

Report Paper for "Political Discourse on Social Media: Echo Chambers, Gatekeepers, and the Price of Bipartisanship" by Garimella, Morales, Gionis and Mathioudakis

Seminar: Big Data, AI and Society
University of Zuerich
Philip Hofmann, 14-710-842

*** Summary of the paper ***

The authors are researching the existence of echo chambers in twitter for political topics. For this, they examine tweets that are shown to, and produced, on a per user basis. Those tweets get assigned with a political leaning, on basis of linkage to a specific news site. Meaning that only tweets were considered that link to one of the news sites in the authors database. With that information, they classified users concerning the political diversity of tweets shown and the political diversity of tweets produced. They categorize them into partisan, bipartisan and gatekeepers. At the same time, they also consider network topology of the user, as well as acceptance or popularity of user's content. They conclude that there in fact is the tendency to echo chambers, that there is a price to bipartisanship, and that gatekeepers differ and should have their own category, but their influence on echo chamber is left unstudied.

*** Three ***positive*** aspects of the paper ***

The authors attributed their source data (where it comes from), what datasets they used, how related work ties (or doesn't) into their own work. They came up with a separate category, namely the gatekeepers. They looked at content on a per-user basis as well as network features. The authors also experimented with categorizing with only profile/network attributes.

*** Three ***negative*** aspects of the paper ***

The authors used a filtering-criteria that might be very restrictive and exclude a lot of data and even effect the results. It might have been interesting if the authors also considered the cases where a user 'consumes' mostly content of one political leaning but then produces mostly content of the other political leaning. Considering the 'reply'-function of twitter, it would be interesting to research how partisan members discuss topics with each other, if other political leanings get introduces through replies, that got filtered out if they do not have a link to one of the news sites of the authors in those replies. It would have been interesting if they also researched what effect an echo chamber would have on a user and if it would be beneficial to seed opposing viewpoints into feeds.

*** Evaluation of related work ***

Related work seems to be sufficiently included and contemplated by the authors.

*** Detailed evaluation ***

The authors explained in detail the methodology, formulas, and datasets and the corresponding sources used in the dataset, also the categorization of twitter-users and what filtering mechanisms were used on the dataset. The filtering applied on the dataset seems very restrictive, only keeping tweets that have a link to one of the news-sites that are actually to be found in the corresponding dataset the authors used. The authors applied this filtering to establish a tweets political leaning since they could query the political leaning from the linked news-source out of their database and then they equated the political leaning of the tweet with the political leaning of the news-site. This approach might have reduced the size of the described dataset to be analyzed massively, depending on the ratio of how many tweets are actually including a link to a news-site found in the dataset. I would have liked the authors to compare the datasets after filtering applied, how many datapoints the filtered dataset still contains. If there is a way to categorize the political leaning of a tweet on content alone, without any third-party dependency like linkage to a external site, for example by using word-based classifiers, the classification might be more accurate (users might tweet a link but in the tweet show a opposing viewpoint to the mentioned link) and the filtering would be more inclusive (needing to include a link to one of 500 news-site seems very restrictive). Also the dataset of the 500 news-sites was specifically geared towards facebook (most shared on facebook) and not twitter, we need to acknowledge that the twitter-userbase might have different preferences concerning news-sites shared in tweets.

The writer of this response paper liked that the authors of the paper took different aspects into consideration, not only tweets but also network topology of the users. Also the consideration of 'gatekeepers' and their resulting 'price of bipartisanship' leaves a good impression on the writer. The authors also present their findings and resulting data not only in textual form but also include a lot of figures with a visual representation of their analysis.

*** Reproducibility ***

The authors clearly described their datasets used and the formulas for the analysis. This work should be reproducible under the assumption that the described datasets are available.

*** Editorial remarks ***

None

*** Standard of writing ***

The paper seems to be well-written. The definitions are clear. The authors explained the filtering used (and the analysis made) on the datasets. The sheer amount of figures on the paper and the detail displayed on the figures (all the data points) in such a small scale feels a little bit overwhelming at first, one might consider abstracting details away of the figures, presenting it with an amount of detail that is sufficient for the

reader to understand but not overwhelming concerning the size of the figures on the paper itself.

*** Overall judgment ***

2

*** Statement of confidence ***

Medium