Detecting and Tracking Polarized Communities in Temporal Networks

(Technical Paper)

Social Media As a Public Good: Finding a Compromise Between Public And Private Interests In Social Media Content Moderation

(STS Paper)

A Thesis Prospectus Submitted to the

Faculty of the School of Engineering and Applied Science

University of Virginia • Charlottesville, Virginia

In Partial Fulfillment of the Requirements of the Degree

Bachelor of Science, School of Engineering

Joseph Spaeth

Fall, 2020

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

Signature ____ Date __________

Joseph Spaeth

Approved ____ Date __________

John R. Hott, Department of Choose Department

Approved ____ Date __________

Kathryn A. Neeley, Associate Professor of STS, Department of Engineering and Society

Introduction

Facebook currently has 2.7 billion users, growing massively since its inception in 2004. This is indicative of the broader trend over the last 2 decades of social media use increasing, and playing a more outsized role in the lives of its users. Particularly over the course of the last six years it has become increasingly clear that in those social networks, different types of communities will form around different groups, the most relevant of which is the "echo chamber" phenomenon, wherein users form in-group communities based upon shared beliefs, and only allow in users sharing those same beliefs, and isolating those users from outside interaction and information (Nguyen, 2019; Dizikes, 2018). Additionally, relationships on social media can often be represented using graphs, wherein the vertices (or nodes) are users and the

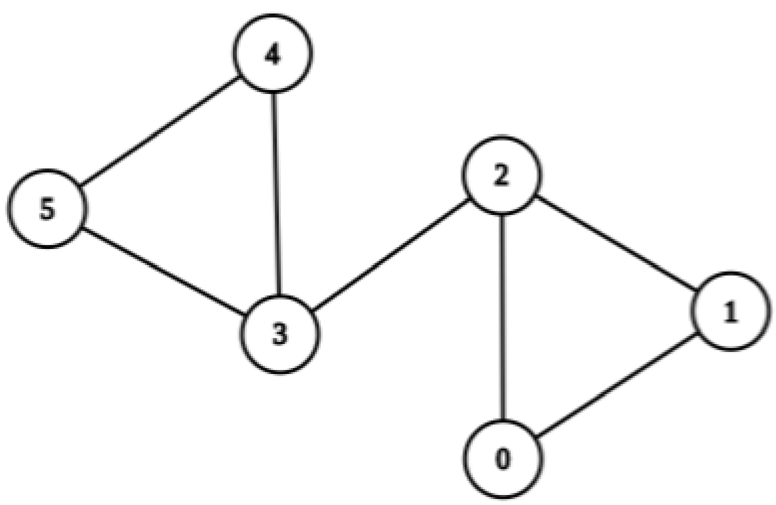

Figure 1: A simple network with nodes labelled 0-5. The edges represent a relationship between nodes. In the case of this project, nodes in the graphs discussed are twitter users, and edges of the graphs are interactions (follows, retweets, mentions) between the users (created by author).

relations between them are the graph's edges, as seen in Figure 1. Echo chambers specifically are densely connected and high modular subgraphs (subsections of the network), where many users within the chamber interact, but very few of those users interact with users outside of that "bubble", as in Figure 2.

