Data Science and Public Policy
Columbia University
April 2021

Leveraging Wikipedia Data to Assess Public Awareness of U.S. Hate Groups

Data Science and Public Policy Final Report

Group 5: Louis Jarvers, Liam Kearney, Michelle Strayer, Jiaxi Liu, and Jiage Song

The years 2018 and 2019 represented ten-year highs in incidences of hate crimes in the United States.[1] Following racist narratives surrounding the COVID-19 pandemic, 2020 and 2021 likely continued this trend, demonstrating spikes in anti-Asian hate crimes across several major US cities including New York, Los Angeles, and Seattle.[2] White supremacist and far-right extremist groups have also seemingly entered public consciousness, given public displays and heightened news coverage. Legal advocacy group Southern Poverty Law Center (SPLC) maintains records on hate crimes in the United States and the activities of hate groups. In a recent report, the group noted that in 2017, 2018, and 2019, they "recorded historically high hate group numbers."[3]

Little academic literature focuses on the role of public knowledge in understanding and preventing hate crimes. Public awareness and education play a role in countering violent extremism (CVE) policies in several countries, though the nature of its inclusion varies.[4] American post-9/11 CVE efforts focused on the role of education in communities targeted for recruitment, largely as a vehicle for achieving upward socioeconomic mobility. Later efforts utilized public awareness to frustrate recruitment efforts, teaching citizens about online recruitment tools to impair their effectiveness.[5] Some CVE theory additionally posits that maintaining open dialogue around violent extremist groups and their activities hinders recruitment and radicalization.

This report focuses on the question of visualizing and measuring public awareness around US hate groups, using Wikipedia data as a proxy for public attention. We draw on previous literature that uses Wiki data to measure public attention in the fields of economics and environmental conservation[6] and publications on the applicability of Wiki data to political science research.[7] This report uses Social

---

[1] Michael Balsamo, "Hate crimes in US reach highest level in more than a decade," *AP*, 16 November 2020.
[2] Hannah Allam, "FBI Report: Bias-Motivated Killings At Record High Amid Nationwide Rise In Hate Crime," *NPR*, 16 November 2020; "Maryland U.S. Attorney's Office and FBI Baltimore Field Office Condemn[...]," The United States Attorney's Office, District of Maryland, 31 March 2021.
[3] Rachel Janik and Keegan Hankes, *The Year in Hate and Extremism 2020*, Report, Southern Poverty Law Center, 1 February 2021.
[4] See, for example: Mecklenburg, Michael Herzog Zu and Ian Anthony, *Preventing Violent Extremism in Germany: Coherence and Cooperation in a Decentralized System*, Report, Stockholm International Peace Research Institute, 2020: 8-18; Omer Taspinar, "Fighting Radicalism, not 'Terrorism': Root Causes of an International Actor Redefined," *SAIS Review* vol. XXIX no. 2 (Summer-Fall 2009).
[5] "Working to Counter Online Radicalization to Violence in the United States," The United States Department of Justice Archives, 5 February 2013.
[6] See, for example: John C. Mittermeier, Ricardo Correia, Rich Grenyer, Tuuli Toivonen, and Uri Roll. "Using Wikipedia to measure public interest in biodiversity and conservation." Conservation Biology (Conservation Methods), Society for Conservation Biology, 22 March 2021; Mirko KJampf, Eric Tessenow, Dror Y. Kennett, Jan W. Kantelhardt, "The Detection of Emerging Trends Using Wikipeida Traffic Data and Context Networks," Plos One, 31 December 2015.
[7] Denis Cohen, Nick Baumann, and Simon Munzert, "Studying Politics on and with Wikipedia," Tutorial, Methods Bites: Blog of the MZES Social Sciences Data Lab, 26 August 2019.

Network Analysis (SNA) and SPLC's list of designated hate group organizations to map the connections between hate groups. The organizations on SPLC's designation list represent first-level nodes in the network, with the individuals (people) connecting these organizations representing second-level nodes. This report then engages Natural Language Processing (NLP) and LDA topic modelling to analyze article content, and time series analysis to examine view data as a public attention proxy. Our findings emphasize the applicability of these tools to other fields and dimensions of research, and present potential policy implications for their use.

Data Source

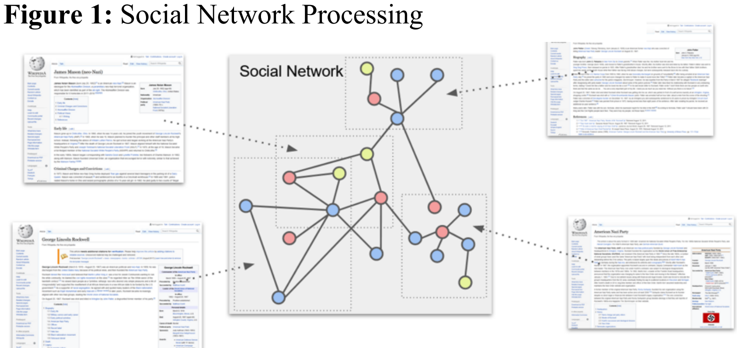
Wikipedia is a free, multilingual, open-collaborative online encyclopedia, created and maintained by a_community of volunteer contributors using a wiki-based editing system through which volunteers share their knowledge.[8] The English Wikipedia, with 6.3 million articles as of April 2021, is the largest of the 321 language editions and represents the largest general reference work on the Internet. In gauging its usefulness as a proxy for public knowledge, it is important to assess potential sources of bias, which may arise from content edits, availability, or other factors (such as Wikipedia's funding model).

Regarding bias induced by editors, although older articles lean Democratic, the overall political standpoint has become more neutral in recent years. Notably, individual articles do not seem to change their bias significantly due to revision; rather, over time, newer articles containing opposing points of view were responsible for centering the overall average. Wikipedia's revenue and balance sheets have grown steadily over time, supported by a growing pool of diverse philanthropic donations.[9] While it is not possible to know what occurs at a boardroom level, there is no immediate evidence to suggest Wikipedia's content is influenced by its donor pool. The scope of data used for subsequent analysis is also an important consideration—we chose to focus on current articles, rather than the content of their revision history.

