Analyzing a Fake News Authorship Network

Cody Buntain[1,2], Jennifer Golbeck[2], Brooke Auxier[2], Biniyam Girum Assefa[2], Karen Boyd[2], Kristen M. Byers[2], Gursimran Chawla[2], Daniel Chen[2], Benjamin J. Cooper[2], Jake Cupani[2], Clay Daetwyler[2], Nicholas DeWitt[2], Suzanne Garcia[2], Christine Hafer[2], Misbah Khan[2], Elo Lewis[2], Marianna J. Martindale[2], Matthew Louis Mauriello[3,4], Helen McNamara[2], Sean Alexander McWillie[2], Daniel Millay[2], Talal Munzar[2], Sean Mussenden[2], Nicholas Orji[2], Lisa Phung[2], Kristine M. Rogers[2], Christopher Anton Rytting[2], Tuba Shadan[2], Subhatra Sivam[2], Koralleen Stavish[2], Aditya Subramanian[2], Sai Tipirneni[2], Rrahul Topiwala[2], Melissa Wagner-Riston[2], Peratham Wiriyathammabhum[2], and Frazer Workneh[2]

[1] New York University, New York NY 10012 cbuntain@nyu.edu
[2] University of Maryland, College Park, College Park MD 20742
[3] Stanford University, Stanford CA 94305
[4] Oregon State University, Corvallis, OR 97331

Abstract. This paper analyzes 246 fake news websites previously identified in three research projects. From this dataset, we extract a set of authors who have written for these sites in 2016, which we make publicly available. Applying a novel shared authorship construct, we analyze a network of fake news sites. This analysis shows a tight cluster of sites, with a trend of article reposting, wherein sites copy content from each other but preserve author bylines. We also show the most central authors, while associated with different sites, share common affiliations with a single site: Infowars.

Keywords: fake news · author networks

1 Introduction

Recent research has studied fake news in online spaces, showing these alternative sources are often well-separated from mainstream media [5]. Such research often focuses on connections among sites, where connections are defined as in-article citations or common users in social networks. In this paper, we instead expose a new network among these sites by studying shared authorship (i.e., two sites are connected if an author writes for both sites). From this view, we identify a trend of article reposting, wherein sites copy content from each other but preserve author bylines. We also study influence among these sites and show the most central authors, while associated with different sites, share common affiliations with a single site: Infowars. After a thorough search, we have identified authorship information for 164 sites out of the 246 on our original list and make this dataset publicly available.[5]

[5] https://github.com/cbuntain/FakeNewsIConf2019

2 Dataset Construction and Author Collection

To construct a network of fake news sites, we synthesize a new dataset of 246 fake/alternative-news sites from recent studies: Golbeck et al. [3], Guess et al. [4], and Starbird [5]. These datasets contain 96, 92, and 79 fake or alternative sites respectively, and we find little overlap across these datasets, with any pair of datasets having no more than nine sites in common. For each of these sites, we have identified authors who published articles during 2016. To generate this author list, a researcher reviews each site to identify articles written during 2016 and the associate authors by the article's byline. Some sites are no longer accessible, however, and for those sites, we rely on the Internet Archive's snapshots to find articles and authors. Article bylines are also sometimes omitted or are not full names, as one might expect from a mainstream media source. Instead, some bylines are screen names (e.g., "admin" or "TASS"), and others are generic organizational names (e.g., "Tax Justice Network").

After collecting these site-author pairs, we construct a bipartite network of fake news sites where a node represents an author or a domain, and edges connect authors and sites based on whether that author has written for that domain. From this bipartite network, we construct two projection networks: one of sites such that two sites are connected if they share a common author, and another of authors where authors are linked if they write for a common site. Edges in these projects are weighted by the number of common authors/sites respectively. For some sites, we have no authorship data, and thus they have no edges and are excluded from our network.

After collection, we have identified authorship information for 152 of the 246 sites, finding 6,387 unique author-site pairs and 4,966 unique authors. After removing isolated nodes, the site-to-site network contains 76 sites and 427 edges, and the author-author network contains 4,929 authors and 1,924,813 edges. In analyzing these networks, we rely on three metrics: betweenness, closeness, and eigenvector centrality [2].

3 Analysis

The site network, shown in Figure 1, has one main cluster, but filtering by edge weight reveals two major collaboration networks: between Geopolitical.ru and katehon.com and four sites around The Millennium Report. Geopolitical.ru and katehon.com share 139 authors, and while they do link with many other domains, these additional ties are much weaker.

Separately, The Millennium Report at themillenniumreport.com has over 80 common authors with theeventchronicle.com, globalresearch.ca, activistpost.com, and sgtreport.com. These domains also have strong links between one another, and theeventchronicle.com shares authors with many of the other domains in our dataset. Network statistics reflect the important role of themillenniumreport.com, as it has the highest betweenness centrality (0.33) and harmonic closeness centrality (0.87), meaning many sites link to it and many shortest paths flow

through it. Most of these sites also aggregate and re-post content without authors' permission. Thus, shared authors may be an indication that the sites are interested in the same topics and re-posting common articles rather than authors choosing to write for the websites.