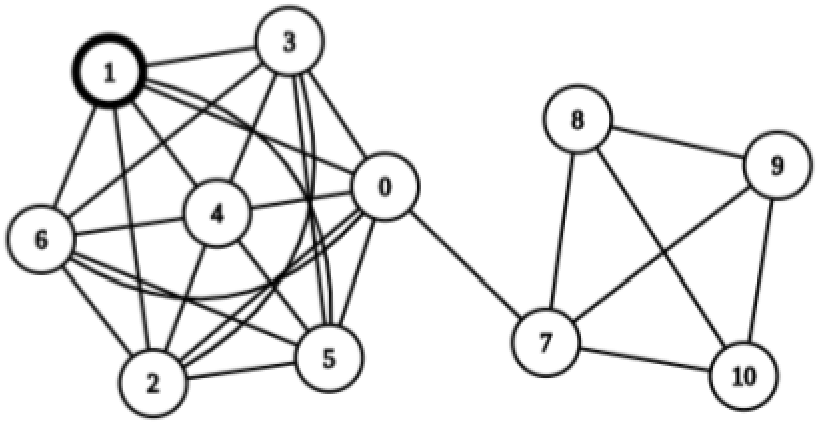

Figure 2: A graph with an echo chamber. The leftmost subgraph of this graph (the subgraph containing nodes 0-6) exhibits the characteristics of an echo chamber. It is modular, not well connected to the rest of the graph, and highly connected within itself (created by author).

Though we understand what these groups look like once they have formed, we lack that same understanding for how they form and evolve over time, and temporal networks can be leveraged to understand that formation and evolution process (Rossetti, 2018, 5-7). Temporal networks are graphs like this, but where nodes can also be connected by edges representing a change in time. There is not yet a complete understanding of this process on the technical side. Accordingly, we intend to gain a better understanding of this phenomenon by applying these models of temporal networks (and the algorithms related to them) to real Twitter data. (Recuero, 2019, 6).

Additionally, the formation of echo chambers is undesirable, as they lead to the development of extremist political rhetoric, as well as the proliferation of false news and conspiracy theories as a result of their insular nature (Dizikes, 2018). If we do not better understand echo chamber development, it will remain difficult to know how these groups form and evolve, and likewise allow for the continued spreading of misinformation and fake news. In order to resolve this, my technical project will apply temporal modularity measures and community detection algorithms to real-world twitter data, which will allow us to detect the groups (communities) through time by optimizing network modularity to better understand group formation. My STS project will explore the consequences of using this in order to quell the formation of echo chambers online, weighing the pros and cons of moderating platforms to prevent the proliferation of echo chambers.

Technical Topic: Developing and Applying Modularity and Community Detection Measures to Temporal Networks

As stated prior, a network consists of a group of vertices (or nodes), V, and a group of edges, E, connecting them. The key difference between this and a temporal network is, naturally, a temporal aspect to the data at hand. For example, one can see the "snapshot" method of modelling a temporal network in Figure 3.

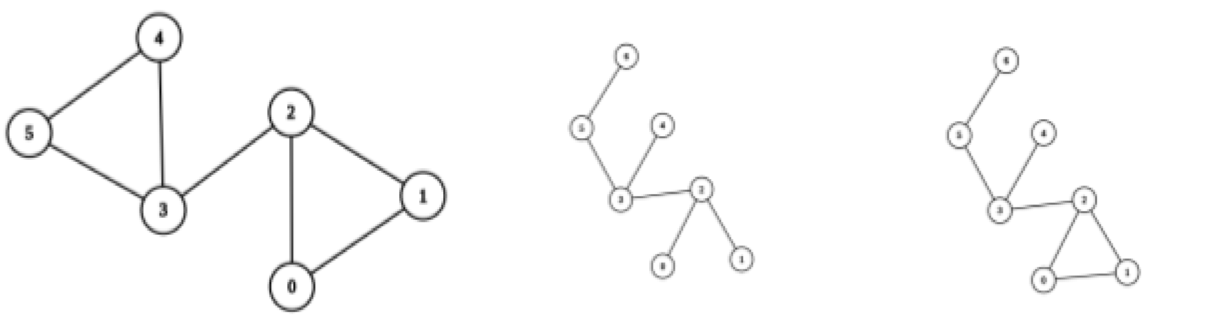

Figure 3. A snapshot based approach to representing a temporal network. This is a single network, viewed across three different timestamps. The left being the earliest, the right being the latest. The snapshot approach represents a temporal network as slices in time. One can think of it as a collection of static networks. In this figure, the graph gains a node, the connection between nodes 4 and 5 is severed, and the edge between nodes 0 and 1 is severed, then reestablished later (created by author).