Data Extraction

At its core, this project seeks to transform Wikipedia article content, and particularly the references between articles, into a social network. While it is theoretically possible to extract multiple entities (persons, organizations, events, locations, etc.) from each article, this visualization approach examines the article and its Wikipedia-internal links first. Accessing the Wikipedia REST API with a body of self-built R functions, we identify relevant articles (here: hate groups listed by the Southern Poverty Law Center) and pull their content, as well as their page ID, revision ID, page view, edit history, and article summary metadata. We use HTML and Wiki-text parsers to identify inter-Wikipedia references and use Wikipedia's article categorization to

Figure 1: Social Network Processing



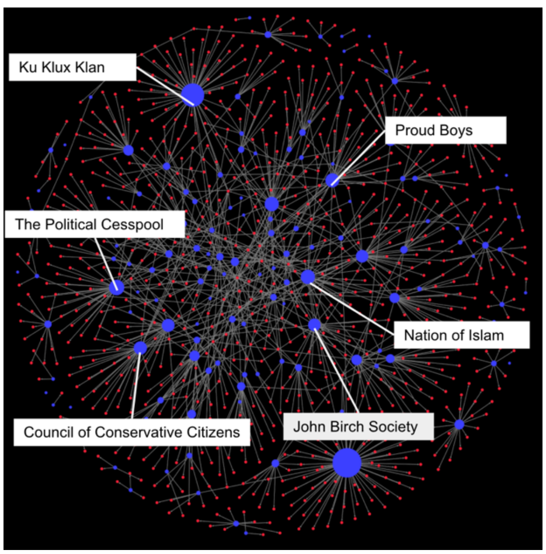
[8] For more information on Wikipedia's publicly available information on multilingualism, open collaboration, and editors, we recommend the following reference pages, respectively: https://en.wikipedia.org/wiki/Multilingualism; https://en.wikipedia.org/wiki/Open_collaboration; https://en.wikipedia.org/wiki/Wikipedians.
[9] "Financial Reports," Wikimedia Foundation (Website), 18 November 2020: https://wikimediafoundation.org/.

isolate pages on individual people. We then process the network data from the resulting articles (node list) and their references (adjacency list) via the open-source software Gephi and analyze the article content and public awareness measures (page view, edit history) with R- and Python-based libraries.

Social Network Analysis (SNA)

Using SNA with Wikipedia data allows us to visualize the focus of public knowledge. The centrality of a group or actor in the network does not necessarily reflect its centrality in real life; rather, it reflects public knowledge of interconnectedness. The following analysis explores central persons and organizations. The nodes in the network (see Figure 2) are divided into two levels: articles on hate group organizations (1st level nodes) and articles on associated people (2nd level nodes). The “edge” between two given nodes represents one article’s reference of another. With a directed network, nodes have two different degrees, the in-degree (the number of incoming edges), and the out-degree (the number of outgoing edges). In our case, the first-level nodes (hate group organizations) only have out-degree attributes, as they all point out to the second level data (associated people).

Figure 2: First Level Out-Degree Network Plot



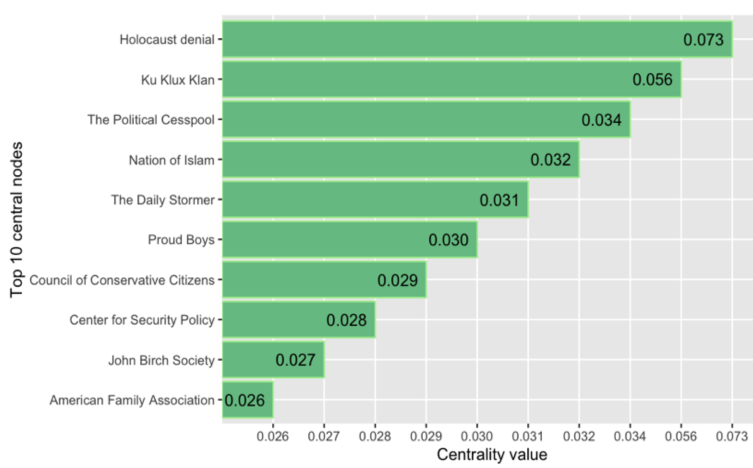
Our network is based on the Frukterman-Reingold (FR) algorithm, originating from Force-Directed Layout. The molecular mechanics model we employ is used in FR algorithms. Each vertex (node) is treated as a molecule. The vertices with edges have attractive force (*f*attr), and the vertices without edges have directly repulsive force (*f*rep). The definition of force conforms to the physical formula of intermolecular force:

*f*attr(u, v) = k^2 / distance (u, v).

We also calculated several basic social network statistics, including degree centrality, closeness centrality, and betweenness centrality (see Figure 3). Degree centrality addresses the question: "Who is the most important or central node in the network?" Closeness, by contrast, can be regarded as a measure of how long it takes to spread information from one node to all other nodes sequentially. Finally, betweenness centrality quantifies the number of times a node acts as a bridge along the shortest path between two other nodes.

The out-degree network plot (Figure 2) focuses on the number of references from first-level nodes to second-level nodes, showing more central and indirectly connected organizations placed in the center. Furthermore, the centrality value graph (Figure 3) identifies Holocaust denial, the KKK, and The Political Cesspool as the top three high centrality nodes.

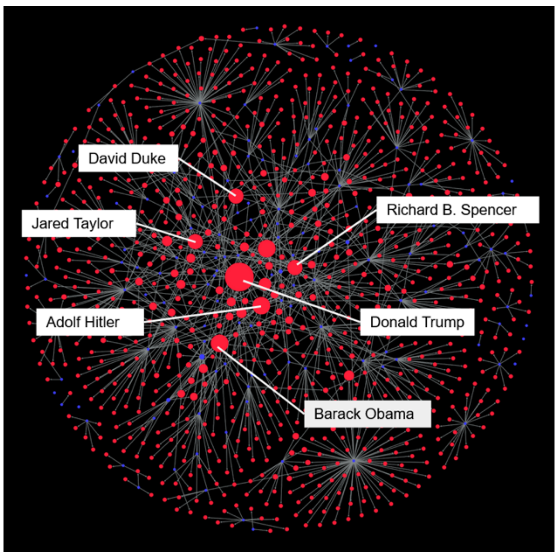
Figure 3: Centrality Value for Top 10 Central Nodes



Next, we visualize in-degree, which counts the incoming edges on a node.

Progressing the network plot from the first level (organizations) to the second (people), we see that ranking by node size identifies prominent individuals referenced in hate group articles. The results show a mixture of politicians (Donald Trump, Barack Obama, Hillary Clinton), historic figures (Adolf Hitler) and far-right activists (David Duke, Jared Taylor, Richard Spencer)—see Figure 4, below. The in-degree visualizations below, here paired with a Frutcherman-Reingold algorithm that sorts the most connected nodes to the center and the least connected to the periphery, identify central actors. Featuring former U.S. President Donald Trump as the most central person in the network raises further questions about interaction type. Here, a juxtaposition of in-degree and pageview might help to separate central actors involved in hate groups from any actors identified as central because of their prominence beyond the hate group milieu.

