

Augmenting Actor Lists with Explorative Network Sampling: Mapping German Right-wing Counterpublics on Twitter

Keywords: *data mining, digital trace data, follow networks, information systems, right-wing counterpublics*

Extended Abstract

Much far-right extremism social media research is based on account and actor lists, built on domain knowledge and years of research experience. While this leads to actor lists with high precision, it is at risk to produce low recall ones. Also, actor lists need to be updated regularly in a time consuming manner which leaves potential for missing out actors and other errors.

We present an approach to evaluate network sampling methods that support augmenting account lists of actors that exhibit stronger than average homophily and is based on a human-in-the-loop design. Our test case to evaluate different approaches are right-wing counterpublics in the German Twittersphere.

Starting with a list of actors, based on a binary coding scheme, we use these accounts as seeds for several topology-based network sampling methods. We adapt these in a self-developed multi-strategy social media network sampling tool, and execute until a predefined sample size is reached. Following a similar approach as Heft et al. (2019), we include actors based on whether they can be categorized as (a) being part of a counterpublic and (b) belonging to the political right. Whereas we understand counterpublics as sub-publics which are in opposition to a perceived or existing political or media mainstream [1]–[3]. Similar to [4], we categorize actors as being part of the political right, if they explicitly declare a right-wing political standpoint.

The basis for the categorization for both criteria is not (only) the account itself, but the actor (individual, group or institution) who is the account owner, and the self-understanding of the group, organization, or ideology the actor might be part of or of which they are supportive. Whenever possible, the classification will rely on existing research and only be conducted by ourselves, in cases when the existing sources are insufficient or not existent.

All algorithms that we will evaluate are all walk/topology-based, for example the following algorithms (if possible evaluation will be conducted over a range of parameters): random walks, forest fire [5], rank degree [6], a breadth-first approach which filters by degree ranks, with rank based on follower numbers and other adaptations for Twitter [7] and spiky-ball sampling [8], a generalisation of some of the above mentioned methods, that allows for more elaborate filters and probabilistic sparsification of the result.

Finally, we evaluate the relative precision and recall of all sampling methods compared to each other based on the results from all methods combined. This leads not only to insights, which the best method for further exploration is in our test case, but also allows us to deduct transferable insights about drawbacks and advantages of every method. As we can draw on recent mappings of the German Twittersphere from 2021/2022 (Fig. 1, Fig. 2, Fig. 3), collected with an adapted version of the rank-degree method [6], [7], [9], we map the

respective result on the overall network structure of the German-speaking Twittersphere. This will help us to understand the performance of single methods based on the cluster- and community-structure of the overall network.

To explore the soundness of our research design, we have assembled a list of 635 actors that tentatively fit our coding scheme for right-wing counterpublic accounts. Of these 635 accounts about a third (217 to be precise) can also be found in our general German Twittersphere sample (Figure 3 & 4). The mapping of these accounts confirms our hypothesis that the qualitative attributes of these actors match with a topologically detectable closeness in the network.

This result has two implications that are conducive to our research design: 1) Walk/topology based methods that sample accounts in proximity to our seeds are very likely to find more actors that fit the coding scheme, and 2) we do not expect to miss many relevant actors when using a method that tries to avoid false positives, e.g. by limiting the diameter of the resulting network sample.

References

- [1] N. Fraser, “Rethinking the public sphere: A contribution to the critique of actually existing democracy,” *Social Text*, no. 25/26, p. 56, 1990, doi: 10.2307/466240.
- [2] A. Rauchfleisch and J. Metag, “The special case of Switzerland: Swiss politicians on Twitter,” *New Media & Society*, vol. 18, no. 10, pp. 2413–2431, Nov. 2016, doi: 10.1177/1461444815586982.
- [3] L. Schwaiger, *Gegen die Öffentlichkeit: alternative Nachrichtenmedien im deutschsprachigen Raum*. Bielefeld: transcript, 2022.
- [4] A. Heft, E. Mayerhöffer, S. Reinhardt, and C. Knüpfer, “Beyond Breitbart: Comparing right-wing digital news infrastructures in six western democracies,” *Policy & Internet*, vol. 4, no. 1, p. 1, 2019, doi: 10.1002/poi3.219.
- [5] J. Leskovec, J. Kleinberg, and C. Faloutsos, “Graphs over time: densification laws, shrinking diameters and possible explanations,” in *Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining*, New York, NY, USA, Aug. 2005, pp. 177–187. doi: 10.1145/1081870.1081893.
- [6] E. Voudigari, N. Salamanos, T. Papageorgiou, and E. J. Yannakoudakis, “Rank degree: An efficient algorithm for graph sampling,” *Proceedings of the 2016 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining, ASONAM 2016*, pp. 120–129, 2016, doi: 10.1109/ASONAM.2016.7752223.
- [7] [REDACTED FOR PEER REVIEW]
- [8] B. Ricaud, N. Aspert, and V. Miz, “Spikyball sampling: Exploring large networks via an inhomogeneous filtered diffusion.” arXiv, Oct. 22, 2020. doi: 10.48550/arXiv.2010.11786.
- [9] [REDACTED FOR PEER REVIEW]

