Mapping the echo-chamber: detecting and characterizing partisan networks on Twitter.

Armineh Nourbakhsh, Xiaomo Liu, Quanzhi Li, Sameema Shah

Research and Development, Thomson Reuters
{armineh.nourbakhsh,xiaomo.liu,quanzhi.li,sameena.shah}@tr.com

1 Background

The phrase "echo-chamber effect" has been widely used to describe the way modern media, especially social networks, have reshaped political discourse. Instead of diversifying the political landscape, modern media is blamed for facilitating the formation of like-minded cliques that have little exposure to alternative opinions. This phenomenon has been studied prior to the rise of social networks. For instance, Gilbert et al. [4] showed that the majority of comments posted on blogs tended to agree with the content of the blog post. Social networks only seem to have aggravated the problem. A study by the think tank *Demos* showed that politically vocal accounts on Twitter largely interacted with like-minded accounts, and shared content from domains that were agreeable to their party [10]. Benkler et al. [1] showed that link-sharing behavior reflected a polarization of political debates during the 2016 presidential elections in the U.S. They also argued that there were differences between the way right- and left-wing media motivated discussions on Twitter and Facebook.

Partisan networks are often oblivious to, or openly hostile towards, alternative perspectives. This renders them prone to spreading misinformation. The guarded nature of the networks helps sustain the propagation of false rumors even after they have been debunked by authoritative sources [14, 15], or even when the rumors' initiators admit to their mistake [12]. They are also susceptible to deliberate attempts by fake news websites, professional trolls and bots who spread misinformation [5, 16]. Identifying such networks can help researchers and analysts locate potential sources of false rumors and conspiracy theories. Studying these networks can also uncover the dynamics of the evolution and operation of the so-called professional fake news industry [6, 3].

2 Problem Statement

Studies focusing on the detection of ideologically biased or politically partisan ideas on social media often follow a supervised model of tone or stance classification [8, 9]. Other studies use network indicators such as clique detection or co-linkage to identify echo-chambers [1]. Our goal is to propose a model for mapping out the partisan networks that satisfies the following criteria:

- Can identify latent networks or communities that engage in partisan discourse (i.e. echo-chambers) without any need for supervision.
- Given an account, a domain, or a sub-network, can identify the level of *bias* or *partisanship* indicated by the homogeneity of content.
- For a given echo-chamber, can automatically identify topics or phrases that the community is vulnerable to spreading misinformation about, by analyzing the distribution of vector representations of messages.
- Does not require any manual labeling of data.
- Is ideology-agnostic, i.e. does not use human labels to identify left-learning and right-leaning accounts.
- Instead of relying entirely on content-based or network-based signals, combines them in a bootstrapping fashion.
- Can be applied in settings where the network is only partially observable, e.g. in studies where the full Twitter firehose is not accessible.
- Can be used to discover partisan sources outside of Twitter (e.g. websites or accounts on other social media) via citation analysis.

3 Methodology

We use the Kaggle fake news dataset[1] as a seed set to collect messages from Twitter. The dataset contains a list of webpages that were crawled between 2016-10-25 and 2016-11-25. Each page is labeled based on its content, using the B.S. Detector plugin for Chrome[2].

We extract the domains cited in the dataset. For each domain, we download tweets from users who have cited the domain at least once during the same period. If there is an indication that an account is affiliated with the domain (e.g. if the domain is mentioned in the user profile or if the user handle is the domain's official handle), the account is removed from the dataset. This resulted in 25,742,785 tweets from 131,552 unique accounts.

Using this dataset, we propose a model for mapping out and characterizing echo-chambers. First, we extract all domains cited within each account. We then calculate two metrics:

1) Bias of a domain: This is calculated based on the similarity of a given domain to the domains included in the Kaggle dataset. Similarity is measured as co-linkage by the same user. So if D^j is the set of domains cited by user j, then for domain d_i, bias is defined as:

$$closest\_kaggle\_domain(d_i) = \arg\max_{d_k} \sum_j Count(d_i \in D^j \wedge d_k \in D^j) \quad (1)$$

$$bias(d_i) = closest\_kaggle\_domain(d_i) * score(closest\_kaggle\_domain(d_i)) \quad (2)$$

[1] https://www.kaggle.com/mrisdal/fake-news
[2] http://bsdetector.tech

where d_k is a domain from the Kaggle dataset, and $score(d_k)$ is the number of times d_k appears in the Kaggle dataset.

Table 1 lists a few domains, their closest Kaggle domain, and their normalized bias score based on the above formula.

Table 1. A few example domains and their standardized bias scores.