Figure 4: Second Level In-Degree Network Plot
Hate group network with Fruchterman-Reingold algorithm and node size by in-degree [red: persons, blue: organizations]



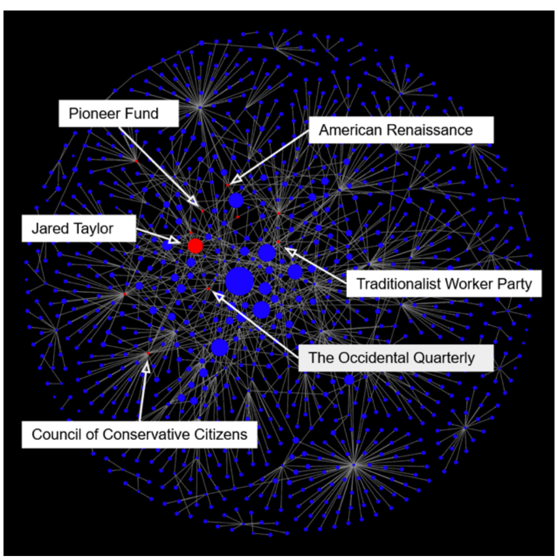


Figure 5: Jared Taylor Focus Plot
Highlighting the network around white supremacist Jared Taylor [red: Taylor and ten referencing organizations, blue: other edges]

Central persons described via article content as far-right or white supremacist activists are also of key interest. In the example of Jared Taylor (Figure 5), an American white nationalist as classified by the SPLC, we observe ten incoming edges from hate groups like the Council of Conservative Citizens and The Occidental Quarterly or American Renaissance. The edges denote that these organizational articles reference Taylor and establish a relationship between organization and activist; however, we are not (yet) able to typify that relationship (e.g., founder, member, influence, opponent, quoted author, etc.). On an aggregate level, the network successfully identifies numerous key far-right actors between U.S. hate groups. The next section of this report looks more closely at their role.

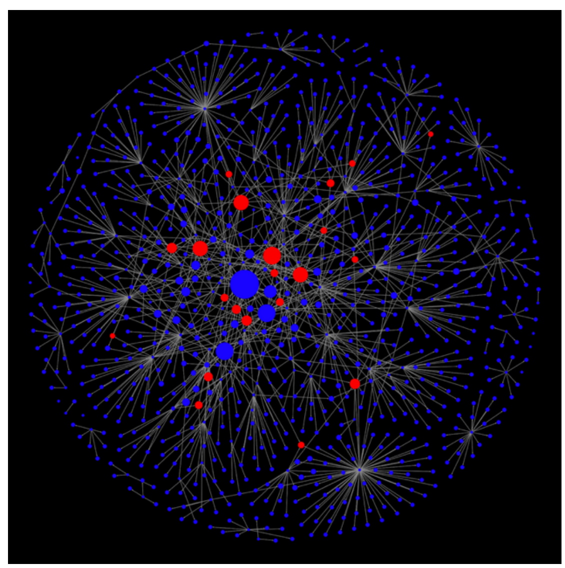


Figure 6: Overall Degree Plot of First and Second Level Far-Right Nodes
Overall degree visualization of first and second level nodes, highlighting central far-right activists in the network [red: far-right activists, blue: other nodes (persons/organizations)]

In the last step of the network analysis, we focus on organizations and actors in the periphery of the network. Instead of highlighting central actors, we evaluate which groups and persons are least connected. Figure 7 shows that religious hate groups (as classified by the SPLC) are pushed to the periphery of the network and do not appear to be heavily connected. As the Wikipedia SNA reflects public knowledge of these groups, and not necessarily their actual behavior, this trend could reflect a lack of real-world organizational networking or could be a result of missing information on Wikipedia.

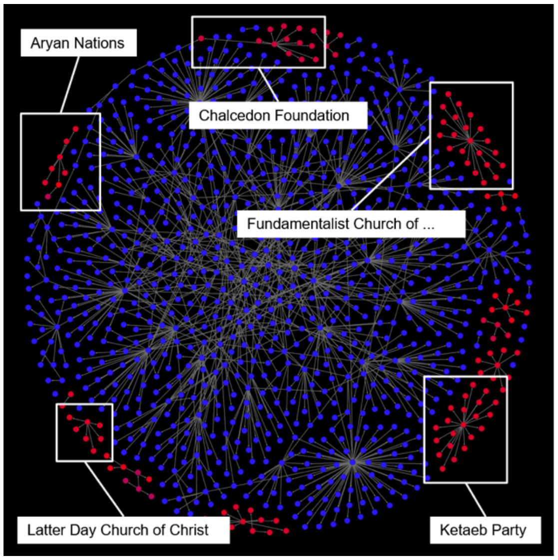
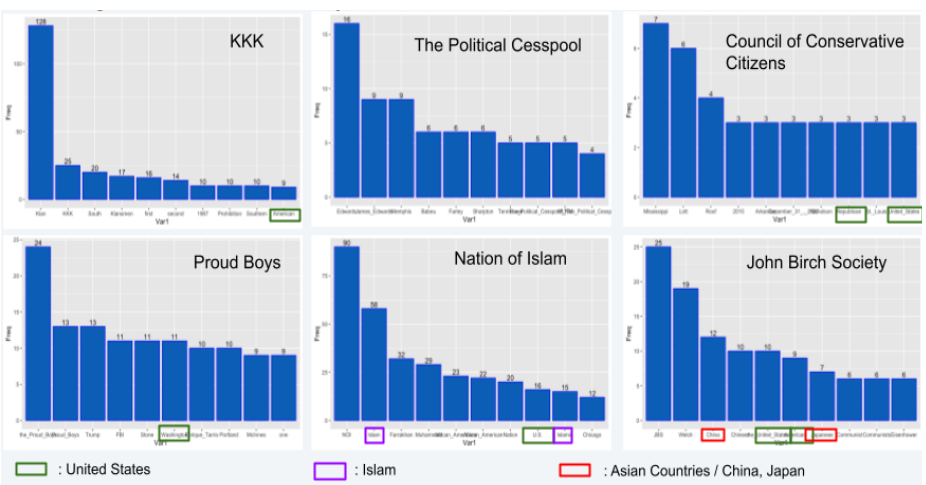


Figure 7: Peripheral Nodes Plot
Visualization of peripheral organizations and persons in the network, showing least connected entities [red: outside organizations and persons, blue: other nodes]

Natural Language Processing (NLP) Analysis

In this initial NLP analysis, we look into top in-degree and out-degree network nodes, and examine the top frequency words that appeared in their respective Wikipedia pages. Moreover, we also analyze common words between them. This analysis thus seeks to identify patterns in hate group article content, thereby detecting potential patterns in group activities or ideologies.

