Developing a Social Media Algorithm that Focuses on Providing Users with a Balanced Information Diet
(Technical Topic)

Evaluating the Merits and Drawbacks of External Review of Social Media Companies
(STS Topic)

A Thesis Project Prospectus
In STS 4500
Presented to
The Faculty of the
School of Engineering and Applied Science
University of Virginia
In Partial Fulfillment of the Requirements for the Degree
Bachelor of Science in Computer Science

By
Alby Alex

May 1, 2024

On my honor as a University student, I have neither given nor received unauthorized aid on this assignment as defined by the Honor Guidelines for Thesis-Related Assignments.

Signed: Alby Alex

Advisors

Kathryn A. Neeley, Department of Engineering and Society

In the past two decades, the development of social media has led it to become the primary news source for a significant number of people. As the number of potential news sources and reported events increased, people began experiencing an "information overload" (Pentina et al., 2014, 221). There was no way to make sense of what information was important, which led people to use their algorithm-formed social media feeds to curate their news consumption. As a result, social media algorithms and their creators were given significant societal influence. Unfortunately, as Pentina (2014) states, the information provided by these algorithms "tends to be filtered through the attitudinal preferences of [the user's social media] network and may not provide an encompassing, balanced or diverse knowledge about civic society" (221). This lack of variation in their information diet leads people to develop narrow worldviews and become politically polarized over time.

As Garcia (2023) states, "voters are polarized in many countries to the point at which they feel greater aversion to people with different political views than affection for like-minded people" (39). This polarization may not solely come from the direct use of social media, but also from second order effects, such as increased political participation inspired by it (Lee et al., 2017). Higher levels of political polarization from voters will lead to the election of more partisan representatives to government, which will lead to legislative gridlock. Congress will not be able to compromise to address important issues (McCarty, 2007) and this polarization may result in further consequences "for the effective function of democracy" (Bail et al., 2018). As such, this polarization from social media must be solved, and the circumstances that led to it must be properly investigated. To achieve this goal, my technical work will aim to develop an algorithm for social media platforms that works to convert users to a balanced information diet,

while my STS research will focus on the concept of external review and its place in the analysis of social media platforms.

Technical Topic: Designing a Social Media Algorithm that Prioritizes a Balanced Information Diet

Political polarization on social media is mainly a consequence of how platform algorithms are designed to maximize engagement, as described by Kim (2017). This algorithm design reflects how social media companies make money. Social media companies offer their services for free, instead opting to serve ads to users to earn revenue from advertisers. Advertisers want to maximize the "impressions", or individual exposure events, that their ads receive, and they are willing to pay more for it. As Kim (2017) states, "social media platforms have an incentive to keep users engaged so they can profit—in some instances by the billions—from hosting ads" (148). They accomplish this through the analysis of large amounts of data that they collect from users during their platform use. During this analysis, the data is sorted into different categories that reveal what the user engages with. This engagement can include active activities like "liking" or "sharing", but it also includes more passive actions like the amount of time that a user spends looking at a post. As explained by Kim (2017), "the more frequent the engagement, the stronger the association the algorithm will make between the user and that content" (149).

Once the platform finishes creating a somewhat comprehensive profile of all the data it has on a particular user, Kim (2017) states that it forms a pool of content that might be potentially interesting to the user. This pool is generated using natural language processing (NLP) algorithms that can determine the meaning behind a certain post and categorize it. In addition, the pool also contains items that friends of the user have engaged with previously. Once

this pool is created, the algorithm goes through and ranks each item based on how likely the user is to interact with it. Finally, it then serves the content to the user in the order of these ranks. By itself, this process, as depicted in Figure 1, does not seem to cause political polarization. However, most Americans visit a social media site daily, with an increasing number choosing to use it as their primary source of news (Bail et al., 2018). With the algorithms' focus on showing users content that they will like, Garcia (2023) states that "a common concern is that news-feed algorithms on social media could have a role in generating polarization by exposing individuals to more information from politically like-minded sources than they would otherwise see, thus creating 'echo chambers' and 'filter bubbles'" (39). This increased exposure to like-minded media leads people to become more politically polarized (Kubin et. al, 2021).