Domain	Closest Domain from the Kaggle dataset	Bias
`newsbusters.org`	`breitbart.com`	0.13
`wakeup-world.com`	`collective-evolution.com`	0.03
`truthuncensored.net`	`truthfeed.com`	0.012
`news.groopspeak.com`	`occupydemocrats.com`	0.5

2) Partisanship of an account: This is calculated based on the similarity of domains shared by the same user. We first calculate global co-occurrence counts for every pair of domains. Each user's partisanship is the sum of the pairwise co-occurrence counts of the user's domains. So for a given user u_z partisanship is defined as:

$$partisanship(u_z) = \sum_{l,m} \sum_{j} Count(d_l \in D^j \wedge d_m \in D^j) \text{ for all } d_l, d_m \in D^z \quad (3)$$

Table 2. A few example users and their standardized partisanship score.

User	Partisanship
sessie03	0.107
tonitrivi	0.059
FordCrews	0.024
foodsociety2011	0.001

Note that an account's partisanship is not necessarily an indicator of its bias. An account can be very homogeneous in terms of the domains that it shares, but none of its domains may be too biased. *Partisanship* here merely reflects the account's tendency to share from similar domains. This is further demonstrated by the low correlation between the accounts' partisanship scores and the cumulative bias of the domains they have shared ($PearsonR = 0.049$ with $p = 5.068 \times 10^{-64}$).

Next, we sampled 20,000 accounts uniformly at random, and used them to map out echo-chambers. To do this, we first identified accounts that had shared identical domains. Figure 1 shows a circular graph of these accounts. Each node represents one account, and its size represents the partisanship of the account. An edge between two accounts indicates common shared domains. The width of each edge is proportional to the number of common domains, and the corresponding bias of those domains. The average clustering coefficient of the network is 0.68. The accounts mentioned in Table 2 are highlighted in the graph.

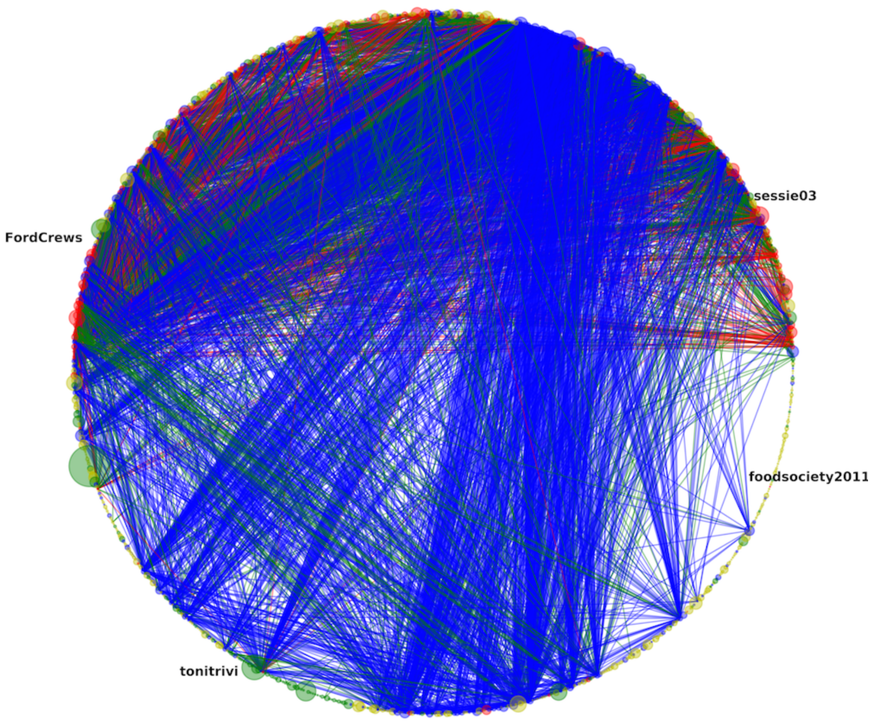

Fig. 1. Circular graph of users from the sample, and the common domains that they have cited.

We used *Python*'s community detection library[3], which in turn uses the Louvain method of identifying communities on a given weighted network [2]. The algorithm identified four communities, which are color-coded in Figure 1. These communities indicate groups of users with large intra-group co-linking behavior, and small inter-group co-linking, which is consistent with the definition of echo-chambers presented in section 1.

Having mined these communities, we would like to know which subjects or topics are prone to be construed differently by a given echo-chamber. In order to do this, we devise a model to identify the semantic deviation of each community compared to the global norm. We use Word2Vec, a method to produce dense, low-dimensional vectors ("embeddings") for words, which maintains composi-

[3] http://perso.crans.org/aynaud/communities/

tionality in some semantic dimensions [13].We use a set of pre-trained vectors created on a corpus of 198 million tweets as the global dataset[4] [11].