Figure 8: Out-Degree Network NLP Analysis Graph

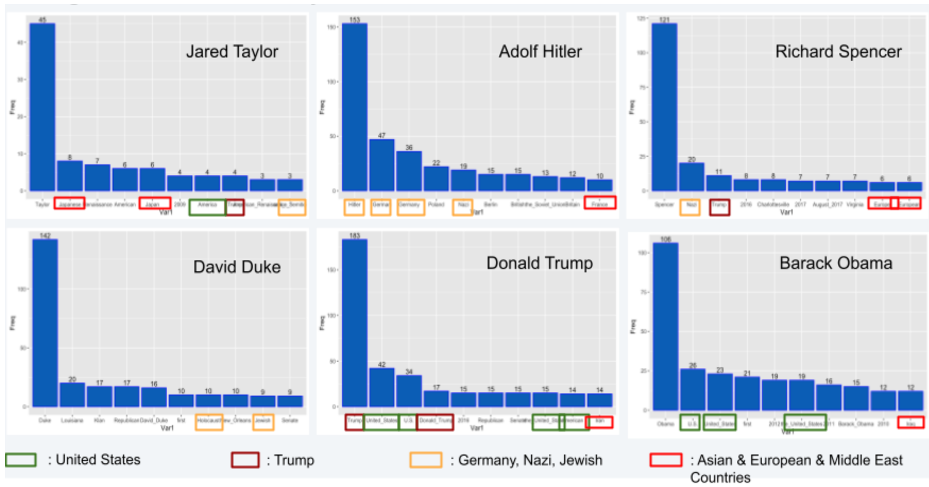


Beginning with the out-degree network, we select nodes with an out-degree of 28 or more. Six labels match this criteria (see Figure 8, below). The graph lists the top ten frequency word groups within these articles, including a number of overlapping words. For example, "United States" appears five times, while words relating to Islam or to Asian countries appear more than twice.

We then examined in-degree networks, selecting nodes with an in-degree of 10 or more (ie. referenced in 10 or more articles). The resulting word labels fall into six groups, as shown in Figure 9, on the next page.

The in-degree graph (Figure 9) shows a greater number of overlapping words. For example, word groups about Trump appear 8 times, while word groups about the United States appear 9 times. Additionally, the top frequency words contain a wider range of geographical regions compared to the out-degree network, including European and Middle Eastern countries. Interestingly, while "Nazi"

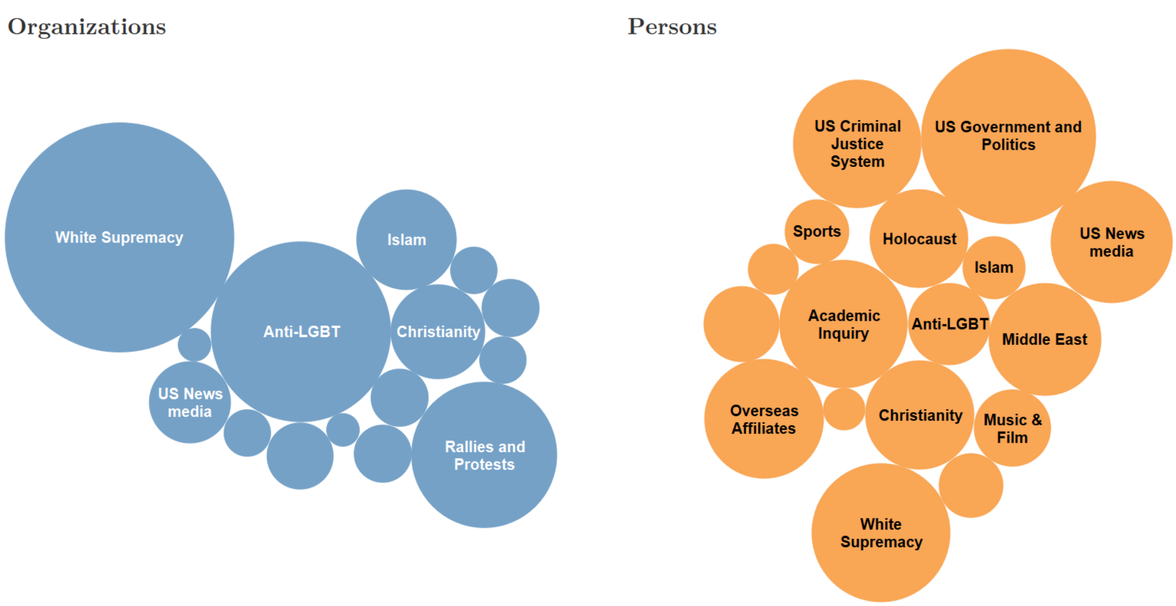
Figure 9: In-Degree Network NLP Analysis Graph



related keywords did not appear in the out-degree network analysis, there are many words in the "Nazi" group in the in-degree network analysis.

We continue to use NLP to explore larger tends, examining features of the article texts in aggregate. First, we applied Latent Dirichlet Allocation (LDA) to automatically determine the main topics of the text. An optimal model was chosen based on both the statistical optimization procedure of Griffiths et. al (2004)[10] and qualitative inspection of resulting topics, varying K (number of topics assigned) and the feature selection thresholds (minimum term frequency and maximum document frequency). Using K=24 and assigning topic labels based on a thematic interpretation of the words in each group, the most likely topics for each article are visualized below (split by organizations and persons, and merging similar topics). In addition to expected topics such as white supremacy and anti-LGBT, the LDA revealed topics including "science" (prominent thinkers giving intellectual ballast to hate groups), "sports", "music and film" and "US news media". These themes may be important directions for further research.

Figure 10: Latent Dirichlet Allocation (LDA) Article Topics Chart
Visualization of topics using K = 24 (assigned article topics), split by article classification (organization, person). Size of bubble denotes number of articles.



[10] Thomas Griffiths and Mark Steyvers, (2004), "Finding Scientific Topics," *Proceedings of the National Academy of Sciences of the United States of America*, 101 Suppl 1: 5228-35.