Figures

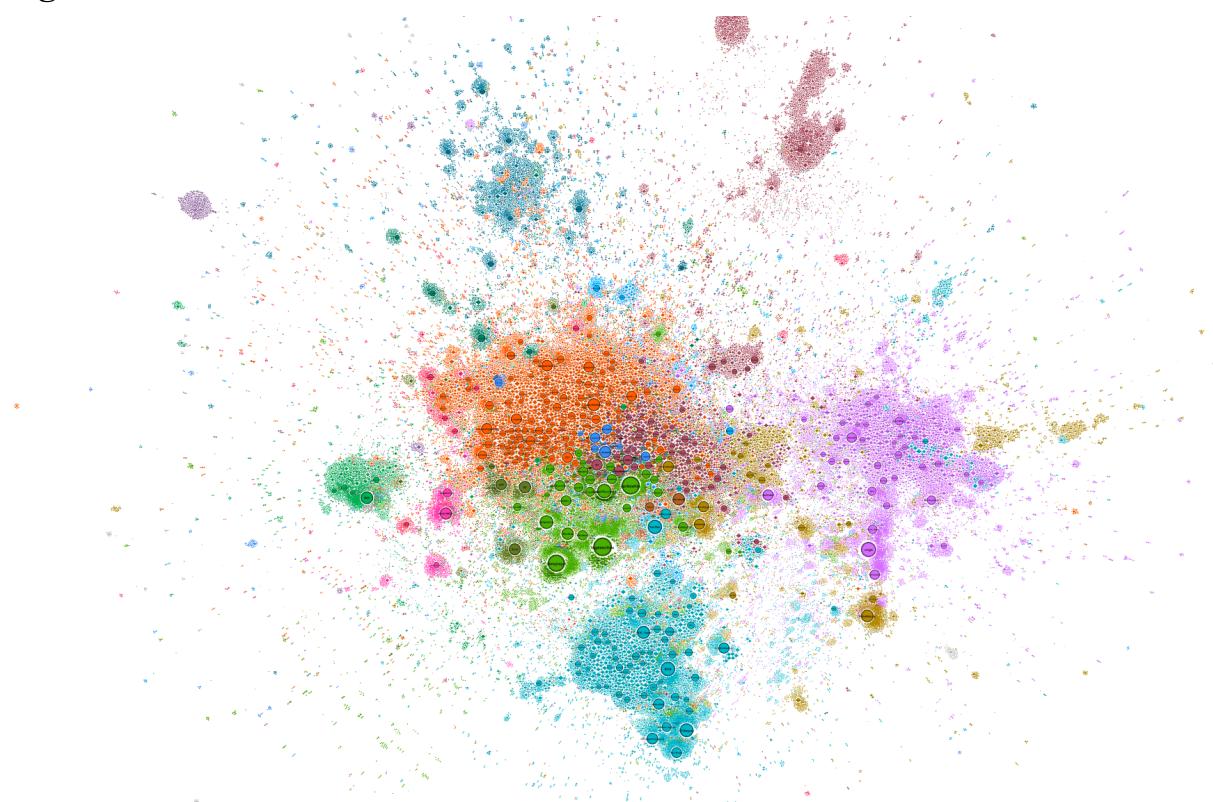


Figure 1: Overview over the German Twittersphere Sample (2-core, collected Dez. 2021 until July 2022)