This could be reflected in the relationships (edges) changing over time, appearing and disappearing, the nodes themselves appearing and disappearing, or both (Rossetti, 2018, 7). In order to better understand group formation within networks, we use measures of modularity and community detection (Himelboim, 2017, 6). Modularity is a measure of how well connected subsections of a given network are. For example, a graph with high modularity will have well connected subgraphs, but not be connected well (or not at all) between those subgraphs. Similarly, community detection algorithms are used to define those subsections and point them out to a user (Casteigts, 2011, 12-15). The goal is to identify where those sections of interconnectedness lie, as they represent communities.

At the moment, we lack the effective modularity analysis and community detection algorithms we have for still networks for temporal networks. The best community detection algorithms lose some or all of the relevant temporal information when finding communities and measuring modularity (a measure explained in Figure 4), which is not helpful (Casteigts, 2011, 10).

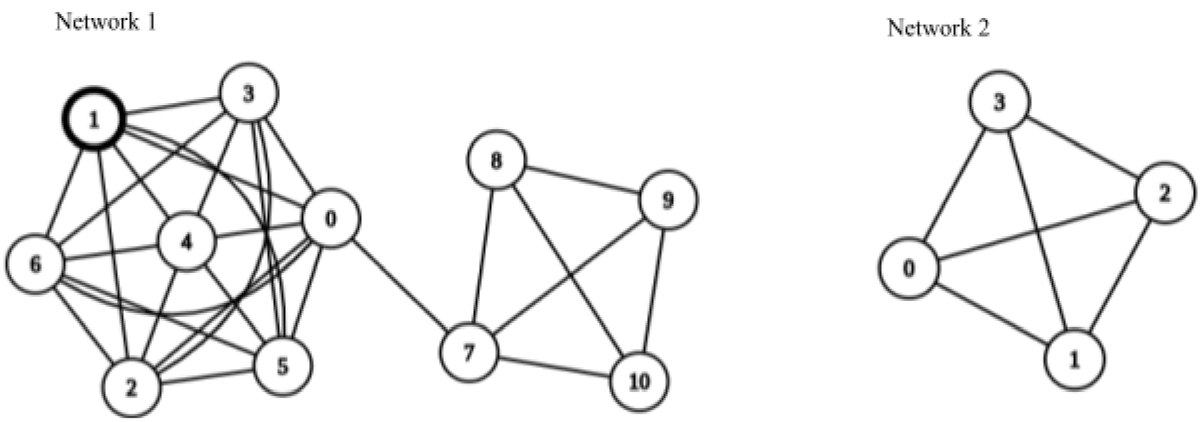

Figure 4. Two networks with varying degrees of modularity. Modularity is a graph metric used to represent how well the graph can be split into well connected subgraphs. Network 1 is quite modular, as it can clearly be split into 2 distinct groups, whereas network 2 is not. Modularity is a numeric measure of how well a network can be evaluated in this manner, and it is often the goal of community detection algorithms to maximize its value (created by author).

For example, Rosetti points out how both the choice of model and algorithm can have a significant effect on how the communities are detected. Additionally, the methods of evaluation proposed by them are untenable for use on social media datasets, as they rely on ground truth values (already knowing how a graph should be grouped) which cannot be easily attained for datasets of the size we will work with (Rosetti, 2018, 6). In addition to the choices of model and algorithm playing a role, one of the best algorithms we have at the moment is suboptimal. Nicosia points out in their paper on graph metrics that “this definition of modularity is invariant under inversion of the sequence of slices which, in the particular case of time-varying graphs, implies invariance under time inversion" (Nicosia, 2013, 23). As a result, when using the algorithm proposed in that paper, information about how the communities form over time is lost, as the algorithm produces the same results if one is analyzing the data backwards in time or forwards in time. This is a demonstration of how temporal information can be lost depending on algorithm choice (a better algorithm would be time variant), and likewise demonstrates the limitations of current methods.

