iS
CHANNEL

Social Media Personalization Algorithms and the Emergence of Filter Bubbles

Sanveer (Sunny) Rehani

MSc Management of Information Systems and Digital Innovation
Department of Management
London School of Economics and Political Science

KEYWORDS

Social media
Content personali-zation algorithms
Filter bubbles
Opinion polarization
Preference data
Platform governance

ABSTRACT

This article investigates how the technical infrastructure of social media platforms, particularly personalization algorithms, enables filter bubbles to emerge. It identifies several possible explanations, namely that platform owners intend this outcome; that the relevant algorithms contain inherent biases; and that user actions amplify the effect of these algorithms. While the true answer cannot be ascertained, the article concludes by outlining important areas for future research and highlighting the implications of social media content personalization.

I. Motivation

Content personalization on social media platforms is widely perceived as a beneficial feature, enabling each of us to have a unique and individualized online experience. What is seldom discussed is the hidden danger this carries: the continuous reinforcement of users' own viewpoints and beliefs. Without any exposure to content that chal-lenges existing viewpoints, it is easy for social media users to obtain a distorted image of the real world due to the limited scope of information that they encounter. This phenomenon, dubbed 'filter bubbles', could have a serious effect on our ability as individuals to form balanced opinions, and could negatively alter the way we react to opinions that contradict our own.

What makes these filter bubbles especially dangerous is our ignorance of their exist-ence. Social media platforms, in an effort to provide users with the content most rele-vant for them, conduct this personalization automatically without informing users. A concept known as "what you see is all there is" (Kahneman, 2011, p. 85) describes a cognitive bias faced by humans whereby we treat the information available to us as if it were the only information to exist, thus making judgements and decisions based solely on that information. When filter bubbles cause all the content that we see to align with our existing ideologies, it becomes evident that this cognitive bias could lead to false illusions of reality.

It is critical that we address the issue of filter bubbles sooner rather than later. As social media platforms are increasingly utilized as sources of news and information, and as data algorithms become more advanced, it seems inevitable that this problem will only worsen. If left unchecked, filter bubbles could propagate misinformation among online communities, therefore serving as a serious threat to democracy and contrib-uting further to a global society that is arguably more polarized and divided than ever before.

Corresponding Author
Email Address: sunny.rehani@gmail.com

II. Literature Review

This section distinguishes between two distinct streams of literature; one that focuses on the outcomes of content personalization, and another that focuses on the tech-nical design of the algorithms that enable content personalization. Together, these two streams comprise the context within which this paper's research is situated.

a. Research on consequences

There exists a vast amount of literature concerning the consequences of social media content personalization. Cass Sunstein pointed out the dangers of personalization as early as 2001, years before the ascent of social media, stating that the internet could "increase people's ability to wall themselves off from topics and opinions that they would prefer to avoid" (p. 202). In more recent literature, a recurring conclusion is that social media content personalization often results in ideological segregation to some degree.

A study by Dylko et al. (2018) demonstrates that the customizability technology em-ployed by popular social media platforms leads to increased selective exposure to at-titude-consistent content, thereby causing increased political polarization. Bakshy et al. (2015), in a study of over 10 million US Facebook users, discovered that "the risk ratio comparing the probability of seeing cross-cutting content relative to ideologically consistent content [on a user's 'News Feed'] is 5% for conservatives and 8% for liber-als" (p. 1131). The considerable majority of information that social media users are exposed to, then, is aligned with their existing beliefs. Additionally, Knobloch-West-erwick et al. (2015) show that source credibility does not influence the impact of po-litical information on political attitude. This is especially noteworthy when discussing social media, as users share not only credible news articles but also opinion pieces. Flaxman et al. (2013) find that articles shared on social media are more ideologically segregating than news consumed directly from news sites, mainly due to opinion pieces.

