ACTIVISM UNDER OPPRESSION

How Internet-Age Social Movements Organize to Resist Structural Attacks, Internal Schisms, and Disinformation

MILO TRUJILLO
RON EGLASH

Abstract

This study explores how Internet-organized social movements react to threats of internal schism, infiltration, and external legal and political pressure. The investigation is focused on styles of leadership and how membership of a movement is defined. Using maps of social interactions within Twitter communities, the study analyzes the structure of social movement leadership and membership, and compares their methods of conflict-resolution against what is predicted by contemporary research on infiltrating and discrediting social organizations. This analysis found that organizations in threat of external attack, by law enforcement or public shunning, rely on an anarchic structure and loosely defined membership to protect the identities of their participants. In contrast, organizations facing internal conflict or infiltration form more hierarchical leadership and tightly controlled membership lists to isolate threats to group consensus. These findings have the potential to aid future movement organizers in making more robust organizations in hostile political climates.

1 INTRODUCTION

Social movements face two broad categories of threats: External, and Internal. External threats are any dangers the movement faces from individuals or groups outside the movement, such as arrest and dissolution by law enforcement, as in the case of social movements advocating illegal activity, such as Antifa or hactivist groups like Anonymous. External threats also include social ridicule and exile, particularly in the case of broadly socially unacceptable movements like white supremacy. Internal threats are any dangers originating from members within the movement, and include disagreements over strategy and methodology, leadership disputes, and disinformation or other undermining campaigns by hostile actors within the movement. While "hostile actors" can include agent provocateurs from an external group, these individuals are by intention indistinguishable from legitimate angry or disillusioned members of the movement, and therefore have to be treated as an internal threat. This study outlines the close connection between the structure of a social movement, and the internal or external threats it is defending against.

Throughout this paper I will refer to both *social movements* and *activist networks*. In this context, activist networks are the communities of participants within a social movement. A movement encompasses both the people in it and a broader political struggle, including objectives and methodology. Therefore the terms are closely related, but not interchangeable.

In this study, I analyze the structure of activist networks on social media, particularly Twitter. These groups are of interest because their use of "digital commons" allows them to organize rapidly, without central leadership, at a massive scale. These groups therefore may present novel techniques for social mobilization, with strengths and weaknesses unseen in earlier social movements.

Of particular interest are the following questions:

1. How is membership of decentralized movements defined? Do formal initiation rituals benefit or limit a movement?
2. How are decisions made in online movements? Is there diffused leadership, consensus, or individual direct action? What effects do these strategies have on group cohesion?
3. How vulnerable are different leaderless designs to infiltration and disruption by external forces? Which organizational strategies are the most resistant to disruption?

Answering these questions will be beneficial for two groups. First, movement organizers will have data to support decisions on group hierarchy and communications methods, hopefully aiding the creation of more efficient and robust social movements. Second, academic researchers will have a starting point for mathematical and computer modeling of social movements, an area so far surprisingly devoid of study. This supports future research which will hopefully be of further benefit to movement organizers.

2 METHODS

I performed a comparative study on the centralization and structure of online activist networks. Rather than interviewing members of movements or analyzing publications by social movements, I chose to use Twitter interactions as an indicator of the structure of activist networks. Participants within a movement may not be able to identify the power dynamics of the group from within the organization, so I believe it is valuable to look at a network's structure from the outside. Social media does not reflect the exact leadership hierarchy of an organization, but it does indicate who is speaking to whom, which in turn can identify social ties and provide hints as to the communication structure of the group (Hargittai and Sandvig 2015). This type of social analysis is not uncommon in network science, and has previously been used to uncover social ties among college students (Blue 2018), and the creation of closed social communities within the virtual reality game *Second Life* (Welles et al. 2014).

To perform this study, I built a map of Twitter users that self-identified as part of social movements, and their associates. I analyzed the structure of connections on this map to ascertain attributes about the information flow in the activist network. Finally, I compared the uncovered structure to the structure described by members of the group, and compare reports on how the group has dealt with internal disagreements with literature on attacking groups with a similar structure.