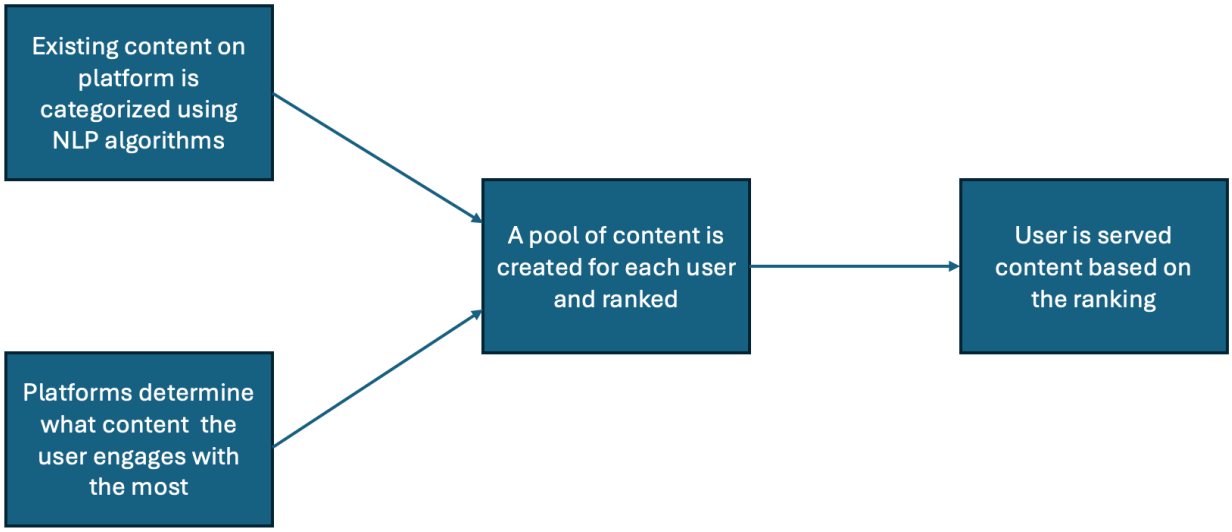

Figure 1. **Social Media Algorithm Process:** This flowchart depicts the steps undertaken by social media algorithms to serve content to users (Created by Author).

Previous efforts to counteract this effect have focused on either creating a chronological feed that tries to be non-partisan (Guess et al., 2023) or a hyper-partisan feed that always provides content with opposing views to the user (Bail et al., 2018). Unfortunately, both strategies have proven to be ineffective at reducing the political polarization of users. In fact, the

results of the second study, as shown in Figure 2, found that being constantly exposed to opposing views made users become more polarized in favor of their original view. These results indicate that a sudden shift in how content is served, regardless of how well-intentioned it is, will not address the problem of polarization. It also shows that a non-curated, chronological feed will also fail at achieving this goal. To address this, my algorithm will focus on producing a gradual transition from the current partisan feed to something that is more balanced.

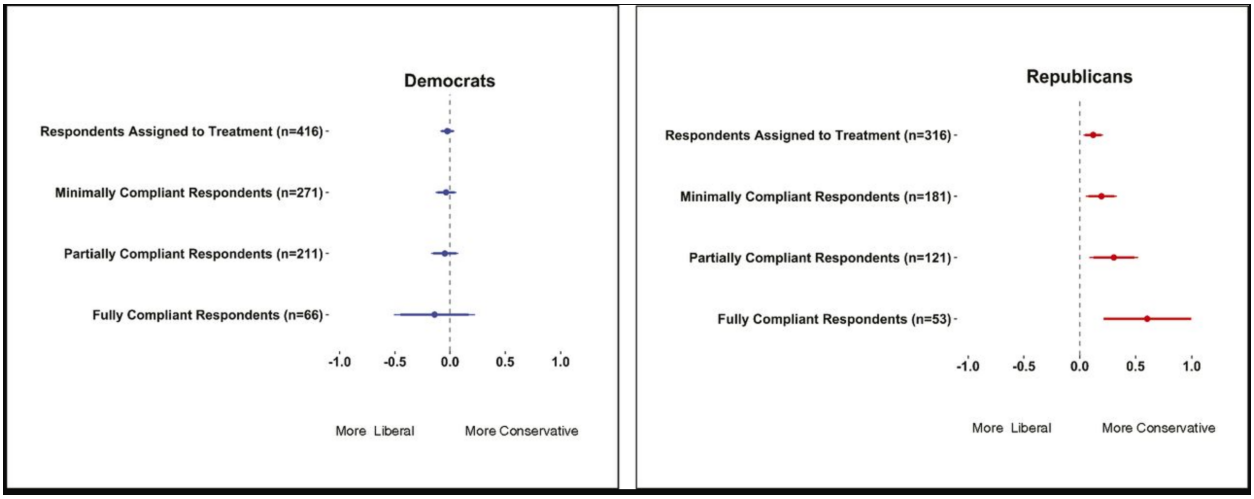

Figure 2. **Results of Potential Depolarization Treatment:** This image depicts the results of a study that only served partisan users content that disagreed with their views. This caused an increase in partisanship among fully compliant participants (Adapted from Bail et al., 2018, 9220).