A key limitation of this set of research is the difficulty and potential subjectivity in-volved in measuring the ideological slant of news sources or pieces of information. However, this weakness is overcome by the frequent recurrence of similar conclusions across the research field.

b. Research on technical elements

There is also an ample amount of literature concerning the technical elements of con-tent personalization. Often this research focuses on the recommender systems of e-commerce sites rather than social media platforms, but as these systems are similar in design, the literature is still useful in this context. Adomavicius and Tuzhilin (2005) classify recommender systems into three categories: content-based recommenders, which utilize users' past behavior to make recommendations; collaborative filtering recommenders, which make recommendations based on the behavior of other users with similar preferences; and hybrid systems, which combine these approaches. Col-laborative filtering is more widespread today, but social media platforms likely employ hybrid approaches to personalize content for users.

This personalization is most commonly enabled by 'item-item algorithms', which de-termine the 'distance' between items based on "how closely users who have rated these items agree" (Alaimo and Kallinikos, 2019, para. 8). This means that items are grouped together into neighborhoods when the same users tend to like or dislike them. The algorithm can then recommend items to users according to other items in the same neighborhood for which they have previously demonstrated a preference. In the context of social media, these items would largely consist of online content (e.g. articles) shared by users; for example, if the Facebook users who tend to 'like' Fox News articles shared on their feed also tend to 'like' Breitbart News articles shared on their feed, then Fox News articles and Breitbart News articles might be grouped into the same neighborhood.

Of course, the act of 'liking' is simply one form of preference data that can be fed into personalization algorithms. Alaimo and Kallinikos (2019) explain that explicit data is collected through "actions that can be straightforwardly linked to preferences" (para. 7), such as the aforementioned Facebook 'like'. Ekstrand et al. (2011) add that pref-erence data can also be collected implicitly from user behavior, e.g. by monitoring clicks, time spent on a page, and so on. Social media platforms presumably collect both kinds of preference data. However, implicit data contains more noise, i.e. mean-ingless information, than explicit data (O'Mahony et al., 2006), which may limit its accuracy in predicting user preferences.

This set of research is somewhat limited by its heavy focus on e-commerce, and rela-tive disregard for social media. While the systems are likely quite similar in any case, it is surprising that there is relatively little research on the technical elements of social media content personalization in particular, given the massive relevance of social me-dia in the modern world. Additionally, much of the research is rather outdated when considering how rapidly these algorithms are being refined and reworked.

c. The gap

Although these are two rich and plentiful streams of literature, it seems that there exists somewhat of a gap between them. There is much discussion on the outcomes of con-tent personalization, as well as on the design of personalization algorithms - but very little on how exactly the latter leads to the former. The aim of this paper is therefore to serve as the bridge between these two fields. This paper will seek to answer the question: *how does the technical infrastructure of social media platforms en-able the emergence of filter bubbles?*

III. Problem Analysis

The existing literature explored in the previous section gives us a good baseline from which we can begin to answer this question. Firstly, it will be helpful to examine why personalization algorithms are implemented into the infrastructure of social media platforms in the first place. It is important to note that today's social media monoliths had not employed any such algorithms in their early days. In fact, user feeds tended to be sorted chronologically. What this provided, from a user's perspective, was a clear finish line for browsing; users could scroll through the new content that had been posted since they were last online, and then stop. However, as these platforms rapidly became more popular, the rate at which digital information was being produced and shared grew exponentially, creating a problem of information overload (Bozdag, 2013).

Information overload causes stress, confusion, and cognitive strain (Eppler and Men-gis, 2004), and can even result in 'social media fatigue' (Bright et al., 2015). Platforms like Facebook, Twitter, and Instagram needed a way to combat this issue in order to keep their massive userbases engaged. Personalization algorithms were a solution that enabled the filtering of information to present only the content relevant to each user. This seems logical, and arguably even necessary, but still begs the question: how and why do these algorithms ultimately equate 'relevant' content to attitude-consistent con-tent?

We identify two possible explanations, both of which will be explored in the following subsections. The first is that social media platform owners are incentivized to inten-tionally program the algorithms in this way. The second – slightly more complex – explanation is that although the outcome is unintentional, it is enabled by inherent biases in the algorithms' design, as well as by the actions of users themselves.

a. Intended design?