2.1 *Data Collection*

To collect data, I began with a group of "seed users", who self-identify as active members of an organization. These will usually be figureheads, spokespeople, or spokesaccounts, broadcasting a public message for the group. I then built a social media analysis system, which given a list of starting usernames, performed the following task:

The system collected a sample of tweets (about 2000) from each seed user, and saved them as an example of discourse from the group. It then read through each of the tweets and extracted the usernames of other accounts *mentioned* or *retweeted* by the seed account. These mentions and retweets constitute network connections between users, and roughly correspond to a conversation or citation between users (Boyd, Golder, and Lotan 2010). The software repeated this process on each of those users, recursively, until I had a network of two to three layers of connections. This provides a sample of accounts connected to the movement.

Data collection cannot usually proceed past the third layer as a result of the *small world problem* (Travers and Milgram 1967). Put simply, after a few levels of association every user is connected to a vast array of others with no distinctly shared social interests. Including this data floods the sample of the social movement (the third layer of the *Black Lives Matter* data contained over one million users), and makes it both technically demanding to analyze, and qualitatively challenging to extract meaning from. While the small world problem traditionally assumes network topologies are either random or uniform, of which Twitter is neither, the high-peer problem remains a challenge even in more complex social media networks (Watts and Strogatz 1998).

2.2 *Network Analysis*

After pruning users with a low number of network connections, which are unlikely to yield insights about the social movement, the collected data is small enough to visualize. I rendered maps of each community, where circles represent Twitter users, and arrows between circles represent social connections (retweets or mentions). The maps were oriented with "edge-repulsive weak-clustering," which generally means that well-connected users are drawn close to one another, users with no connections are drawn far apart from one another, and users are positioned to minimize overlaps in the graph. This produces visually helpful representations of the data, where an observer can see clustering indicating sub-communities within the activist network, in which users communicate among themselves more than to the rest of the network.

2.3 *Comparison to Literature*

Using the above maps, I compared the structure of each activist network to social movements discussed in literature. I am particularly interested in three questions:

- How does the social-media network structure of the organization compare to their self-described leadership model? Are "leaderless" movements as decentralized as they claim?
- How has this organization dealt with dissent? How does it mediate disputes and make decisions? Does its success with group cohesion match what we would expect from its network structure?

2.4 *Groups to Examine*

There are several communities established on Twitter that I examined:

- Hactivists, including Anonymous and Telecomix
- Alcoholics Anonymous
- Antifa
- Black Lives Matter
- Hate Groups

These groups were chosen for their wide recognition, large membership, and, except for Alcoholics Anonymous, their significant social media presence. Alcoholics Anonymous is included as a control-group for comparison to other organizations: It is a thoroughly studied and well-understood social movement with a smaller online presence.

Since I was not familiar with the names of a wide array of hate groups to search for, I instead began with a list of hate groups from the Southern Poverty Law Center (SPLC), a non-profit organization that tracks the growth of hate groups throughout the United States. I included only hate groups with an active Twitter presence. The selected hate groups include the American Nazi Party, ACT for America, American Patrol, American Family Association, FAIR Immigration, the Oathkeepers, the Nation of Islam, WorldNetDaily, White Lives Matter, and the Westboro Baptist Church.

2.5 *Selection of Seed Users*

To gather my data I manually selected "seed users" from each social movement, then generated a map of who these seed users interact with, who those users interact with, and so on. Seed users were generally chosen by a Google search for "<social movement> Twitter," then selecting Twitter users with a high number of Tweets and followers. This methodology is appropriate, since I believe it is also how an interested user would learn more about or join an online social movement.

2.6 *Limitations*

Of chief concern is establishing "ground truth." I can analyze Twitter networks endlessly, but cannot easily confirm whether the leadership structure presented online is equivalent to the leadership structure used in physical space. For this reason I focus on Hactivism and Black Lives Matter; movements organized primarily publicly and online, where I can minimize the possibility of gathering misleading data. While it is possible for there to be a "shadow leadership" for these groups, it is unlikely, as hactivist organizers already use psuedonyms on Twitter and gain little by running a second communication hierarchy, and Black Lives Matter organizers have a reduced audience if they do not communicate on Twitter.

Some organizations, like Antifa and Alcoholics Anonymous, may organize their local chapters offline, and only use Twitter for interactions between chapters. This means these analysis techniques cannot provide insight in to the organizational strategies of social movements at a local level, but can still be used to investigate the movement at a national or international scope.