My algorithm would still produce the user profiles that platforms are currently creating, but it will pay specific attention to any political activity by the user. Once the user's current political leaning is determined, the algorithm will serve the user one piece of content that disagrees with their current views. If the user interacts with this post, then the next feed refresh will include more of this content. This process will continue until the feed reaches a predefined threshold for information balance. I hope to accomplish this by repurposing some of the same NLP services that platforms use to profile content and users. It takes time to change user beliefs and views, so this gradual approach will allow for the time that people need to change. Any new users that join platforms with this new algorithm will have a balanced feed from the start, leading

them to avoid the polarization problems altogether. While user engagement with the platforms may fall, this trade-off is worthwhile to ensure a well-informed population that wants to work with fellow citizens instead of arguing with them in perpetuity.

STS Topic: External Review and its Potential Effects on Social Media Platforms

As mentioned before, the attention-maximizing function of social media resulted in negative consequences like political polarization because companies did not properly consider downstream effects in their pursuit of profit. According to Andrejevic (2013), "platforms like Twitter and YouTube are proprietary, and can only be organised by those who control the platform" (129). As mentioned before, a growing number of Americans acquire their news from social media (Bail et al., 2018) and use social media as a way of curating the news media that they consume (Pentina et al., 2014), meaning that social media companies have the sole power to control how users are shown information. Any small changes they make to their algorithms directly affect how millions of people perceive the world. Unfortunately, it is extremely difficult to analyze the effects of any of these algorithm changes without express permission from the subject companies, as evidenced by a meta-analysis of social media studies on polarization that concluded that Twitter was overrepresented in studies due to the ease with which researchers can scrape information from the platform without cooperation (Kubin et al., 2021).

However, Meta, the owner of two of the largest social media platforms, recently partnered with researchers to better understand some of the social effects of their algorithms. These researchers were able to directly alter algorithm parameters and measure the effect of the changes on user behavior and political polarization (Guess et al., 2023) (Garcia, 2023). Before, this type of testing would occur internally within the company and the results would never be released. Instead, the access provided by Meta enabled independent researchers to conduct

proper research and publish their results to the public. As a result, the public was able to gain a stronger understanding of the societal effects of the algorithms. This adds to the uncertainty about whether all social media platforms should be subjected to this type of external review.

Historically, the process of peer review, as described by Gustafson (1990), has had significance in other engineering disciplines like civil engineering, where the community adopted these processes due to safety incidents such as a 1981 Kansas City walkway collapse and a 1987 collapsing slab in Bridgeport. These incidents led to more than 100 deaths, and the civil engineering community convened in multiple conferences to determine a potential solution to "improve the quality of the constructed project" (350). They came up with the concept of project peer review, which can be defined as "an additional separate step in the design process that is intended to provide an evaluation of the concept of design" (351). The American Society for Civil Engineers Task Committee on Project Peer Review stated that projects that affect either public health and safety or national defense should be reviewed. In addition, projects with a unique design or construction process should also be reviewed (351-352).

The creation of these review processes has resulted in better outcomes for civil projects. In one instance, the project peer review of a dam in the late 1970s led to numerous corrections of design inconsistencies and the usage of alternative construction methods that saved time and money on the project (Gustafson, 1990, 353). After civil engineers experienced multiple disasters, they came together to establish peer review to increase the number of people involved with designing and building a large public work. Social media platforms may benefit from a similar development.

My STS research will focus on understanding obstacles to the implementation of a peer review system for social media platforms and their algorithms. Figure 3 depicts the general

process that I will utilize. Civil engineers build projects that affect significantly less people than social media, yet the external review process for their work is much more robust than anything that exists for software. This system has quickly gained widespread acceptance and support in the civil engineering community due to its effectiveness. In fact, one of the ASCE's conclusions in the wake of the Katrina disaster, as stated by Anderson et al. (2006), was that "many of the major deficiencies in New Orleans's hurricane protection system could have been avoided if the engineering plans and designs had undergone high-level, independent engineering review by external experts" (10). Through careful analysis of social media platforms using Actor Network Theory, I can understand all the parties that would be affected by the establishment of this type of peer review network. I can then apply the findings from various STS studies on peer review in civil engineering to predict the potential positive and negative effects this would have on social media algorithms, companies, and users.