One rationalization for the emergence of filter bubbles is the deliberate programming of personalization algorithms to predominantly present content that users are likely to agree with. Research has shown that disagreement in an online setting leads to negative emotion and aggression (Masullo Chen and Lu, 2017). It is safe to assume that plat-form owners, i.e. administrators, wish to avoid inciting these sentiments among their userbase due to the risk of losing user engagement. By extension, then, administrators are incentivized to minimize the level of disagreement occurring on their platform (Chitra and Musco, 2020). Cross-cutting content is naturally a major source of disa-greement and conflict and is thus undesirable to administrators. Therefore, administrators may purposefully implement algorithms designed to present less cross-cutting content and more attitude-consistent content.

Chitra and Musco (2020), in a study of Twitter and Reddit, demonstrate that when a network administrator is able to actively filter social content in an effort to present users with content that matches their beliefs, the measure of opinion polarization across that network increases significantly (i.e. filter bubbles emerge). This certainly gives some weight to the explanation discussed here.

b. Or a symptom of other factors?

It is also entirely possible that filter bubbles emerge inadvertently as a byproduct of other factors. Firstly, algorithms may not be as objective as we would like to believe. Gillespie (2014) argues that the functions performed by personalization algorithms “always depend on inscribed assumptions about what matters, and how what matters can be identified” (p.177). Human biases manifest themselves in the design of the algorithms that dictate what content we see and do not see. Social media companies are careful to declare these algorithms as neutral and objective, but in reality, this is neither true nor possible (Gillespie, 2014). Although our description of item-item al-gorithms in the previous section is relatively straightforward, this is a gross simplifica-tion of what is, in practice, an incredibly complex black-box technology. So complex, in fact, that the engineers behind them may not even fully understand what they have evolved into. Paul Haahr, an engineer at Google involved with ranking algorithms, has openly stated that Google does not fully comprehend the way in which its ranking system works (Schwartz, 2016).

A prime example of human biases being unwittingly ingrained into algorithms is given by Ananny (2011), who discovered when installing Grindr, a dating app for gay men, that the Android Market inexplicably suggested a sex offender locator app under the ‘related’ applications. It is possible that the algorithm employed by Android was able to “identify a subtle association that, though we may not wish it so, is regularly made in our culture, between homosexuality and predatory behavior” (Gillespie, 2014, p. 190). Perhaps, then, the emergence of filter bubbles is simply a reflection on a flawed human way of thinking that has unknowingly been embedded into personalization algorithms - that the information we agree with is the most relevant information.

Alternatively, filter bubbles may have more to do with the choices of users than with the technical infrastructure of social media platforms. Referring back to content-based recommenders, we can assume that personalization is at least partly based on the be-havior of users. It is conceivable, then, that users may indeed initially be presented with a balanced range of content but choose only to engage with attitude-consistent content, thus conditioning the algorithm over time to present more attitude-consistent and less cross-cutting content. This theory is evidenced by Munson and Resnick (2010), who attempted to present more diverse content to challenge-averse people and found that information consumption habits were not significantly affected, regardless of the presentation method. Another study discovered that while over 95% of Face-book users are exposed to at least some amount of cross-cutting content in their feeds, less than 55% of users choose to engage with (i.e. click on) this content (Bakshy et al., 2015).

While user behavior is plausibly at least partly to blame, this selective exposure may in fact be exacerbated by the algorithm’s ranking system. In a study of search engines, Joachims and Radlinski (2007) demonstrate a clear negative correlation between search result rank number and frequency of clicks (p. 35). Moreover, they use eye-tracking technology to show that over half of the time, users do not even look at results below the third rank (p.35). If we extrapolate this behavior to a social media context, we can infer that users will tend to ignore cross-cutting information if it is not displayed near the top of their feed. So, although personalization algorithms may indeed be including cross-cutting content, it may be futile if this content is presented significantly lower in the feed than attitude-consistent content.