The initial seeds for data collection were chosen subjectively. It is conceivable that by choosing poor starting points I may have skewed the communities I examined, and miss sub-groups that are important to the structure of the movement. This is of limited concern, because the small-world effect suggests that even with non-ideal starting points my data sets will quickly include nearby sub-groups.

Twitter significantly limits the data available for this study, and provides only close to the 2000 most recent Tweets by a user. As a result, this study is limited to temporally-recent communication, and cannot show how the leadership of movements has changed over time. A longer-term study could repeatedly collect Twitter data on the same organizations and produce such results in real time as the social movement changes, but that is outside the scope of the current study.

In some cases, the communities I monitored changed or collapsed as I attempted to analyze them. This is particularly prevalent in the hate group data set, where Twitter banned central organizing accounts from the *American Nazi Party* and the *New Black Panthers* just before I began data collection. Twitter has also dissolved parts of the Antifa user base during their "Bot Purge" of early 2018 (Gallagher 2018). In these cases my analysis of leadership may be hindered, but observing the attempts of these groups to re-organize around their missing leadership also provides a valuable case-study in hierarchy reacting to targeted damage.

3 ANALYSIS

3.1 *Hactivism*

Below is a minimal map of hactivism communities, mostly centered on Anonymous and Telecomix. To seed data collection I began with eleven users that self-identified as hactivist news groups, speakers representing the organizations, or journalists and academic researchers that study hactivism.

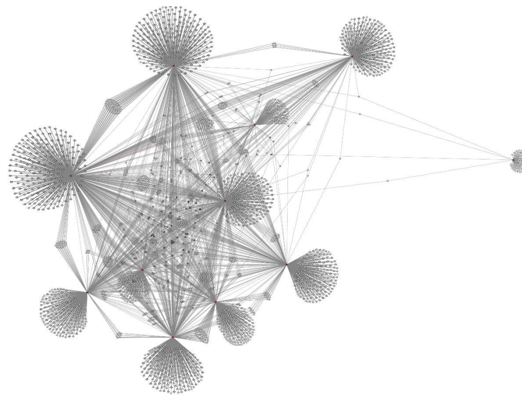


Figure 1: Hactivism, One Layer from Seed

We can immediately see clusters around each of the seed nodes, suggesting the presence of distinct communities. Two clusters in the center are focused around "anonyops" and "anontvofficial", accounts self-identifying as news groups within Anonymous. These accounts do *not* represent independent communities, but are instead are followed by many people also connected to separate hactivist groups on the periphery of the graph.

The community isolated far to the right is the Chaos Communications Congress, a German political hacker conference with less overlap than the intersection of predominantly American and Western-European hactivist groups at the left of the graph.

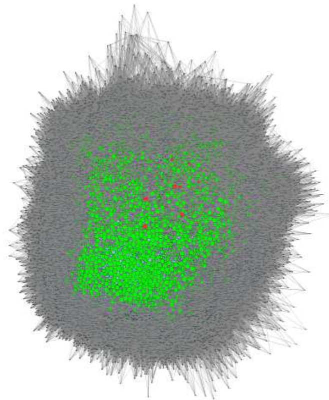


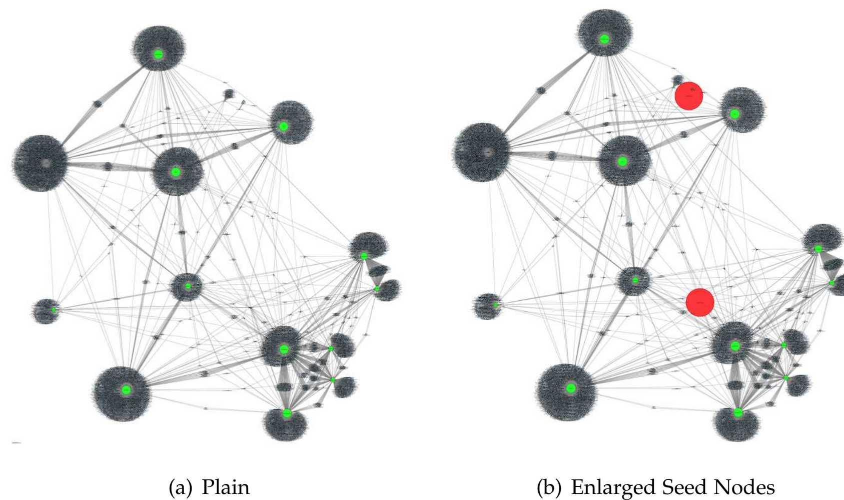
Figure 2: Hactivism Retweets, Two Layers from Seed

