

Background

The Observatory for Online Human and Platform Behavior (NSF award number 2131929) is a large-scale project that is currently constructing “a secure, privacy-protecting, ethically robust, scientifically valid online behavioral research observatory” The platform will allow researchers to better understand the algorithmically driven decision-making that governs much of what we see on the Internet. Time-stamped data from users’ social media feeds will be available, allowing us to learn which accounts are amplified algorithmically and what a true ego-view of a social networking site looks like in aggregate. Algorithmic curation influences both voice and attention online. It is difficult to truly replicate the behavior of a curatory algorithm *in situ*; moreover, while algorithmic audits often focus on the (more easily measurable) outcomes of algorithmic decision-making, the process that generates outcomes may itself be biased. **Online discourse is an increasingly important way that coalitions are built and ideas are amplified, but many of the forces that govern how these things happen are, at best, opaque to the average user.** Because social networking sites are fundamentally networked in nature, it can be difficult to disentangle networked effects from algorithmic effects when the two are so intertwined. Edges in a network can encode many interactions or connections; in this proposal, I focus on edges representing one-time interactions (e.g. a user commenting on another user’s post) as well as more permanent follower/friend relationships. The likelihood of any relationship forming is dependent on both individual agency as well as factors endogenous and exogenous to the platform. **In this research, I will highlight situations in which algorithmic curation influences audience-building; examine how power and community are built; and replay community-building in real time to see what role initially weak ties might play.**

Proposal Part 1: Outliers in Algorithmic Curation

Data that is currently being collected through the Observatory for Online Human and Platform Behavior allows us to see users’ Twitter feeds when they visit and scroll through the website. These feeds are, by default, algorithmically curated; information may appear from accounts the user does not, in fact, follow. While some curatory affordances are built to encourage triad closure, others’ origins are more opaque. Using Twitter feed data, I plan to gather aggregate statistics about account appearance rates on user timelines, then attempt to correlate them with follow velocity at the time of their appearance; total account following; and number of recent viral tweets. This approach should yield outliers: accounts for whom traditional, crude measures of recent virality and popularity fail to correctly account for their frequency of appearance on algorithmically curated feeds. I then will compare aspects of the outliers’ network neighborhood and individual Twitter behavior with that of otherwise similar accounts. One of the stated intentions of the Observatory’s data collection is to “[capture] the algorithmic curation of content by third parties,” and I propose this research in part to further this goal. Social networks are shaped by exogenous and endogenous forces; questions of autonomy and choice in online power are fraught when algorithmic curation is optimized to generate click-throughs and ad revenue. **Exposing the ways that power, in the form of online audience, is built and distributed allows us to better understand the limits and potential encoded in our own autonomous decisions.**

Proposal Part 2: Networked Sites of Community

Building an audience is one aspect of building online power, but an audience must be networked within itself in order to be an effective community. To better understand how power is built on social networking sites, I plan to construct multiplex networks of interactions between users on subreddits on Reddit and in a number of salient Twitter hashtags. Reddit and Twitter incorporate different affordances for user interaction and involve different levels of algorithmic curation customized to the user. Constructing multiplex networks representing connection and interaction on these platforms will allow for a rich comparative approach. I plan to use unsupervised clustering methods, alongside features derived from natural language processing methods, to characterize edges in the multiplex network. After constructing the multiplex interaction networks, I will compute descriptive statistics and run robustness analyses on each network, with the aim of understanding whether node degree distributions resemble those seen in other social networks; how robustness to edge deletion varies from between platforms or communities

and between edge types; and what types of communities emerge within each network. I am particularly interested in communities emphasizing care and mutual aid; marginalized groups frequently form such communities online, but endurance and efficacy vary. **This research will surface scientifically validated characteristics of robust, enduring online communities and allow us to better understand how successful communities leverage the affordances of the platforms they call home.**

Proposal Part 3: The Strength of Surprising Ties

Granovetter (1973) posits that “weak” ties – ties that bridge different communities – are key in the information spread and social capital transfer. However, some ties are “weak” initially, by Granovetter’s measure of network overlap, but then become “strong” later on, as a community knits itself together. I suggest that these ties are both weak *and* strong, depending on when you look at them, but they certainly are “surprising” initially. By manipulating the cursors used to fetch following data from the Twitter API, it is possible to put a lower time bound on the following events that are returned: in other words, if we know when a community formed, it is theoretically possible to retrace its formation to arbitrary precision. If a community forms around a new need or activist goal, the initial ties may seem weak or baffling to an observer who only knows about the network’s history and present. Given recently-formed communities, such as those formed around relatively novel hashtags, I will observe the formation of “surprising” ties, quantify how surprising they are, and analyze the role that these “surprising” ties play in the nascent community. **I believe this research will put accurate information on the ways that communities can coalesce around seemingly unlikely connections into the hands of marginalized users looking to amplify their ideas and grow their power.**

Intellectual Merit

If this research is successful, it will indicate which accounts are disproportionately impacted by algorithmic curation and suggest ways that accounts might leverage (or subvert) algorithmic curation to build audience and power. I also plan to create actionable insights on the formation of enduring communities for marginalized groups focused on care, support, and mutual aid. Additionally, replaying the history of communities’ formation will intervene in the existing literature on weak ties, which prioritizes static networks, and answer important questions about the future lives of ties that initially seem weak. **Funding from the NSF GRFP would allow me to make a significant contribution to our current understanding of online social network formation.**

Broader Impacts

Since my work on the Observatory for Online Human and Platform Behavior will be that of an early adopter, I expect that much of the data infrastructure that I create will be intended for repeated reuse and adaptation within the project. I have previously worked as a data engineer, so I anticipate bringing that aspect of my training to bear in creating reliable, reusable data pipelines as I gather and process data for my research. This will in turn facilitate other researchers’ use of the data the Observatory collects, making it a more robust resource for our understanding of social media platforms. I also anticipate using this work to do community outreach for marginalized groups who use social networking sites; disseminating knowledge of the impacts of curatorial algorithms on the formation and robustness of networks may allow them to build stronger networks of care and spread their content further on their platform(s) of choice.

References

- [1] NSF Award Search: Award # 2131929 - Mid-Scale RI-1 (M1:IP): Observatory for Online Human and Platform Behavior.” https://www.nsf.gov/awardsearch/showAward?AWD_ID=2131929 (July 19, 2022).
- [2] Raji, Inioluwa Deborah et al. 2020. “Saving Face: Investigating the Ethical Concerns of Facial Recognition Auditing.” In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*, New York NY USA: ACM, 145–51. <https://dl.acm.org/doi/10.1145/3375627.3375820> (July 19, 2022).
- [3] Granovetter, Mark S. 1973. “The Strength of Weak Ties.” *American Journal of Sociology* 78 (6): 1360–80.