We also utilized keyness analysis to identify relatively more frequent words in target versus reference article groups (see Figure 11, below). As expected, "white" and "Aryan" are more common in white supremacy articles relative to other articles in the network; interestingly, so are words related to anti-immigration ideology. There also appears to be a relative increase in the use of "terror" and "terrorist" in articles created in 2017 or later.

Figure 11: Keyness Comparison Charts

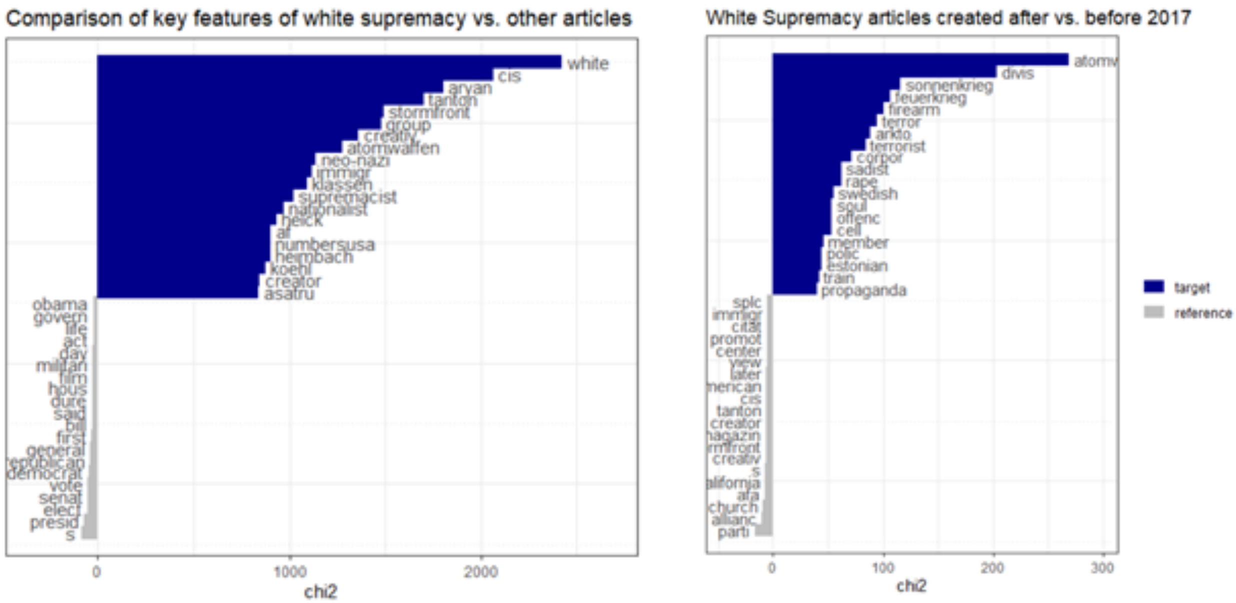
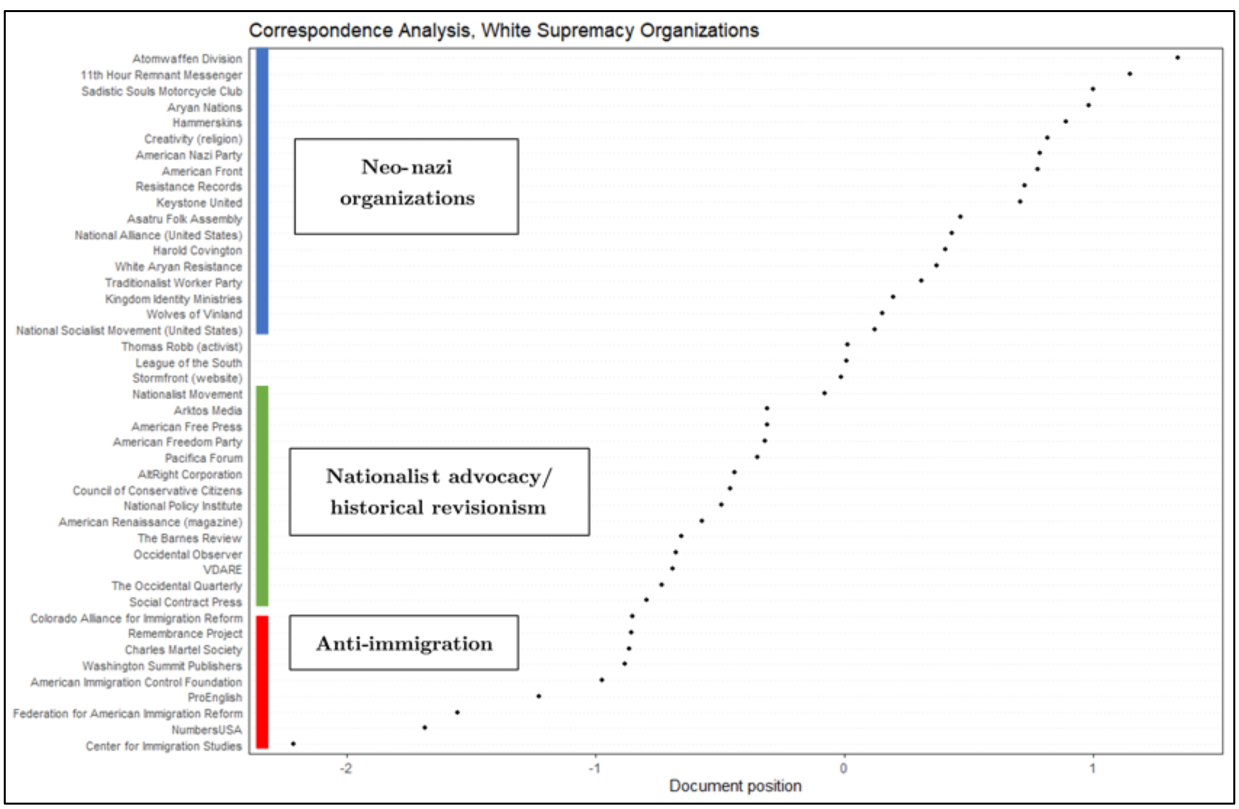
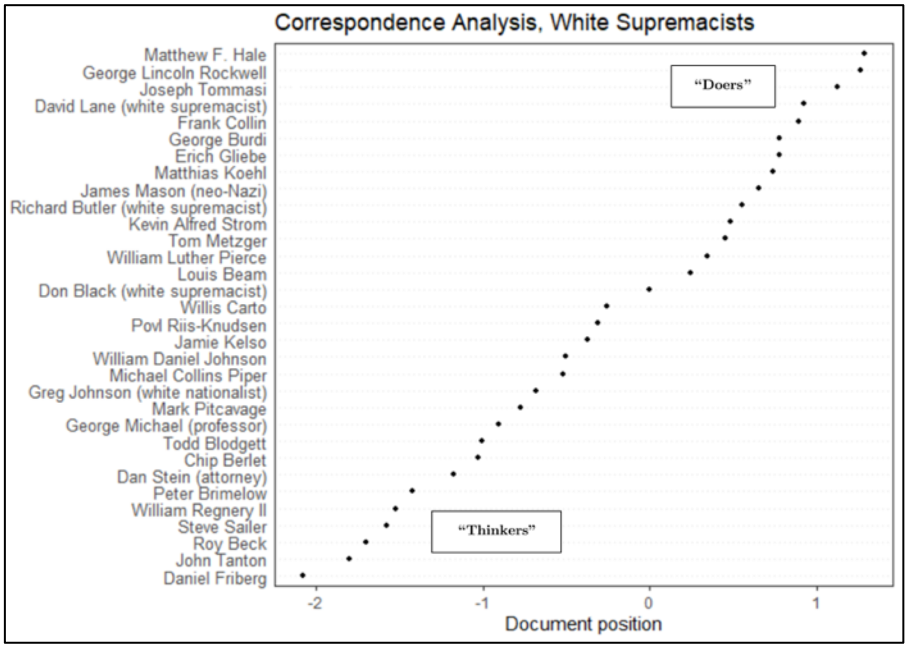