At a higher level of analysis these initial clusters melt together (Figure 2). This suggests that while there may be some variation in hactivism giving the appearance of distinct groups, the community is quite fluid, with members participating in, or moving between, many different sub-communities. This matches the findings of Anonymous researcher Gabriella Coleman, who defines Anonymous as a loosely-defined "scene", where "many Anonymous-based nodes and collectives, whether small teams, larger networks, or simply groups of loosely connected Twitter accounts, form, disband, and regroup in new ways in the course of weeks or months. Others have existed in relatively stable shape now for 5 years" (Coleman 2017).

Coleman's description of Anonymous appears to generalize to most hactivist communities, and echoes the prose of the oft-referenced *Hacker Manifesto*, which includes "We exist without skin color, without nationality, without religious bias, ... You may stop this individual, but you can't stop us all... after all, we're all alike" (The Mentor 1986).

3.2 *Alcoholics Anonymous*

Alcoholics Anonymous presents themselves as a leaderless organization. Members are forbidden from using their names when representing AA, to prevent the rise of unintentional power-dynamics and potential corruption (Wilson 1953). Accordingly, I expect Alcoholics Anonymous to have a minimal online presence used for spreading information, but without a distinguished "social movement" centered around their accounts.



(a) Plain (b) Enlarged Seed Nodes

Figure 3: Graphs of Alcoholics Anonymous and Surrounding Accounts

Above are graphs of two layers out from five prominent AA accounts online, found by their number of Tweets and use of hashtags associated with Alcoholics Anonymous. Three of the five AA accounts examined did not participate in social networking whatsoever, and did not retweet or mention any other users, isolating them from the graph. The remaining two users *did* retweet and mention other accounts, but are dwarfed by the social traffic of their peers. In the left graph, where the size of nodes is determined by their number of connections to other accounts in the graph, the Alcoholics Anonymous accounts are not easily discernible. In the right graph, the size of AA accounts has been artificially inflated to make their locations apparent.

The large clusters in this network represent a range of interest groups and pop culture icons, ranging from the musician "P!nk", to a former administrator at Twitter, to YouTube. These groups do not appear to have any shared objective with Alcoholics Anonymous, and were likely included because they are interests of individual people mentioned or retweeted by the AA accounts.

Clearly, Alcoholics Anonymous does not have a social media community centered around them. This is the behavior one would expect, and suggests that the clustering seen in the hactivist and other data sets *is* indicative of community, and is not implicitly part of Twitter.

3.3 *Antifa*

Antifascism is a relatively new phenomenon in the United States, started in response to Nazi rallies in late 2016. This is in stark contrast to European

Antifascism, which originated in the second World War resistance of Axis powers. As a result, I was not surprised to see vibrant United States Antifa activity on Twitter, that appears to be largely self-contained without obvious links to traditional European Antifascism. What I did not expect to see was clear bifurcation within the U.S. Antifa network (Figure 4).

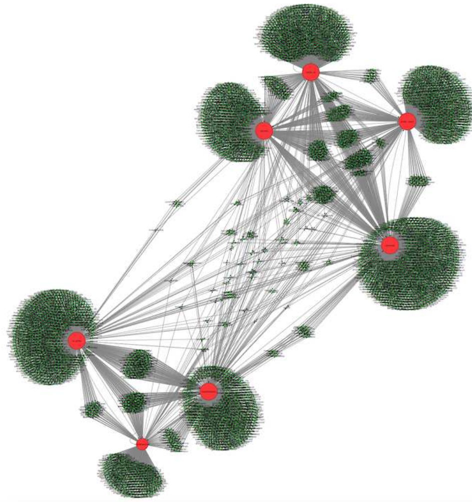


Figure 4: American Antifascist Activity, One Layer from Seed

In the graph, the Antifa accounts in the lower left have federated with what they refer to as *The Antifa Army*. Their website features a list of "confirmed comrades", suggesting this may be an attempt to counteract disinformation campaigns from fake Antifa Twitter accounts in mid-2017 (Gallagher 2017).