Currently there are not many pieces of software available which efficiently and/or effectively handle large temporal networks. The best available software is the now deprecated (no longer up to date) PathPy2, and the new version of that software (PathPy3) is still in development (Scholtes, 2020). Other options include using the deprecated version of that same library, Gephi, and NetworkX. Gephi and PathPy2 are both deprecated, and additionally, Gephi's support for dynamic networks is seriously lacking, as it consumes large amounts of computational resources for little payoff in terms of what it can actually do. NetworkX though, is a popular and effective python package used for static networks. However, it is only effective with time-slice style temporal networks, as it does not come with any support for temporal networks baked in. This leaves its potential applications limited as well, and my work will ideally provide a more accessible means of large scale temporal network analysis.

To alleviate this issue, I propose building on the prior writing on the topic of temporal networks by using the modularity measures and community detection measures (particularly Nicosia and Rossetti's work) to either add functionality to PathPy3 and to make it useful for analyzing large temporal networks, or to write wrapper code for other network analysis software to improve their usability with temporal networks. Though difficult, this will allow users to better understand how large structures in social networks form.

STS Topic: Implications of Using Community Detection Measures to Track Echo Chambers on Social Media

It is currently established and accepted knowledge that fake/false news spreads through social networks primarily through interactions between human actors, acting out of confirmation bias instead of a concern over the truth (Dizikes, 2018; Nguyen 2019). As these act as a vector for which false information and extremist ideology can quickly spread, understanding how echo chambers in networks form and evolve is vital to gaining a better comprehension of how misinformation will flow through a social network. With the ability to track these things comes the question of how to best apply that. It forces the user of the relevant software to determine what news is genuine and what news is not, and what constitutes extremist ideology.

This makes the question of if and how social media sites should police their content a vitally important question. Nguyen notes how echo chambers online are often marked with extremist rhetoric, confirmation bias, and the speedy dissemination of information within those subsections of larger networks (2019). Particularly in a time where it is becoming easier to produce doctored audio and visual content, it is of vital importance that we gain an understanding of how information flows in these networks, in order to prevent the further spread of fake news. Also, conspiracy theories will bubble up on the internet (Nguyen, 2019; Salam 2020). These serve two purposes, they radicalize people in the in-group, and then attempt to grow the group further. As these groups operate on confirmation bias, it becomes very easy to produce false or misleading content, and ensure the others in one's group believe it. Nguyen uses Rush Limbaugh and Fox News as an example of this, demonstrating how the news network manipulates who its viewers and readers trust, which in turn gets them to only see some sources of news as legitimate.

An example of echo chambers in action, and the dangers they cause is the Qanon conspiracy and community surrounding it. Qanon started as a small fringe group in 2017, and has grown rapidly since. There are now 40 different politicians running for office attached to the group, and Salam points out how well they spread misinformation saying “The group has effectively spread dangerous misinformation, particularly during the coronavirus pandemic” (Salam, 2020). As well as the networks themselves encouraging the spread of misinformation, social media algorithms play a large role in it. As most social media sites operate on showing the user posts the site believes that user will interact with, they will be shown posts which align with the things a user has already reposted, commented on, liked, etc (Staff, 2016). This in conjunction with the behavior that echo chambers inherently encourage will only allow this issue to grow. The current solution of banning groups off of a platform well after they have caused damage (for example, the recent purging of Qanon content from Facebook) is not strong enough to fix the issues at hand. Preventing the spread of misinformation well after it spreads is ineffective policy at best.

