed by taking data from people who tweeted about a Delhi assembly debate in 2013. The Reddit network was constructed using data from individuals who posted in a politics sub-forum on the platform. The

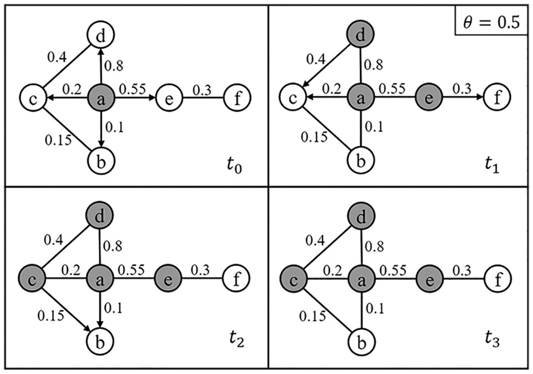


Figure 2: Depiction of the Linear Threshold model

blogs network is different than the last two as its' nodes do represent individuals but blog webpages about the 2004 US election, where every block was either classified as 'conservative', 0, or 'liberal', 1. This also explains the high polarization compared to the other networks.

3.1.5 Considering opinion confidence

Wu et al. proposed a custom model, called the Depolarization Model (DM), based on the Linear Threshold (LT) and the Hegelsmann-Krause (HK) model. The linear transformation model, opposed to the other models presented above, differentiates between active and inactive nodes at time t. In this model a subset $I_{active} \subset I$ is constituted and based on a confidence bound ε, similar to HK it can activate connected nodes in a state transition.

The DM additionally includes the notion of a self-belief and opinion influence parameter, $\alpha \in [0; 1]$ & $\mu \in [0; 1]$ resp., which take the place of ε in LT and control how active nodes can activate other nodes in state transitions. The two parameters interplay in a way such that a high self belief can only be 'persuaded' of other opinions if the individual holding this opinion has a high opinion influence value.

Using this model Wu et al. proposed three strategies of dealing with polarization in this model. For this they divided the set of individuals I into three subsets according to their self-belief α. The groups reflect individuals who are either *open-minded*, *moderate* or *stubborn*, also named as low, medium and high temperature:

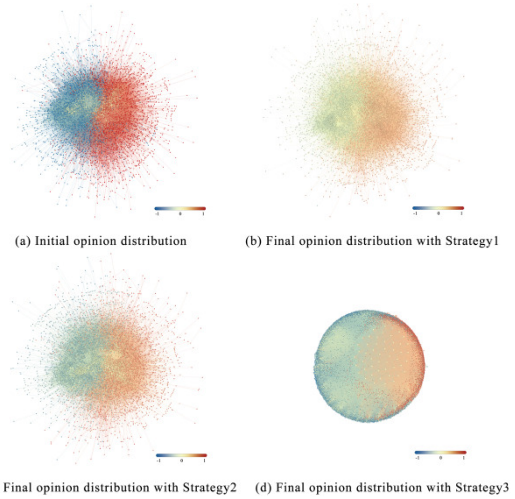
$$\begin{aligned} I_{low} &= i \in I : \alpha_i \in [0; 0.35] \\ I_{medium} &= i \in I : \alpha_i \in (0.35; 0.7] \\ I_{high} &= i \in I : \alpha_i \in (0.7; 1] \end{aligned} \tag{13}$$

Open-minded individuals in the network, as they are most capable of accepting differing opinions, they should be used to act as moderators between heterogenous groups in the graph. This suggests, that adding edges between the most open-minded individuals in these heterogenous groups will reduce polarization through the propagation of the differing opinions from other groups inside a group itself, event to medium and high temperature individuals.

Moderate individuals, as they can only accept moderately differing opinions, can help to reduce polarization by strongly connecting them to an individual with a neutral opinion.

Stubborn individuals can only accept differing opinions of highly social influential individuals. For that it is most productive if new connections between them and highly interconnected (i.e. high μ) individuals of neutral opinion are introduced.

In the experiments conducted both α and μ are set globally, i.e. every individual posesses the same value for these parameters: $\forall i \in I : \alpha_i = \alpha \land \mu_i = \mu$.