Figure 12: First and Second Level Correspondence Analysis Graph: White Supremacy Organizations



Finally, we employ correspondence analysis to visualize the positions of articles along a single dimension, split by persons and organizations (see Figure 12, on the previous page, and Figure and 13, below). Correspondence analysis (a technique similar to principal component analysis for categorical data) allows us to map public knowledge of individual people and organizations along a single spectrum without imposing theoretical constraints (i.e. based solely on the article texts). Combined with LDA, this helps identify key properties of subgroups within the network. For persons classified as "white supremacists", the document position maps loosely from alt-right and anti-immigration thinkers and writers (-2) to neo-Nazi leaders and activists (+1); for classified white supremacist organizations, the document maps from anti-immigration think tanks (-2) to nationalist advocacy and historical revisionist media (-1 to 0) and on to neo-Nazi criminal and terrorist organizations (0.5 to 1). Figures 12 and 13 thus reflect public knowledge of white supremacist ideology.

Figure 13: First and Second Level Correspondence Analysis Graph: White Supremacist Individuals



Time Series Analysis

Given available Wikipedia data, we created a dataset that tracks Wiki page views over time for the first- and second-level network nodes. Using page views as a proxy for public attention, we are able to examine shifts in public awareness trends over time. The following analysis focuses on temporal trends in public awareness of far-right hate and extremist groups, as a subset of the hate group population.

Figure 14, on the next page, shows the time series distribution of view data for 11 high-indegree far-right associated pages in the second level of our network.[11] The graph shows frequent fluctuations in views for individual pages, with the lowest monthly average (Tom Metzger, December 2016) under 10, and the highest monthly average (Anders Behring Breivik, October 2018) over 32,000. Figure 15, on the next page, shows the time series distribution of the same 11 pages, using individually standardized monthly average view scores in place of average monthly view count.[12]

---

[11] We isolated high-indegree (in this case, three or more connections) far-right associated second level nodes by filtering their Wiki page descriptions according to the presence of far-right terms such as 'neo-Nazi', 'right-wing', 'KKK', and 'white supremacist'.

[12] To examine these trends more closely, we standardized average monthly view data by creating individually standardized view scores for each page.

Figure 14: High Indegree View Time Series plot (annotated)