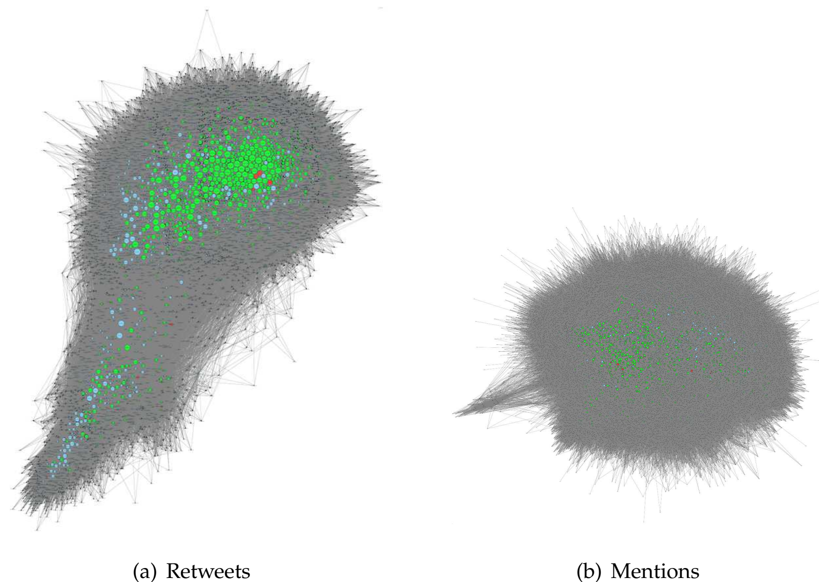
In fact, this could not be further from the truth. Many of the disinformation accounts mentioned in Gallagher's study *are the same accounts* as in this "Antifa Army", and are present on a community blocklist of fake antifa accounts. The blocklist is maintained by a Twitter account named "Antifachecker", and the Antifa Army maintains a parallel account named "Antlfachecker" (with an 'l' in place of a similar-looking 'i'), which works to discredit other Antifa accounts.

In other words, the "Antifa Army" is a concerted effort to give legitimacy to false-flag accounts by creating a parallel sphere of Antifa social activity, where fake accounts mention and retweet one another to appear as part of a larger movement. The tactic is effective at least to cursory observation; I unwittingly selected several Antifa Army accounts as seed nodes since they appeared in the top hits on a Twitter search for "Antifa".

The Antifa accounts in the top right of the graph are not associated with this "Antifa Army", but are on average more active on social media, and have larger peer networks as a result.

3.3.1 *Analysis of Layer 2*

When I extended analysis out one layer further the bifurcation became more complicated. The extra layer increased the number of users in the map from 4500 to over one million, so I have split the map in to "mentions" and "retweets" and employed extensive data pruning on each:



(a) Retweets (b) Mentions

Figure 5: Pruned Graphs of Antifa, Layer 2

In the retweet data the bifurcation is still clear. The dominant community to the top is the legitimate Antifa activity, while the smaller stem on the bottom of the graph is the "Antifa Army". The distinction is muddier, presumably because some observers cannot distinguish between the legitimate and false Antifa accounts and retweet both. However, in general, members of the Antifa community promote either the message of the legitimate accounts or the deceptive accounts, creating visually distinct communities.

In the mention data, this bifurcation is completely lost. The two groups appear to be engulfed in one amorphous community. I believe this is because of the distinction in meaning between a mention and a retweet. Retweets further the spread of a message without modification, which implies support of the message. Mentions are only a way of messaging another user. An informal sampling of Tweets directed at accounts in the Antifa Army indicates that many are denouncements of Antifa Army, requests to stop their activity, and threats. Tweets *from* the Antifa Army frequently reference legitimate Antifa accounts, congratulating them on morally-repugnant acts that often never occurred (Silverman 2017). These mentions do not indicate a single cohesive community, but rather two communities engaged in media combat.

3.3.2 *Formal Membership*

The network diagrams above indicate that Antifa is largely successful at thwarting the disinformation campaign of "Antifa Army", because the disinformation users form a distinct community rather than seamlessly infiltrating other Antifa social groups, but it does so by formalizing membership and centering activity around trusted hubs.