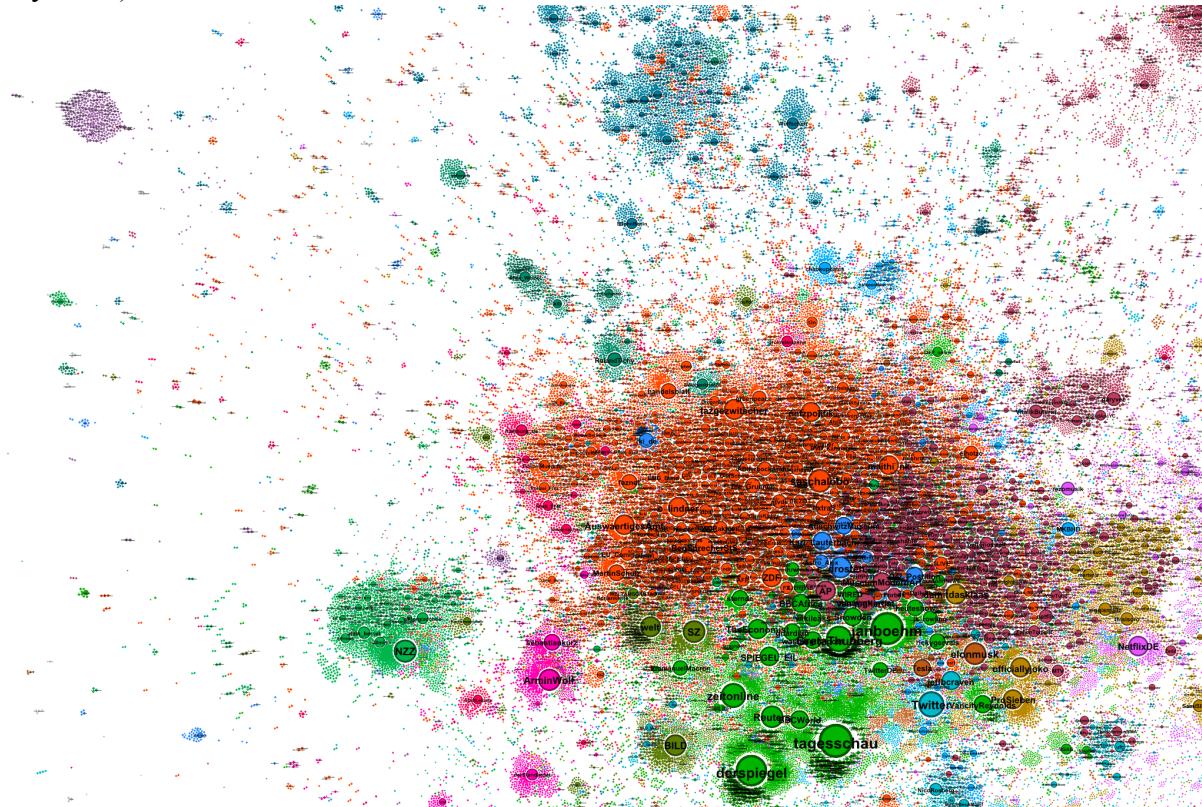


Figure 2: The “mainstream politics” (orange), “mainstream news” (green), and “right-wing” (dark turquoise) clusters of the German Twittersphere Sample

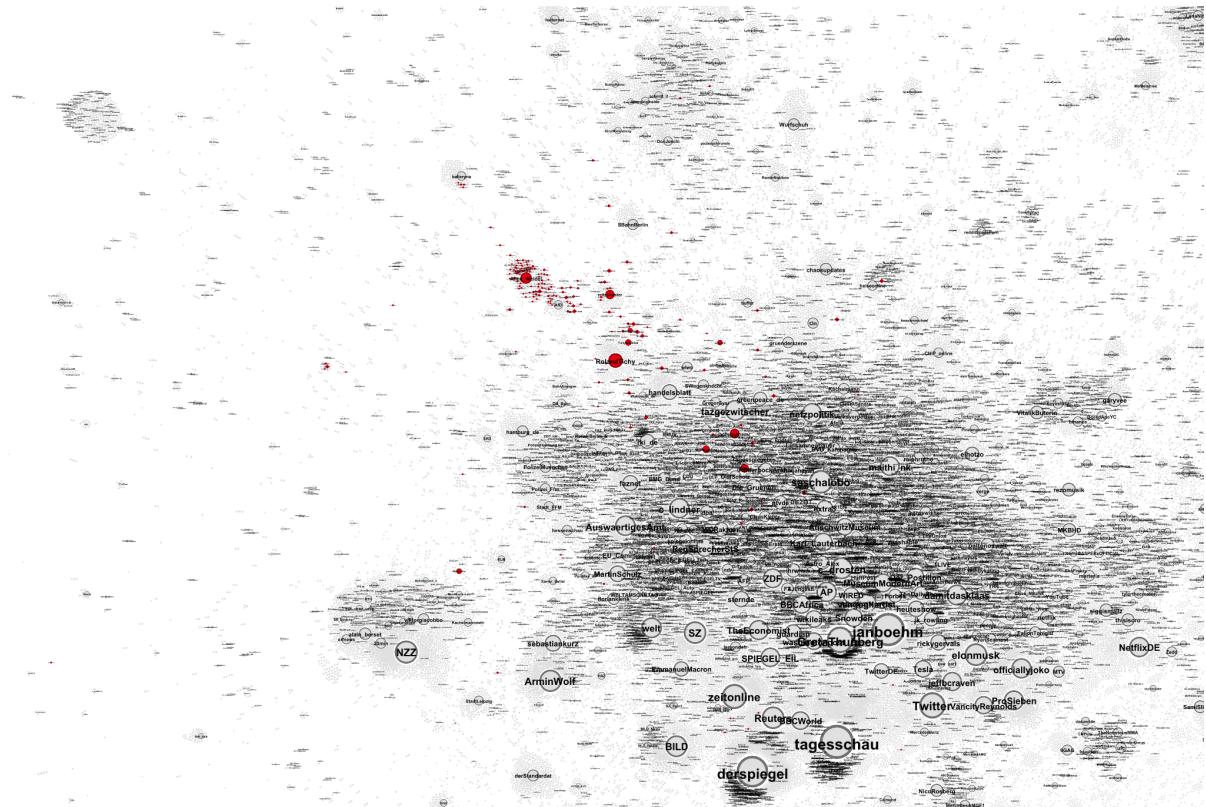


Figure 3: Locations of accounts in the seed list that are also found in the German Twittersphere sample in red (217 accounts)

Figure 4: Accounts in the seed list that are also found in the German Twittersphere sample in the same layout as in Fig. 3, but filtered to seed list accounts only, including sampled connections (217 nodes)