With this in mind, I propose the application of ANT and the knowledge gained from the technical part of this project to track the formation of echo chambers. This information will allow social media networks to make informed decisions related to potentially shutting them down prior to them reaching critical mass in the manner Qanon did. ANT will be used as a framework to attempt to balance the interests of free and protected speech, social media users, and the promotion of safe and civil discourse. The goal of this project is to build on prior research such as that done at MIT by Dizikes, showing how quickly misinformation can spread through social networks, and to help make value based decisions on how to best handle the rapid spread of misinformation and conspiracy theories on social media platforms. Ideally, this will be able to help answer the question of how one can manage concerns of free speech while simultaneously promoting safe and civil discourse online.

Conclusion

The anticipated technical deliverables for this project are a paper on the subject, detailing the findings about the viability of using community detection to find echo chambers before they grow, any and all relevant code and visualizations used, and dataset(s) created by myself and Prof. Hott. This will provide a better understanding of structures within temporal networks form and evolve, hopefully shedding more light on the echo chamber phenomenon. The STS deliverable is a paper, which will use the knowledge gained from the technical project in conjunction with ANT to explore the costs and benefits to tracking social networks in this manner such that we can have a better understanding of the proliferation of misinformation online.

The proper implementation of both of these things would provide us with an improved means of finding communities in temporal networks. This can then be applied and generalized to the understanding of in-group/out-group behavior in social networks, allowing us to better understand why echo chambers form in the manner they do. Furthermore, that would allow for social media sites to quickly catch and prevent the proliferation of hate speech and fake news.. This however, requires massive amounts of content moderation, which presents challenges in its own right. This is where the STS portion of the project becomes relevant, as it serves as a means of better determining when and how a platform owner should police its content, and how to best compromise between that and the interests of free speech.

References

Bi, Y., Wu, W., & Wang, L. (2013). Community expansion in social networks. *Database Systems for Advanced Applications Lecture Notes in Computer Science,* 41-55. doi:10.1007/978-3-642-37487-6_6

Casteigts, A., Flocchini, P., Quattrociocchi, W., & Santoro, N. (2011). Time-Varying graphs and dynamic networks. *Ad-hoc, Mobile, and Wireless Networks Lecture Notes in Computer Science,* 346-359. doi:10.1007/978-3-642-22450-8_27

Dizikes, P. (2018, March 8). Study: On Twitter, false news travels faster than true stories. Retrieved October 09, 2020, from https://news.mit.edu/2018/study-twitter-false-news-travels-faster-true-stories-0308

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter topic-networks using social network analysis. *Social Media + Society, 3*(1), 205630511769154. doi:10.1177/2056305117691545

Nguyen , C. T. (2019, October 31). The problem of living inside echo chambers. Retrieved October 09, 2020, from https://theconversation.com/the-problem-of-living-inside-echo-chambers-110486

Nicosia, V., Tang, J., Mascolo, C., Musolesi, M., Russo, G., & Latora, V. (2013). Graph metrics for temporal networks. *Understanding Complex Systems Temporal Networks,* 15-40. doi:10.1007/978-3-642-36461-7_2

Recuero, R., Zago, G., & Soares, F. (2019). Using social network analysis and social capital to identify user roles on polarized political conversations on Twitter. *Social Media + Society, 5*(2), 205630511984874. doi:10.1177/2056305119848745

Rossetti, G., & Cazabet, R. (2018). Community discovery in dynamic networks. *ACM Computing Surveys*, *51*(2), 1–37. https://doi.org/10.1145/3172867

Salam, K. (2020, October 12). *What Does the QAnon Conspiracy Mean for Politics in the United States—and Abroad?* Foreign Policy. https://foreignpolicy.com/2020/10/12/qanons-sound-and-fury/

Sholtes, Inigo. PathPy3 [2020]. University of Zurich. Retrieved from https://github.com/uzhdag/pathpy

Staff, N. (2016, July 24). The Reason Your Feed Became An Echo Chamber - And What To Do About It. Retrieved October 09, 2020, from https://www.npr.org/sections/alltechconsidered/2016/07/24/486941582/the-reason-your-feed-became-an-echo-chamber-and-what-to-do-about-it