Research Project Proposal
Algorithmic Echo Chambers? Mapping the German Twittersphere

Background, Relevance, and Research Problem

Twitter arguably plays a major role in public communication worldwide as a preferred medium of politicians, journalists and celebrities to directly communicate with or to their intended audiences. As such the structures and dynamics on the platform stand exemplarily as a source of empirical evidence for recent developments in theories about a networked public sphere (Bruns, 2008; Bruns & Highfield, 2016). Not only is it possible to identify characteristic communication structures regarding specific topics, such as polarised crowds or community clusters (Himelboim, Smith, Rainie, Shneiderman, & Espina, 2017), but also to investigate the macro-structures of whole national Twitterspheres. This has been already done for Australia (Bruns, Moon, Münch, & Sadkowsky, 2017) and Norway (Bruns & Enli, 2018).

The analysis of these national Twitterspheres revealed possibilities to investigate the existence and properties of so-called echo chambers or filter bubbles (Bruns, 2017). Filter bubbles and echo chambers are constructs based on the selective sharing of media items that confirm existing beliefs of social media users, or the homophilic subscription, or following, of channels and users that are similar to each other (Bruns, 2017). There is evidence that both phenomena are amplified by search engines and social media (Flaxman, Goel, & Rao, 2016) and affect the health of a public sphere that is meant to enable its participants to engage in informed political deliberation and participation processes (Boutyline & Willer, 2017) by isolating parts of the public political discourse from each other.

However, a similar overview of a national Twittersphere is missing for Germany, which hinders the investigation of these phenomena on Twitter in a German context. This is mainly due to the restrictive access policies of Twitter and its data vendors that make the gathering of a comprehensive dataset, as has been done for the Australian and Norwegian cases, a costly and/or time-consuming task.

Objectives

Instead of attempting the collection of the follow connections of all German-speaking Twitter accounts, this project will generate a large sample of the follow network of the most influential accounts in the German Twittersphere. It will identify topical and social communities of these influential accounts in the German Twittersphere, and, in order to investigate the existence and properties of potential echo chambers, investigate the inward and outward orientation as well as the inner density of these communities. Furthermore, by reconstructing the development of the network, this project will deliver insights into the question of how Twitter's recommendation algorithm for new followers has affected the development of the network, and whether it has intensified the evolution of structural properties that can be identified with echo chamber or filter bubble effects.

Data Sources

This project, as part of a collaboration with the Digital Media Research Centre (DMRC) at Queensland University of Technology (QUT), will make use of a dataset that has been collected by Bruns et al. (2017), comprising account details of all publicly accessible Twitter accounts in 2016. These data allow filtering for accounts by interface language, timezone, and user-provided location and description details. Therefore these data can provide a baseline dataset for assessments regarding the sample size as well as a collection of seeds for the sampling process (see below).

Furthermore, an application for the newly introduced Premium API by Twitter has been successful, which allows (paid) access to comprehensive datasets of historical tweets. This will be useful for the identification of topical communities in the network of sampled accounts.

The follow network will be collected via the public Twitter API.

Research Plan and Methods

DATA COLLECTION

The data collection will be a collaboration of researchers at HBI and DMRC who are interested in working with this data. As the objective is not to collect the whole German-speaking network but only the most influential users, while filtering for accounts using German and/or being identified as located in Germany, the data collection process will follow the rank degree method (Salamanos, Voudigari, & Yannakoudakis, 2017a). This method has been shown to facilitate the identification of the most influential nodes in epidemic network models with samples of only small subgraphs of the full network, while preserving characteristic network measures (Salamanos, Voudigari, & Yannakoudakis, 2017).

DATA ANALYSIS

Community detection algorithms have been successfully used to investigate polarised crowds on Facebook (Del Vicario, Zollo, Caldarelli, Scala, & Quattrociocchi, 2017) and to assess the connectivity of densely linked groups of accounts on Twitter (Bruns et al., 2017). However, community detection algorithms have often ignored underlying definitions (Coscia, Giannotti, & Pedreschi, 2011) that affect the results, leading to epistemological problems especially when assessing the phenomena of filter bubbles and echo chambers (Münch, 2018). Therefore this project will employ three different community detection algorithms: a parallelised version of the Louvain method (Staudt, Sazonovs, & Meyerhenke, 2016; Blondel, Guillaume, Lambiotte, & Lefebvre, 2008), the Infomap method (Rosvall, Axelsson, & Bergstrom, 2009), and the inference of hierarchical stochastic block models (Peixoto, 2018). A comparison of their results provides a further opportunity to assess the effects of algorithms on the perceived structure of public spheres in research itself. An automated keyword extraction method (Münch, 2018) will be used to investigate the topical foci in tweets by accounts in the identified groups. Network measures such as density, E-I-indices, or clustering coefficients (Wasserman & Faust, 1994) will be employed to analyse the connectivity of the groups, while the k-coreness will be used to identify the possibly most influential disseminators in the network (Kitsak et al., 2010). Changes in the structure of the public sphere can be assessed by approximating the date accounts followed each other (Bruns & Woodford, 2014), and comparing results for structures before and after the introduction of follow recommendations by Twitter with help of the Infomap algorithm (Rosvall & Bergstrom, 2010).

TIMEFRAME

The timeframe for this project is estimated to be 9 months. The preferred commencement date is between September and November 2018. The computational resources required are planned to be provided as part of a collaboration with the DMRC after an application for a project on the Australian NeCTAR Research Cloud.