Antifa is an inherently open-membership loosely-affiliated group: All that membership requires is taking anti-fascist action, so someone can be a self-identified member without associating with any other Antifa chapters or individuals.

However, in an attempt to denounce bots, propaganda, and false-flag operations, many prominent Antifa individuals have collaborated on the "Antifa Blacklist", which formally defines a list of non-Antifa-individuals. This explicitly places the prominent Antifa accounts in a position of authority where they can define who is and is not part of the movement. These gatekeepers utilize shared language, symbolism, and historical references to distinguish legitimate members from impersonators, which places pressure on members to conform to a monoculture within antifascism.

3.4 *Black Lives Matter*

Before I began analysis of Black Lives Matter (BLM), I anticipated the group to be amorphous, much like the second layer from the hactivism data set (Figure 2). The accounts I used for seed nodes did not identify themselves as being local chapters, bound to physical regions or a particular sub-topic. Rather, each had thousands of followers and identified itself as a main news anchor for the entire social movement. Therefore, I expected significant overlap between the communities of each account. Instead, I found fragmentation (Figure 6).

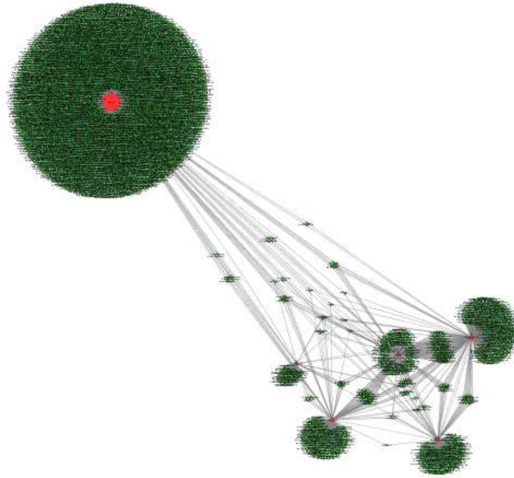


Figure 6: Black Lives Matter, One Layer from Seed

One layer further out, Black Lives Matter solidifies in to two broad camps (Figure 7) suggesting that the first layer clustering is not a fluke, but represents at least two distinct ideological or political differences.

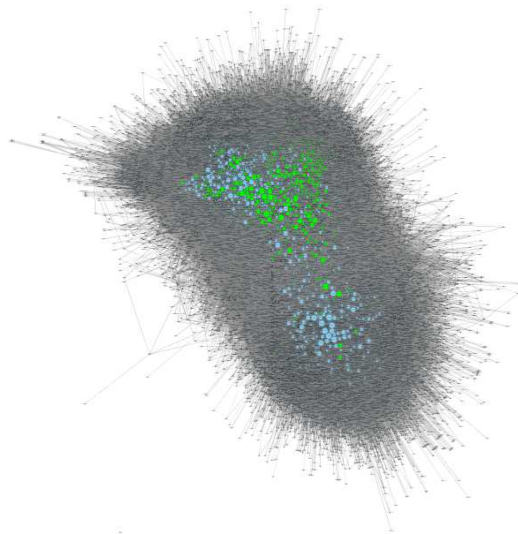


Figure 7: Black Lives Matter, Two Layers from Seed

Further analysis unveiled that my understanding of Black Lives Matter was severely flawed, likely stemming from ignorance as a researcher that has not been embroiled in race movements. BLM began as a slogan at protests, and a hashtag rallying cry, gaining notoriety after the 2014 Fergusson protests and leading to the creation of many local groups focused on issues of racial inequality and police brutality. However, Black Lives Matter rapidly institutionalized in the lead-up to the 2016 presidential election. This organization, nominally calling itself *The Black Lives Matter Global Network*, but often simply

referred to as *Black Lives Matter*, coordinates local chapters with an alert system and shared resources.

The federated model of BLM Global Network, while less formalized than a parliamentary system, still implements a degree of leadership hierarchy, with organizational teams, initiatives, and action platforms. These include formally-defined national organizations like *The Movement For Black Lives* (Often referred to by the acronym "M4BL"), which organizes petitions, town hall meetings, and electoral engagement.