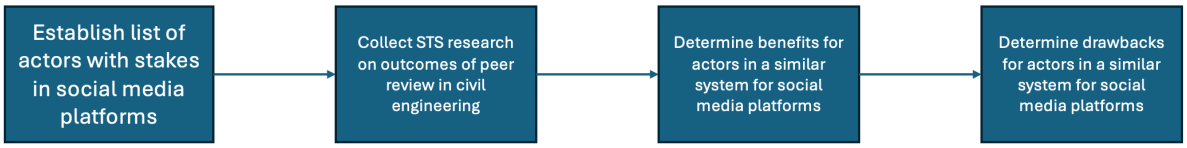

Figure 3. **STS Research Process**: This flowchart shows the general process that the author will utilize in their STS research (Created by Author).

Conclusion

Through my technical work, I hope to develop an algorithm for social media that prioritizes providing a balanced information diet for social media users over maximizing engagement. My STS research focuses on the potential outcomes of implementing a civil engineering style peer review system for changes in social media algorithms. If the technical work is successful, then social media platforms will have the chance to reduce political polarization in the world with minimal expended effort, reducing the existential threat that it

poses to democracy. If the STS research is successful, new guidelines can be created to establish a new peer review system for social media companies. This system will allow for a greater societal understanding of social media and help avoid any negative second-order consequences from potential algorithm changes. Social media platforms today have tremendous influence on society, emphasizing the need for a comprehensive understanding of the extent of this influence and how negative consequences can be mitigated. (2038 words)

References

Andersen, C. F. et al. (2006). Hurricane Katrina: One year later: What must we do next? A statement by the American Society of Civil Engineering Hurricane Katrina External Review Panel. Reston, VA: American Society of Civil Engineers.

Andrejevic, M. (2013). Public service media utilities: Rethinking search engines and social networking as public goods. *Media International Australia*, 146(1), 123–132. https://doi.org/10.1177/1329878x1314600116

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. B., Lee, J., Mann, M., Merhout, F., & Volfovsky, A. (2018). Exposure to opposing views on social media can increase political polarization. *Proceedings of the National Academy of Sciences*, 115(37), 9216–9221. https://doi.org/10.1073/pnas.1804840115

Garcia, D. (2023a). Influence of facebook algorithms on political polarization tested. *Nature*, 620(7972), 39–41. https://doi.org/10.1038/d41586-023-02325-x

Guess, A. M., Malhotra, N., Pan, J., Barberá, P., Allcott, H., Brown, T., Crespo-Tenorio, A., Dimmery, D., Freelon, D., Gentzkow, M., González-Bailón, S., Kennedy, E., Kim, Y. M., Lazer, D., Moehler, D., Nyhan, B., Rivera, C. V., Settle, J., Thomas, D. R., … Tucker, J. A. (2023). How do social media feed algorithms affect attitudes and behavior in an election campaign? *Science*, 381(6656), 398–404. https://doi.org/10.1126/science.abp9364

Gustafson, G. D. (1990). Why project peer review? *Journal of Management in Engineering*, *6*(3), 350–354. https://doi.org/10.1061/(asce)9742-597x(1990)6:3(350)

Kim, Sang Ah. (2017). Social media algorithms: Why you see what you see. *Georgetown Law Technology Review*, 2(1), 147-154.

Kubin, E., & von Sikorski, C. (2021). The role of (social) media in political polarization: A systematic review. *Annals of the International Communication Association*, 45(3), 188–206. https://doi.org/10.1080/23808985.2021.1976070

Lee, C., Shin, J., & Hong, A. (2018). Does social media use really make people politically polarized? Direct and indirect effects of social media use on political polarization in South Korea. *Telematics and Informatics*, 35(1), 245–254. https://doi.org/10.1016/j.tele.2017.11.005

McCarty, N. (2007). The policy effects of political polarization. In P. Pierson & T. Skocpol (Eds.), *The transformation of American politics: Activist government and the rise of conservatism* (pp. 223–255). Princeton, NJ. Princeton University Press. http://www.jstor.org/stable/j.ctt7rwcc.14

Pentina, I., & Tarafdar, M. (2014). From “information” to “knowing”: Exploring the role of social media in contemporary news consumption. *Computers in Human Behavior*, 35, 211–223. https://doi.org/10.1016/j.chb.2014.02.045