Benefits and Outcomes

The project

- provides a first authoritative sample of influencers in the German Twittersphere which enables a multitude of further research at HBI and DMRC;
- reveals community structures of influencers in the German Twittersphere;
- enables inferences about changes in follow behaviour after recommendation algorithms were introduced on Twitter.

Planned publications:

1. conference contribution regarding the sampling process and first results on test samples;
2. journal article regarding the community structures and their connectivity;
3. journal article regarding the community structure of the German Twittersphere over time and the influence of recommendation algorithms.

References

Blondel, V. D., Guillaume, J.-L., Lambiotte, R., & Lefebvre, E. (2008). Fast unfolding of communities in large networks. http://doi.org/10.1088/1742-5468/2008/10/P10008

Boutyline, A., & Willer, R. (2017). The Social Structure of Political Echo Chambers: Variation in Ideological Homophily in Online Networks. Political Psychology, 38(3), 551–569. http://doi.org/10.1111/pops.12337

Bruns, A. (2008). Life beyond the public sphere: towards a networked model for political deliberation. Information Polity, 13, 71–85.

Bruns, A. (2017). Echo Chamber? What Echo Chamber? Reviewing the Evidence. In Future of Journalism 2017. Cardiff, UK. Retrieved from https://eprints.qut.edu.au/113937/

Bruns, A., & Highfield, T. (2016). Is Habermas on Twitter? Social Media and the Public Sphere. In The Routledge Companion to Social Media and Politics (pp. 56–73). Retrieved from https://eprints.qut.edu.au/91810/

Bruns, A., Moon, B., Münch, F. V., & Sadkowsky, T. (2017). The Australian Twittersphere in 2016: Mapping the Follower/Followee Network. *Social Media + Society*, *3*(4). http://doi.org/10.1177/2056305117748162

Bruns, A., & Enli, G. (2018). The Norwegian Twittersphere: Structure and Dynamics. Nordicom Review, 1–20. http://doi.org/10.2478/nor-2018-0006.1

Bruns, A., & Woodford, D. (2014). Identifying the Events That Connect Social Media Users: Charting Follower Accession on Twitter. SAGE Research Methods Cases. London: SAGE Publications, Ltd. http://doi.org/10.4135/978144627305013516710

Coscia, M., Giannotti, F., & Pedreschi, D. (2011). A classification for community discovery methods in complex networks. Statistical Analysis and Data Mining, 4(5), 512–546. http://doi.org/10.1002/sam.10133

Del Vicario, M., Zollo, F., Caldarelli, G., Scala, A., & Quattrociocchi, W. (2017). Mapping social dynamics on Facebook: The Brexit debate. Social Networks, 50, 6–16. http://doi.org/10.1016/j.socnet.2017.02.002

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. Public Opinion Quarterly, 80 (Special issue 1), 298–320. http://doi.org/10.1093/poq/nfw006

Himelboim, I., Smith, M. A., Rainie, L., Shneiderman, B., & Espina, C. (2017). Classifying Twitter Topic-Networks Using Social Network Analysis, 1–38. http://doi.org/10.1177/2056305117691545

Kitsak, M., Gallos, L. K., Havlin, S., Liljeros, F., Muchnik, L., Stanley, H. E., & Makse, H. A. (2010). Identification of influential spreaders in complex networks. Nature Physics, 6(11), 888–893. http://doi.org/10.1038/nphys1746

McQuail, D. (2010). Mass communication theory (6th ed.). SAGE Publications.

Münch, F. V. (2018). Measuring the Networked Public – Exploring Network Science Methods for Large Scale Online Media Studies (forthcoming doctoral dissertation). Queensland University of Technology, Brisbane, Australia

Salamanos, N., Voudigari, E., & Yannakoudakis, E. J. (2017). A graph exploration method for identifying influential spreaders in complex networks. Applied Network Science, 2(1), 26. http://doi.org/10.1007/s41109-017-0047-y

Salamanos, N., Voudigari, E., & Yannakoudakis, E. J. (2017a). Deterministic graph exploration for efficient graph sampling. Social Network Analysis and Mining, 7(1), 24. http://doi.org/10.1007/s13278-017-0441-6

Smith, M. A., Rainie, L., Shneiderman, B., & Himelboim, I. (2014). Mapping Twitter topic networks: from polarized crowds to community clusters. Pew Research Internet Project. Retrieved from http://www.pewinternet.org/2014/02/20/mapping-twitter-topic-networks-from-polarized-crowds-to-community-clusters/

Staudt, C. L., Sazonovs, A., & Meyerhenke, H. (2016). NetworKit: A tool suite for large-scale complex network analysis. Network Science, 4(4), 508–530. http://doi.org/10.1017/nws.2016.20

Peixoto, T. P. (2018). Bayesian stochastic blockmodeling. Retrieved from http://arxiv.org/abs/1705.10225v4

Rosvall, M., Axelsson, D., & Bergstrom, C. T. (2009). The map equation. European Physical Journal: Special Topics, 178(1), 13–23. http://doi.org/10.1140/epjst/e2010-01179-1

Rosvall, M., & Bergstrom, C. T. (2010). Mapping change in large networks. PLoS ONE, 5(1). http://doi.org/10.1371/journal.pone.0008694

Wasserman, S., & Faust, K. (1994). Social network analysis methods and applications. Cambridge University Press.