Using the dictionary and the vectors provided by the global dataset as initial parameters, we train a model on the tweets posted by each community. In the resulting vector-set, the deviation of a term t from community c is defined as:

$$dev(t, c) = consine\_distance(v_t^c, v_t^g) \tag{4}$$

where v_t^c is the vector representation of t in the model created for community c, and g is the global model. We define the maximum deviation of a term as the community with the largest deviation for that term:

$$max\_dev(t) = \arg\max_c dev(t, c) \tag{5}$$

Next, we find terms that show marked deviation in a certain community, but not in others. This is defined as clusters of 3 neighboring terms that all have the highest distance from the global vector, and show a median absolute deviation of 3.5 or higher [7]. Table 3 shows a few examples of such clusters. As can be seen from the table, the deviating communities express familiar patterns of political discourse around controversial subjects and conspiracy theories.

Table 3. Clusters of deviating terms in the communities from Figure 1

Community	**Deviating Terms**
green	russian, Russia, lobbyist wikileaks, alleged, brothers vaccines, reckless, study
blue	leftists, poll, absurd MAGA, race, rights campaign, budget, Nordstorm
red	Tory, legal, politician Corbyn, #LastMinuteCorbynSmears, @jeremycorbyn bbcsp, industry, MSM

4 Conclusion

In this study, we show that echo-chambers can be mined without manual labeling of data and without making presumptions about the users' political preferences. We use co-linking analysis and a community-detection algorithm to find networks of like-minded users who spread similar information without engaging with other domains or communities. We also offer a method for automatically identifying topics or subjects that those communities are vulnerable to misrepresenting. Due to space limitations, we leave a more detailed description of the models and results to the poster session.

[4] Dataset 7, available at `http://doi.org/10.5281/zenodo.581402`

References

1. Yochai Benkler, Robert Faris, Hal Roberts, and Ethan Zuckerman. Study: Breitbart-led right-wing media ecosystem altered broader media agenda. `https://www.cjr.org/analysis/breitbart-media-trump-harvard-study.php`, Mar 2017.
2. Vincent D Blondel, Jean-Loup Guillaume, Renaud Lambiotte, and Etienne Lefebvre. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment*, 2008(10):P10008, 2008.
3. Jestin Coler. A former fake news creator on covering fake news. `http://niemanreports.org/articles/a-former-fake-news-creator-on-covering-fake-news`, May 2017.
4. Eric Gilbert, Tony Bergstrom, and Karrie Karahalios. Blogs are echo chambers: Blogs are echo chambers. In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference on*, pages 1–10. IEEE, 2009.
5. Joshua Gillin. Fake news website starts as joke, gains 1 million views within 2 weeks. `http://www.politifact.com/punditfact/article/2017/mar/09/fake-news-website-starts-joke-gains-1-million-view`, Mar 2017.
6. Andrew Higgins, Mike McIntire, and Gabriel J.X. Dance. Inside a fake news sausage factory: 'this is all about income'. `https://www.nytimes.com/2016/11/25/world/europe/fake-news-donald-trump-hillary-clinton-georgia.html`, Nov 2016.
7. Boris Iglewicz and David Hoaglin. *Volume 16: how to detect and handle outliers, The ASQC basic references in quality control: statistical techniques, Edward F. Mykytka*. PhD thesis, Ph. D., Editor, 1993.
8. Mohit Iyyer, Peter Enns, Jordan Boyd-Graber, and Philip Resnik. Political ideology detection using recursive neural networks. In *Proceedings of the Association for Computational Linguistics*, pages 1113–1122, 2014.
9. Kristen Johnson and Dan Goldwasser. Identifying stance by analyzing political discourse on twitter. *NLP+ CSS 2016*, page 66, 2016.
10. Alex Krasodomski-Jones. Talking to ourselves?: Political debate online and the echo chamber effect. `https://www.demos.co.uk/wp-content/uploads/2017/02/Echo-Chambers-final-version.pdf`, 2017.
11. Quanzhi Li, Sameena Shah, Xiaomo Liu, and Armineh Nourbakhsh. Data sets: Word embeddings learned from tweets and general data. In *ICWSM*, pages 428–436, 2017.
12. Sapna Maheshwari. How fake news goes viral: A case study. `https://www.nytimes.com/2016/11/20/business/media/how-fake-news-spreads.html?_r=0`, Nov 2016.
13. Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*, pages 3111–3119, 2013.
14. Ben Nimmo. Spread it on reddit: How a fake story about angela merkel led to a far-right cluster on reddit. `https://medium.com/dfrlab/spread-it-on-reddit-3170a463e787`, Feb 2017.
15. Craig Silverman. How the bizarre conspiracy theory behind "pizzagate" was spread. `https://www.buzzfeed.com/craigsilverman/fever-swamp-election?utm_term=.pjMPq6JGW#.tp3682e0q`, Nov 2016.
16. David Weigel. The seth rich conspiracy shows how fake news still works. `https://www.washingtonpost.com/news/the-fix/wp/2017/05/20/the-seth-rich-conspiracy-shows-how-fake-news-still-works/?utm_term=.75fcafc5a6b2`, May 2017.