There is now a schism in the Black Lives Matter community between two groups: Those that adhere to a revolutionary, self-organized political model, and those that have nationalized and work towards improvements within the current political system. Proponents of the former group see the later as co-opting the black liberation movement and its accomplishments for their own agenda, and intercepting media attention and funding that would otherwise support local movements (Black Lives Matter: Cincinnati 2018). More broadly, there are accusations of the later group "selling out" and joining a national trend of corporatized activism (Dauvergne and LeBaron 2014) that reinforces and works within the existing power structures of the country without challenging the framework of oppression itself.

3.5 *Hate Groups*

Hate groups present an interesting case study, as they have similar organizational and coordination problems to other social movements, but face heavy public criticism and frequent conflict with law enforcement.

While different hate groups have dramatically different identities, with focuses ranging from opinions on sexuality and abortion, to religion, to immigration and race, hate groups tend to be tightly interconnected (see Figure 8).

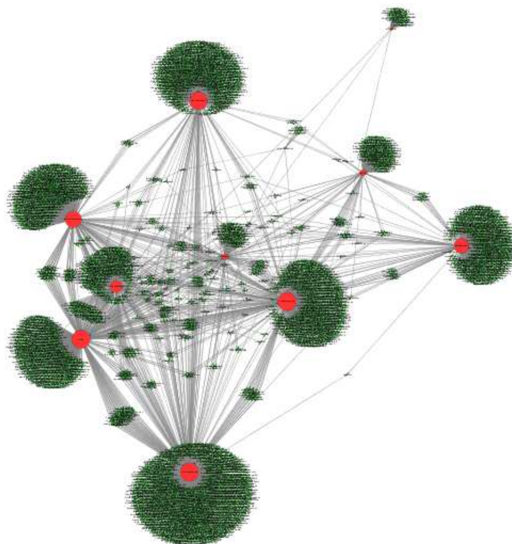


Figure 8: Hate Groups, One Layer from Seed

Remarkably, this network includes not only the white-male-Christian dominated groups described above, but also the Nation of Islam (NOI), a black Muslim extremist group. While it is somewhat estranged from the rest of the groups (NOI is the small community in the upper-right), it engages in a similar political space, and interacts with several of the same accounts as the other hate groups.

Social factions crystalize one layer further out (see Figure 9). In the retweet map, most hate groups have unified in to a single nondescript community. Two outliers exist - the *Nation of Islam* (NOI), protruding to the right, and the *Westboro Baptist Church* (WBC), near the bottom of the map.

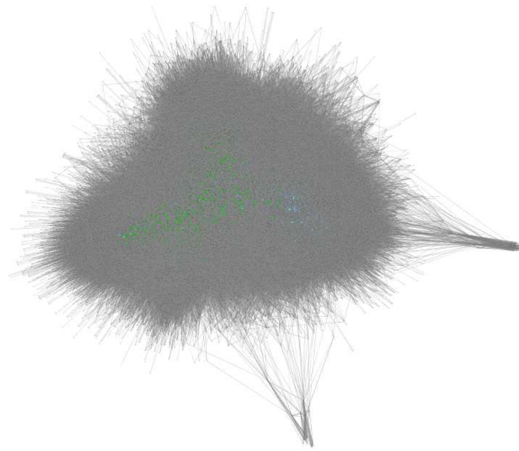


Figure 9: Hate Group Retweets, Two Layers from Seed

The large cluster makes up what is commonly refered to as *The Alt-Right*. It is an interdisciplinary coalition of activists in politics, religion, race, and sexuality, featuring unsavory elements including white supremacists, Nazis, anti-LGBT groups, and anti-immigration militias. The two outsiders to this community have distinct ideologies - the Nation of Islam has little in common with white-christian groups, and the Westboro Baptist Church is an isolationist organization which does not communicate with anyone outside the WBC. It is linked to the graph only because other hate groups retweeted messages from the WBC Twitter account.

What this suggests is that the social structure of the alt-right is similar to that of hactivist communities (Figure 2): Membership is loosely defined, group boundaries are porous, and isolating a single subcommunity is almost impossible. As in hactivism, many hate-group participants rely on anonymity to engage in socially isolating or blatantly illegal activities.

4 CONCLUSION

This study has three general conclusions regarding group membership and leadership paradigms:

1. The "default" model in Twitter-based social movements is an anarchic, self-organized design in which there are no fixed leaders, but organizers that help retweet proposals so they "ripple" through the network. This design is beneficial, because shared leadership means the movement can work on a wide array of tasks at once, and there is no single organizer or sub-group that can be dismantled to scatter the movement. However, this model is vulnerable to internal attack wherein agent provocateurs embed themselves in the movement and spread harmful messages.