IV. Facebook: a brief case study

Though the details of social media personalization tend to be kept confidential, Face-book has occasionally published informative blog posts on their ‘Newsroom’ or ‘En-gineering’ domains, i.e. their platforms for communicating with the public. These will allow us to examine the extent to which the concepts discussed here are manifesting in reality.

In a blog post titled “Recommending items to more than a billion people”, Facebook reveals that they employ a collaborative filtering technique in order to recommend pages and groups to users (Kabiljo and Ilic, 2015). A separate post about a ‘News Feed’ revamp claims to “learn from you and adapt over time” (Mosseri, 2016, para. 5), indicating that content-based recommendation is utilized in their content ranking algorithm. This confirms our prior assumption that social media platforms adopt hy-brid approaches to personalization.

In the same post, Facebook asserts their impartiality, and later states that their “aim is to deliver the types of stories we’ve gotten feedback that an individual person most wants to see. We do this not only because we believe it’s the right thing but also because it’s good for our business.” (Mosseri, 2016, para. 10). Interestingly, this lends itself to the theory that social media companies intentionally design personalization algorithms in a way such that ideological segregation is made inevitable.

Facebook also discusses their collection of preference data. As far as explicit data, they admit that identifying positive signals (e.g. ‘liking’, joining a group, etc.) is much more straightforward than negative signals. Regarding implicit data, they claim their ap-proach is to “treat the data as a combination of binary preferences and confidence values” (Kabiljo and Ilic, 2015, para. 38). In other words, their algorithm analyzes implicit signals (e.g. time spent viewing a post, etc.) to estimate the likelihood that a user will find a given recommendation useful, even when no explicit preference has been put forth. This is significant because it verifies that Facebook can infer more about users than what they are willing to divulge. That is, even if users make a con-scious effort to diversify their information consumption, these algorithms can still es-timate their ideological leaning based on implicit activities, and subsequently adjust the type of content they are exposed to in the future.

V. Conclusions

We have conducted an in-depth exploration of the way in which the technical infra-structure of social media platforms (specifically their personalization algorithms) ena-ble the emergence of filter bubbles. Several possible explanations were identified, namely that this outcome is intentional on behalf of platform owners; that the algo-rithms have unavoidable inherent biases; and that user decisions amplify the effect of the algorithms. Of course, the true answer remains unknown. These algorithms are trade secrets, and the chances of social media companies revealing their inner work-ings are slim.

This research has naturally been constrained by the lack of official documentation regarding these algorithms, but it has nonetheless shed light on a largely overlooked phenomenon that affects nearly everyone. Going forward, there should be more em-pirical research conducted to determine potential countermeasures to online filter bubbles. Munson et al. (2013) developed a browser widget that tracks a user’s internet history and continually displays how ideologically balanced their information con-sumption is. This is a fascinating idea, and there should be further studies that explore a social-media-specific solution, given that people increasingly consume information in bite-sized amounts (i.e. tweets, Facebook statuses, news headlines) as they scroll through their feeds.

It is easy for some to disregard this issue, claiming that they use social media merely to keep in touch with friends and that they consume news and information elsewhere. The reality, though, is that social media platforms have reworked the fabric of the internet as a whole. Through social plugins (one-click widgets on other websites to ‘like’ content through your Facebook account, share content on your Twitter profile, etc.), social media and the external web are “becoming increasingly interconnected with each other, as the activities performed in one space will affect the other, rendering both more open and relational” (Gerlitz and Helmond, 2013, p. 1354). The implica-tion here is that no matter where on the internet one chooses to go, social media platforms are observing, learning, and personalizing.

References

Adomavicius, G., & Tuzhilin, A. (2005). Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. IEEE transactions on knowledge and data engineering, 17(6), 734-749.

Alaimo, C., & Kallinikos, J. (2019). Recommender systems.