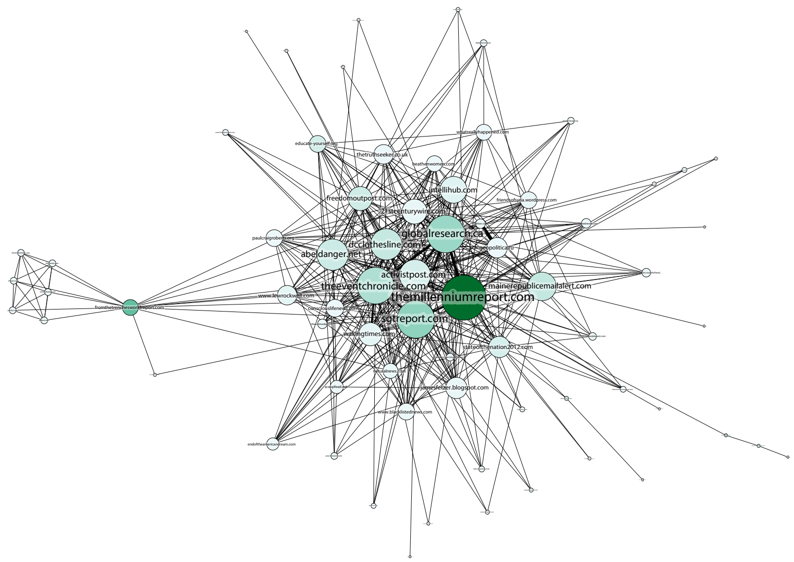

Fig. 1. The network of domains linked by shared authors. Nodes are sized by degree and colored by betweenness (darker indicates higher betweenness). The dark green node in the lower right of the main cluster is The Millennium Report.

This central network illustrates the interconnected nature of the fake news ecosystem. While many sites share only one author with another site, 126 pairs share 5 or more authors (see Figure 2). Additionally, as preferential attachment would suggest [1], only a few sites are responsible for the majority of connections. Removing the 14 nodes with the highest degree removes 83% of the edges.

Initial investigation suggests this core set of sites is interlinked not because they have many of the same authors on their payroll, but rather because they freely post content taken from other sites. They appear to be finding posts on smaller sites, adding their own headline, posting it, while still crediting the original author. This result suggests the small core of sites with many shared authors has many common topics of interest and shares posting behavior, but not that they have large shared staffs of writers producing unique content.

Examining the author projection, this network is much denser than the site network (0.15 vs. 0.04 respectively), a result of the high connectivity in sites, where each site has an average of 42 associated authors. The top five most central authors (by eigenvector centrality) are Christina Sarich, Claire Bernish, Kurt Nimmo, Jon Rappoport, and Justin Gardner. These authors have all written for Infowars, a far-right site, suggesting it exerts significant influence.

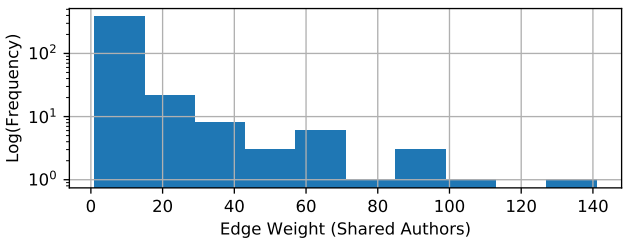

Fig. 2. Distribution of shared authors across domains, presented in log-scale.

4 Conclusions

In this project, we investigate 246 fake news websites synthesized from three extant datasets [4,3,5]. We manually identify authors who have written for these sites during 2016 and release this dataset of author connections as part of our contribution in this work. We then construct a network of fake news sites based on shared authors, finding a single dense cluster with a small core set of sites with dozens of authors in common. Sites in this core cluster appear to post content taken from other sites, preserving author bylines, creating an artificially dense network of shared authorship from borrowed content. Analysis of the most popular authors suggests high influence of the alt-right platform Infowars, where all of the most central authors have published.

References

1. Albert, R., Barabási, A.L.: Statistical mechanics of complex networks. Reviews of modern physics **74**(1), 47 (2002). https://doi.org/10.1016/j.jwb.2010.05.014
2. Chakrabarti, D., Faloutsos, C.: Graph Mining: Laws, Tools, and Case Studies, vol. 3. Morgan & Claypool Publishers (2012). https://doi.org/10.2200/S00449ED1V01Y201209DMK006
3. Golbeck, J., Mauriello, M., Auxier, B., Bhanushali, K., Bonk, C., Bouzaghrane, M., Buntain, C., Chanduka, R., Cheakalos, P., Everett, J., Falak, W., Gieringer, C., Graney, J., Hoffman, K., Huth, L., Ma, Z., Jha, M., Khan, M., Kori, V., Lewis, E., Mirano, G., Mohn, W., Mussenden, S., Nelson, T., Mcwillie, S., Pant, A., Shetye, P., Shrestha, R., Steinheimer, A., Subramanian, A., Visnansky, G.: Fake news vs satire: A dataset and analysis. WebSci 2018 - Proceedings of the 10th ACM Conference on Web Science pp. 17–21 (2018). https://doi.org/10.1145/3201064.3201100
4. Guess, A., Nyhan, B., Reifler, J.: Selective exposure to misinformation: evidence from the consumption of fake news during the 2016 U. S. presidential campaign. Tech. Rep. 682758, European Research Council (2018), https://www.dartmouth.edu/~nyhan/fake-news-2016.pdf
5. Starbird, K.: Examining the Alternative Media Ecosystem through the Production of Alternative Narratives of Mass Shooting Events on Twitter. International AAAI Conference on Web and Social Media (ICWSM) (2017), http://faculty.washington.edu/kstarbi/Alt_Narratives_ICWSM17-CameraReady.pdf