2. In response to internal attack, social movements institutionalize and centralize leadership, allowing identification and excommunication of harmful individuals. This formalization of membership means a small number of individuals are "gatekeepers" to the movement, who can define who qualifies as a member. These gatekeepers are defacto leaders and organizers of the movement, and introduce the possibility for corruption, schisms, and targetting by external attackers.

3. An alternative to either organizational model is the creation of an immutable "guiding document". Anyone following the rules of the document is a member of the social movement, and anyone violating the rules is explicitly not a member. This can prevent abuse of the movement name by provocateurs without the use of an authority, but it leaves the social movement inflexible and unable to respond to

> scenarios outside the scope of its founding document. This works for single-issue social movements like *Alcoholics Anonymous* (Section 3.2), but is not applicable to broad-objectived social movements like *Black Lives Matter* (Section 3.4). To widely deploy this model would require making many small and specific social organizations that layer to create a broad social movement.

This study also has a specific conclusion regarding the use of "hybrid" leadership models: While it is possible for two leadership models to co-exist in a single social movement, it obscures the goals and methods of the movement, and can hide one side of the movement from media attention. This is based on the particular study of *Black Lives Matter* (Section 3.4), where the movement is largely referenced in media and popular discourse as a self-organized, leaderless movement, but in fact has an internal schism between self-organized and highly centralized components. The two groups have differing scopes, objectives, and methods, but share a name to the detriment of both sides.

REFERENCES

Black Lives Matter: Cincinnati. 2018. "Why Black Lives Matter: Cincinnati is Changing its Name". Visited on 03/29/2018. `http://blacklivescincy.com/home/2018/03/28/why-black-lives-matter-cincinnati-is-changing-its-name/`.

Blue, Alexis. 2018. "Researcher Looks at 'Digital Traces' to Help Students". Visited on 03/19/2018. `https://uanews.arizona.edu/story/researcher-looks-digital-traces-help-students`.

Boyd, Danah, Scott Golder, and Gilad Lotan. 2010. "Tweet, tweet, retweet: Conversational aspects of retweeting on twitter". In *System Sciences (HICSS), 2010 43rd Hawaii International Conference on*, 1–10. IEEE.

Coleman, Gabriella. 2017. "From Internet farming to weapons of the geek". *Current Anthropology* 58 (S15): S91–S102.

Dauvergne, Peter, and Genevieve LeBaron. 2014. *Protest Inc.: The Corporatization of Activism*. Polity. ISBN: 0745669492.

Gallagher, Erin. 2017. "Fake Antifa Twitter". Visited on 02/18/2018. `https://medium.com/@erin_gallagher/fake-antifa-twitter-be3a05f4d2ca`.

Gallagher, Sean. 2018. "Twitter "bot" purge causes outcry from trollerati as follower counts fall". Visited on 02/21/2018. `https://arstechnica.com/tech-policy/2018/02/twitter-suspends-thousands-of-accounts-for-bot-behavior-some-cry-censorship/`.

Hargittai, Eszter, and Christian Sandvig. 2015. *Digital research confidential: The secrets of studying behavior online*. MIT Press.

Silverman, Craig. 2017. "Fake Antifa Twitter Accounts Are Trolling People And Spreading Misinformation". Visited on 04/12/2018. `https://www.buzzfeed.com/craigsilverman/fake-antifa-twitter-accounts`.

The Mentor. 1986. "The hacker manifesto". *"The Conscience of a Hacker", Phrack Magazine, Volume One*, no. 7.

Travers, Jeffrey, and Stanley Milgram. 1967. "The small world problem". *Phychology Today* 1 (1): 61–67.

Watts, Duncan J, and Steven H Strogatz. 1998. "Collective dynamics of 'small-world' networks". *nature* 393 (6684): 440.

Welles, Brooke Foucault, et al. 2014. "Dynamic models of communication in an online friendship network". *Communication Methods and Measures* 8 (4): 223–243.

Wilson, Bill. 1953. "Twelve steps and twelve traditions". *The AA Grapevine, Inc. and Alcoholics Anonymous*.