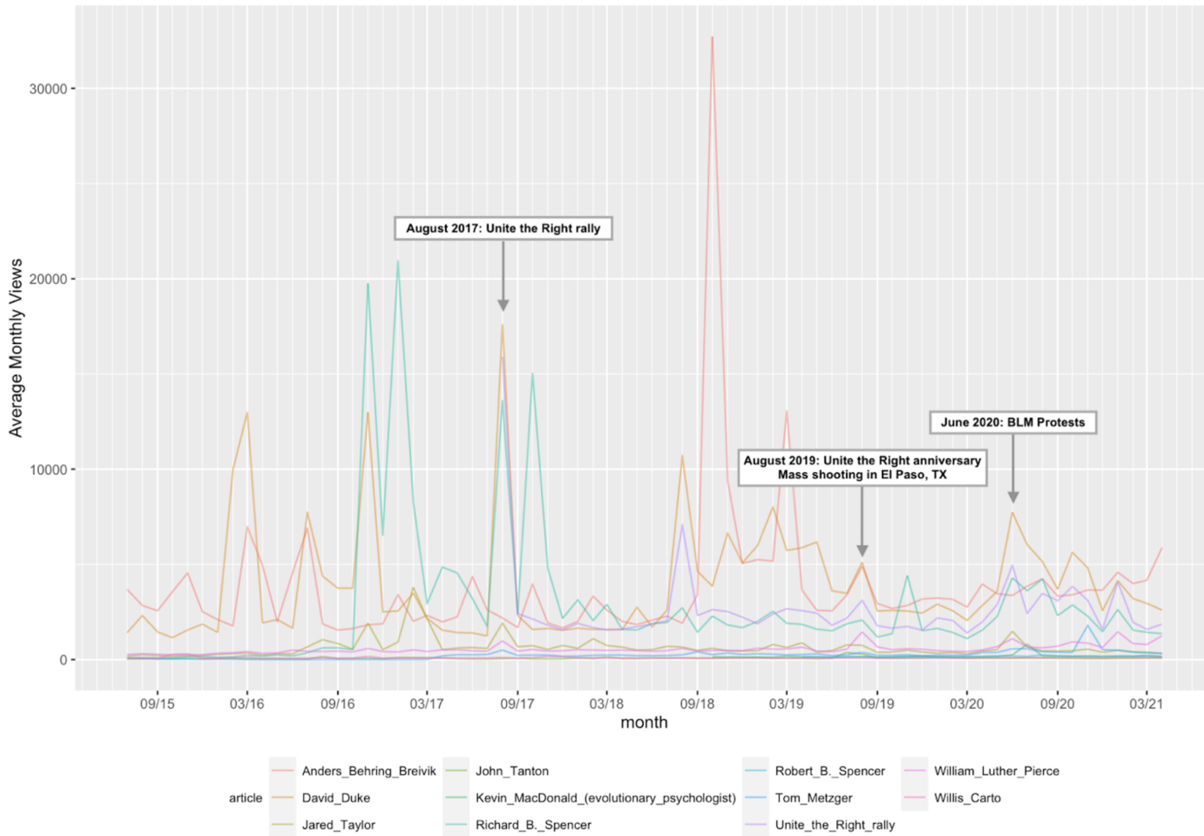
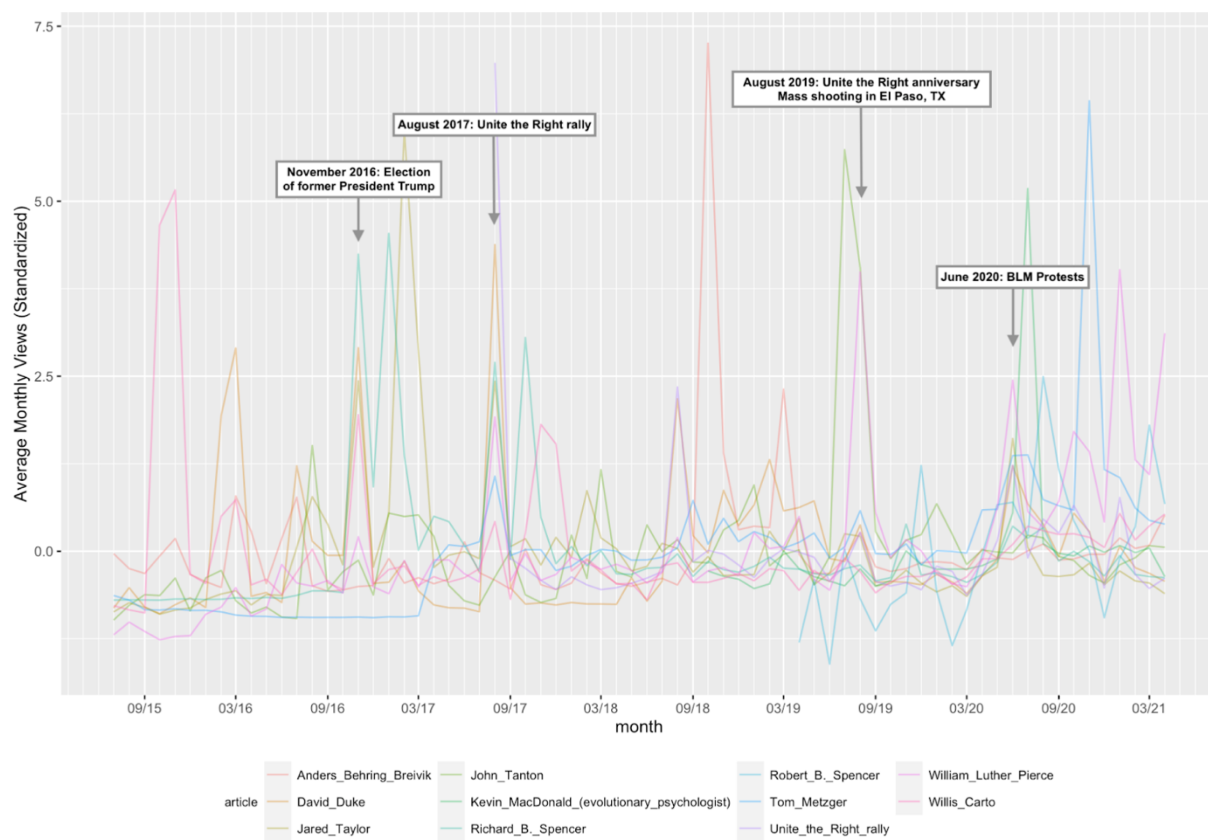


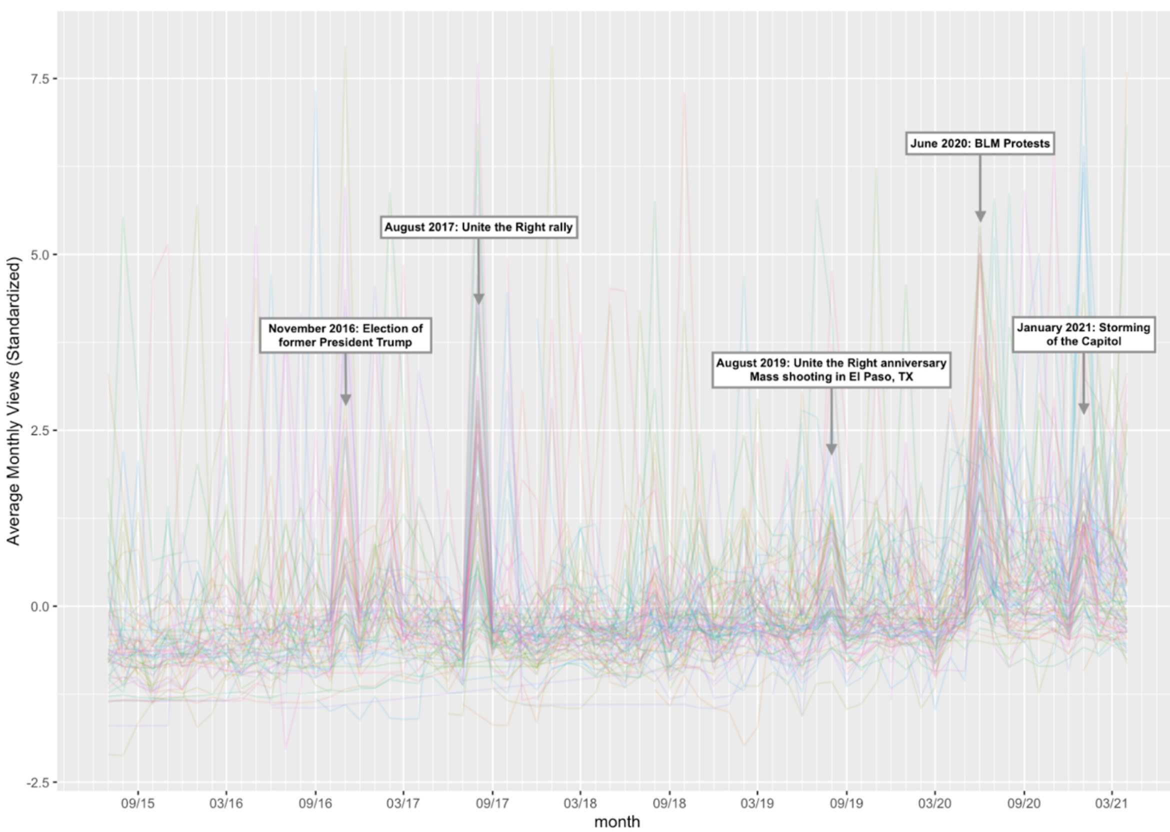
Figure 15: Standardized High Indegree View Time Series plot (annotated)