Ananny, M. (2011, April 15). The Curious Connection Between Apps for Gay Men and Sex Of-fenders. Retrieved April 15, 2020, from https://www.theatlantic.com/technology/ar-chive/2011/04/the-curious-connection-between-apps-for-gay-men-and-sex-offenders/237340/

Bakshy, E., Messing, S., & Adamic, L. A. (2015). Exposure to ideologically diverse news and opin-ion on Facebook. Science, 348(6239), 1130-1132.

Bozdag, E. (2013). Bias in algorithmic filtering and personalization. Ethics and information tech-nology, 15(3), 209-227.

Bright, L. F., Kleiser, S. B., & Grau, S. L. (2015). Too much Facebook? An exploratory examina-tion of social media fatigue. Computers in Human Behavior, 44, 148-155.

Chitra, U., & Musco, C. (2020, January). Analyzing the Impact of Filter Bubbles on Social Net-work Polarization. In Proceedings of the 13th International Conference on Web Search and Data Mining (pp. 115-123).

Dylko, I., Dolgov, I., Hoffman, W., Eckhart, N., Molina, M., & Aaziz, O. (2018). Impact of cus-tomizability technology on political polarization. Journal of Information Technology & Poli-tics, 15(1), 19-33.

Ekstrand, M. D., Riedl, J. T., & Konstan, J. A. (2011). Collaborative filtering recommender sys-tems. Foundations and Trends® in Human–Computer Interaction, 4(2), 81-173.

Eppler, M. J., & Mengis, J. (2004). The concept of information overload: A review of literature from organization science, accounting, marketing, MIS, and related disciplines. The information society, 20(5), 325-344.

Flaxman, S., Goel, S., & Rao, J. M. (2013). Ideological segregation and the effects of social media on news consumption. Available at SSRN, 2363701.

Gerlitz, C., & Helmond, A. (2013). The like economy: Social buttons and the data-intensive web. New media & society, 15(8), 1348-1365.

Gillespie, T. (2014). The relevance of algorithms. Media technologies: Essays on communication, materiality, and society, 167, 167-193.

Joachims, T., & Radlinski, F. (2007). Search engines that learn from implicit feedback. Com-puter, 40(8), 34-40.

Kabiljo, M., & Ilic, A. (2018, December 3). Recommending items to more than a billion people. Retrieved April 22, 2020, from https://engineering.fb.com/core-data/recommending-items-to-more-than-a-billion-people/

Kahneman, D. (2011). Thinking, fast and slow. New York: Farrar, Straus and Giroux.

Knobloch-Westerwick, S., Johnson, B. K., & Westerwick, A. (2015). Confirmation bias in online searches: Impacts of selective exposure before an election on political attitude strength and shifts. Journal of Computer-Mediated Communication, 20(2), 171-187.

Masullo Chen, G., & Lu, S. (2017). Online political discourse: Exploring differences in effects of civil and uncivil disagreement in news website comments. Journal of Broadcasting & Electronic Media, 61(1), 108-125.

Mosseri, A. (2019, November 7). Building a Better News Feed for You. Retrieved April 24, 2020, from https://about.fb.com/news/2016/06/building-a-better-news-feed-for-you/

Munson, S. A., & Resnick, P. (2010, April). Presenting diverse political opinions: how and how much. In Proceedings of the SIGCHI conference on human factors in computing systems (pp. 1457-1466).

Munson, S. A., Lee, S. Y., & Resnick, P. (2013, June). Encouraging reading of diverse political viewpoints with a browser widget. In Seventh International AAAI Conference on Weblogs and Social Media.

O'Mahony, M. P., Hurley, N. J., & Silvestre, G. C. (2006, January). Detecting noise in recom-mender system databases. In Proceedings of the 11th international conference on Intelligent user interfaces (pp. 109-115).

Schwartz, B. (2016, March 8). Google's Paul Haahr: We Don't Fully Understand RankBrain. Re-trieved April 18, 2020, from https://www.seroundtable.com/google-dont-understand-rankbrain-21744.html

Sunstein, C. R. (2001). Republic.com. Princeton University Press.