(a) Initial opinion distribution

(b) Final opinion distribution with Strategy1

(c) Final opinion distribution with Strategy2

(d) Final opinion distribution with Strategy3

Figure 3: Visualization of opinion distribution with different strategies

4 Results/Discussion

The models and strategies presented in this paper aim to understand and potentially mitigate polarization in social networks. The results from various approaches indicate that increasing connectivity and strategically altering network structure can reduce polarization to some extent. This discussion will delve into the key insights, critical analysis, and broader implications derived from the research.

Key Results and Interpretation:

The experiments highlight the role of edge weight in influencing network polarization. Among the heuristics explored, the Disagreement Seeking (DS) approach, which targets nodes with significant opinion differences, showed to have the strongest effect in reducing polarization. This aligns with the intuitive understanding that connecting individuals of differing opinions will reduce polarization in the network.

Another approach, the Coordinate Descent (CD), also demonstrated efficacy in reducing polarization by adjusting edge weights in the direction that most reduces variance. This heuristic, while technically grounded, requires more computational effort, given its dependence on network structure analysis.

The Fiedler Difference (FD) heuristic, which connects distant nodes in the network, showed moderate success. This heuristic, rooted in algebraic graph theory, emphasizes the importance of connecting isolated clusters to foster cohesion within the network.

The results of these heuristics suggest that polarization can be minimized by enhancing cross-cluster connections, thereby promoting a more cohesive network structure. This finding has implications for social media platforms that wish to encourage diverse interactions among users.

Assumptions and Limitations:

The analysis assumes that increasing connectivity in the network leads to reduced polarization. While the results support this assumption, it does not account for the complexities of real-world social networks, such as user behavior, algorithmic bias, or external societal influences. Moreover, the heuristics focus on structural changes, but do not consider the content or context of interactions. This limitation implies that while network structure can influence polarization, other factors might play a significant role.

Another limitation is the assumption that nodes (representing individuals) can be connected without resistance or unintended consequences. In real-world social networks, users might resist connections that contradict their beliefs, leading to potential backfire effects. This resistance underscores the need for a nuanced approach when applying these findings to real-world platforms.

Comparisons with Other Techniques:

The strategies for reducing polarization, especially those focusing on increasing edge weights, align with existing literature that advocates for diverse interactions to combat echo chambers and filter bubbles. However, this approach differs from more content-focused techniques, such as promoting balanced content or curating news feeds. The network-based approach offers a structural perspective but should ideally complement other methods to achieve a comprehensive solution to polarization.

Advantages and Disadvantages:

The advantage of the proposed heuristics lies in their simplicity and adaptability to various network structures. They offer a scalable approach to reducing polarization without requiring significant alterations to existing platforms. However, the disadvantage is that they may overlook user behavior dynamics and the potential for unintended consequences. Additionally, the reliance on network structure alone may not address deeper social issues contributing to polarization.

Implications for Social Network Platforms:

For social network platforms, these results suggest that fostering diverse connections can reduce polarization. However, implementing these strategies requires careful consideration of user privacy, autonomy, and platform algorithms. Social network operators should approach these changes with sensitivity to avoid user backlash or unintended outcomes.

Overall, while the heuristics provide a promising approach to mitigating polarization, they should be part of a broader strategy that considers user behavior, platform design, and societal context. Further research is needed to explore these dynamics and develop comprehensive solutions to the complex issue of polarization in social networks.

5 Conclusion

The study of opinion formation and polarization in social networks has shown several promising strategies for mitigating the effects of polarization. Through the exploration of various models, including the Friedkin-Johnsen model, the Hegelsmann-Krause model, and the Depolarization Model, we've identified methods that focus on increasing edge weights, bridging opinion gaps, and fostering more interconnected networks to reduce polarization.

The Disagreement Seeking (DS) heuristic emerged as a particularly effective strategy, targeting nodes with the most significant opinion differences. This approach encourages interaction across polarized groups, promoting a more cohesive network. However, implementing these strategies in real-world social networks requires careful consideration of user behavior, resistance to unwanted connections, and broader societal influences.