Figures 14 and 15 demonstrate similar trends across individual view data; for example, Anders Behring Breivik's distribution (red/orange) shows similar fluctuation, with a clear outlying peak in October 2018 and a subsequent smaller peak in March 2019. Further, the graphs show several periods in which all eleven pages demonstrate view increases: August 2017, August 2019, and June 2020 (annotated). The standardized time series (Figure 15) additionally shows a peak in November 2016 (annotated). The aforementioned annotations highlight these periods and note corresponding incidences of widely-reported events and far-right activity. These include the Unite the Right rally (August 2017), the 2019 mass shooting in El Paso, TX (August 2019),[13] the Black Lives Matter or George Floyd protests (May 26, 2020 through August 2020) and the U.S. election of former president Donald Trump (November 2016).

Figures 16 (below) and 17 (on the next page), show standardized average monthly view data across far-right associated pages in the first- and second-level network nodes. In expanding the page sample, we examine trends across both personal and organizational page types; Figures 16 and 17 map standardized view data for the resulting 89 pages.[14]

Figure 16: Standardized View Time Series plot (annotated)

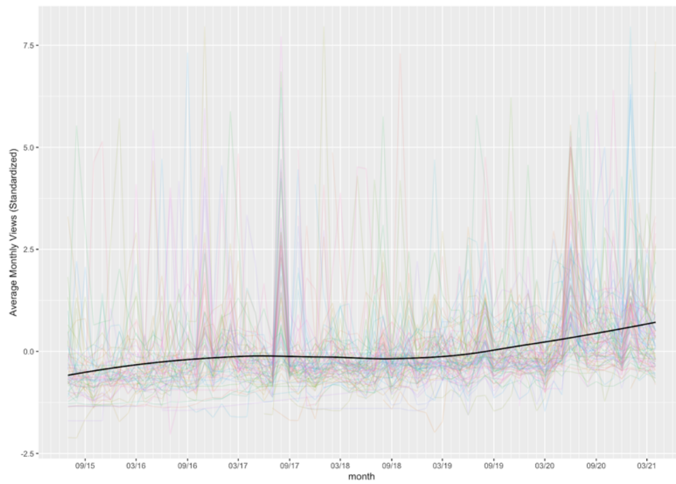


[13] The August 2019 rise coincides both with the two-year anniversary of the Unite the Right rally and a mass shooting in El Paso, TX. The El Paso shooting is currently considered an act of far-right terrorism, though as a result of U.S. domestic terrorism laws, it will be tried as a racially motivated hate crime. (See: Smith, Rojas, and Robertson, "Dayton Gunman Had Been Exploring 'Violent Ideologies,' F.B.I. Says," *The New York Times*, 6 August 2019.

[14] In order to expand our page sample and examine potential trends across organizational page types, we ran page descriptions for the entire first- and second-level network through the same far-right filter.

Figure 16 shows four clear spikes: November 2016, August 2017, June 2020, and January 2021 (annotated). The modest August 2019 rise is more subtly evidenced (annotated). Unlike Figures 14 and 15, Figure 16 shows a clear rise in public attention with the storming of the U.S. Capitol building in January 2021. Moreover, the period between May 2020 and March 2021 evidences a higher concentration of individual page spikes, suggesting more peaks in public attention to individual far-right actors or groups. Figure 17 further shows an overall increase in views after 2019; the trendline in Figure 17 highlights this trend, demonstrating a gradual rise and decline between approximately September 2015 and September 2018, followed by a clear positive trend in page views beginning around October 2018 and lasting through March 2021 (the end of the recorded data).

Figure 17: Standardized View Time Series plot with trendline



In sum, time series analysis suggests trends in public attention to far-right hate groups rise alongside relevant major national events, like the Unite the Right rally, the BLM protests (and counterprotests), and the Storming of the Capitol. Further, overall public attention to far-right groups seems to be increasing over time.

Findings

Through several types of analysis, this report identified a number of trends among hate groups. Social Network Analysis (SNA) visualized the focus of public knowledge on hate group interconnectedness, highlighting far-right and white supremacist groups as the most central, and Christian hate groups as the most peripheral. Natural Language Processing (NLP) isolated divergent word groupings between in- and out-degree networks. In addition, Latent Dirichlet Allocation (LDA) identified expected classifications amongst organizations and persons (white supremacy, anti-LGBT, rallies and protests, etc.), and helped reveal other important themes in the network, including U.S. domestic politics, criminal justice, overseas affiliates, music and film, and academic inquiry. Further visualization through Correspondence analysis mapped public knowledge of the ideological scale of hate groups. Finally, time series analysis suggested public attention of far-right and white supremacist hate groups rises around associated major national events, and noted an overall increase in public awareness over time.

Policy Implications and Next Steps

Wikipedia represents not only an encyclopedic reference, but also a valuable tool for assessment of public attention. As public awareness plays a vital role in shaping policy, Wikipedia can inform policy implementation and communication efforts. Social network Wiki analysis can also serve as a starting ground to identify common denominators between topics (evidenced, this report, through its usefulness in detecting persons between hate groups). Wiki data SNA can also help identify group clusters and degrees of group centrality for certain organizations.

Given the time limitations of this project, our group intends to continue our analysis in future. We plan to expand our work to compare U.S. public attention to domestic vs. international hate groups. Further, as Wikipedia supports multiple languages, we plan to compare hate group types in English and German, focusing on public awareness in the United States and Germany. Finally, we plan to utilize news monitoring platforms (ie. Factiva) to evaluate the strength and duration of public attention spikes given news coverage.

Importantly, we finally note suggestions for further research outside the scope of our analysis. Further research might engage Google Trends data as a separate or additional platform. Research might also engage further languages to expand scope, or utilize geographic analysis via geographic Wikipedia trends data to more precisely measure public attention trends in certain regions.

Contribution Statement

This report, and its preceding presentation, were produced with a joint, collaborative effort on every level. Every group member was involved in analysis, data collection, coding, visualization creation, presentation, and composition of the final report.