The key takeaway from this paper is that polarization can be reduced by encouraging diverse interactions and breaking down echo chambers in social networks. However, achieving this requires a nuanced approach, considering both network structure and the context of interactions. Social network operators must balance these structural changes with user autonomy and privacy, ensuring that any changes do not result in adverse effects.

It has to be noted though that the research shown in this paper focuses the studies on social networks regarding a single topic of interest and the opinions expressed and the polarization that is aimed to be reduced in the model are only applicable to this topic. It is possible that the alterations made in the network which reduce polarization for the topic of interest might cause a surge of polarization in other topics.

Future research should focus on integrating these structural approaches with content-based strategies to address polarization comprehensively. Moreover, studies should consider user behavior dynamics and explore methods to overcome resistance to diverse interactions. By combining structural changes with a deeper understanding of social dynamics, we can develop more effective solutions to reduce polarization in social networks.

References

Guillaume Deffuant, D. Neau, Frédéric Amblard, and Gérard Weisbuch. Mixing beliefs among interacting agents. *Advances in Complex Systems*, 3:87–98, January 2001.

Morris H. DeGroot. Reaching a Consensus. *Journal of the American Statistical Association*, 69 (345):118–121, 1974. ISSN 0162-1459. doi: 10.2307/2285509. URL https://www.jstor.org/stable/2285509. Publisher: [American Statistical Association, Taylor & Francis, Ltd.].

Noah Friedkin and Eugene Johnsen. Social Influence and Opinions. *Journal of Mathematical Sociology - J MATH SOCIOL*, 15:193–206, January 1990. doi: 10.1080/0022250X.1990.9990069.

Rainer Hegselmann and Ulrich Krause. Opinion Dynamics and Bounded Confidence: Models, Analysis and Simulation. *Journal of Artificial Societies and Social Simulation*, 5(3), 2002. URL https://philarchive.org/rec/RAIODA.

Stefan Neumann, Yinhao Dong, and Pan Peng. Sublinear-Time Opinion Estimation in the Friedkin–Johnsen Model, April 2024. URL http://arxiv.org/abs/2404.16464. arXiv:2404.16464 [cs].

Pew Research Center. Social Media and News Fact Sheet, November 2023. URL https://www.pewresearch.org/journalism/fact-sheet/social-media-and-news-fact-sheet/.

Rüdiger F. Pohl. *Cognitive Illusions: A Handbook on Fallacies and Biases in Thinking, Judgement and Memory*. Psychology Press, December 2012. ISBN 978-1-135-84495-0. Google-Books-ID: MS5Fr8safgEC.

Miklos Z. Racz and Daniel E. Rigobon. Towards Consensus: Reducing Polarization by Perturbing Social Networks, December 2022. URL http://arxiv.org/abs/2206.08996. arXiv:2206.08996 [cs].

Aaron Shaw. Social media, extremism, and radicalization. *Science Advances*, 9(35):eadk2031, August 2023. ISSN 2375-2548. doi: 10.1126/sciadv.adk2031. URL https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10468141/.

Petter Törnberg. How digital media drive affective polarization through partisan sorting. *Proceedings of the National Academy of Sciences*, 119(42):e2207159119, October 2022. doi: 10.1073/pnas.2207159119. URL https://www.pnas.org/doi/10.1073/pnas.2207159119. Publisher: Proceedings of the National Academy of Sciences.

E. S. Volkova, L. A. Manita, and A. D. Manita. Hegselmann-Krause model of opinions dynamics of interacting agents with the random noises. *Journal of Physics: Conference Series*, 1163 (1):012064, February 2019. ISSN 1742-6596. doi: 10.1088/1742-6596/1163/1/012064. URL https://dx.doi.org/10.1088/1742-6596/1163/1/012064. Publisher: IOP Publishing.

Yue Wu, Linjiao Li, Qiannan Yu, Jiaxin Gan, and Yi Zhang. Strategies for reducing polarization in social networks. *Chaos, Solitons & Fractals*, 167:113095, February 2023. ISSN 0960-0779. doi: 10.1016/j.chaos.2022.113095. URL https://www.sciencedirect.com/science/article/pii/S0960077922012747.