A TALE OF TWO TWITTERS? IDENTIFYING BRIDGES BETWEEN LANGUAGE-BASED TWITTERSPHERES

Felix Victor Münch
Leibniz Institute for Media Research | Hans-Bredow-Institut

Luca Rossi
IT University Copenhagen

## Transnational digital public spheres

Recent global political developments – such as Brexit, the retreat of the USA from the Paris Agreement, or mass migration due to humanitarian crises – are accompanied and influenced by communication that spans across borders of national public spheres. Despite early promises of a global public sphere, questions whether and how online media create a global space that sustains a deliberative discourse of national and global interests on the citizen level, remain understudied due to the complexity and the costs of research with the appropriate geographical representativeness (Hänska & Bauchowitz, 2019).

Addressing this is especially relevant for the European Union, which is, language- and culture-wise, a most diverse global power. Recently, the EU's unity and political integrity has suffered, affecting its social, political, and cultural cohesion. This is also due to polarised public online discourses within and between countries, which are sometimes even actively manipulated (Bastos & Mercea, 2019). Therefore, an overview of public online communication beyond national borders and language barriers is essential.

This paper contributes to this overview by exploring long-term relations between the Italian and German-tweeting Twittersphere, while addressing the following research questions:

1. What is the overall structure of this bilingual network?

2. Are there bridges between these language communities in the form of single accounts and how can they be described?

3. Are there bridges in the form of groups and what are they tweeting about?
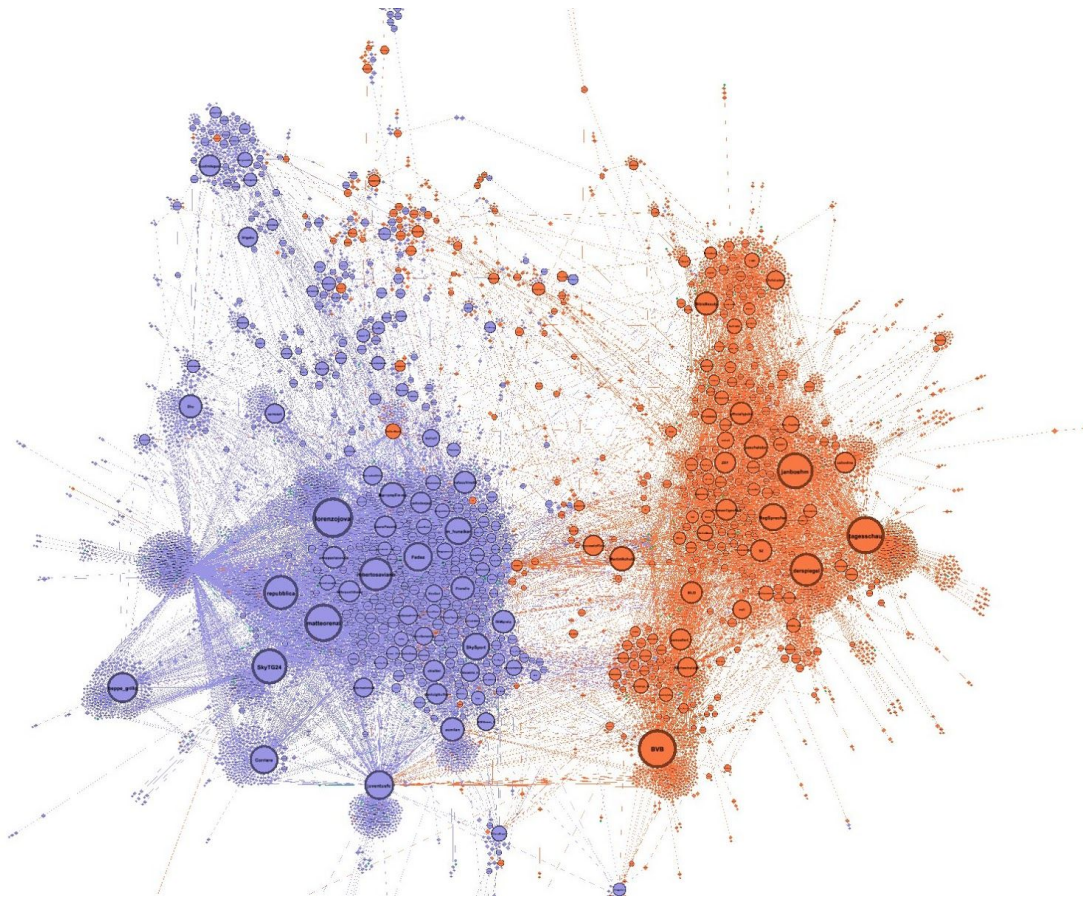
Figure 1: *Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (purple) - German (orange) follow network sample. Nodes sized by betweenness centrality (Brandes, 2001).*

The Italian-German Twittersphere is especially interesting because it reflects much of the EU's complexity: its language-communities are of comparable size, share several national borders, belong to different language families (preventing mutual understability), and the countries involved exhibit distinct cultural differences, with

Germany and Italy often being referred to as exemplary Central and Southern European states, respectively.

## Method and Preliminary Results

We build on an innovative network crawling strategy for language-based Twitter follow networks (Münch et al., 2019). We developed it further to combine strengths and mitigate weaknesses of rank degree (Salamanos et al., 2017), snowball (Goodman, 1961), and forest fire sampling (Leskovec & Faloutsos, 2006). Thereby, we collect a sample of the best connected and centrally positioned accounts in the Italian-German Twittersphere. With 100 random accounts as seeds for each language and 200 crawlers, we achieved a seed pool size of over 1 million accounts and collected a network sample of 14,685 nodes within one day as a feasibility study, which this abstract is based on. Our conference contribution will draw on several months of the ongoing collection.
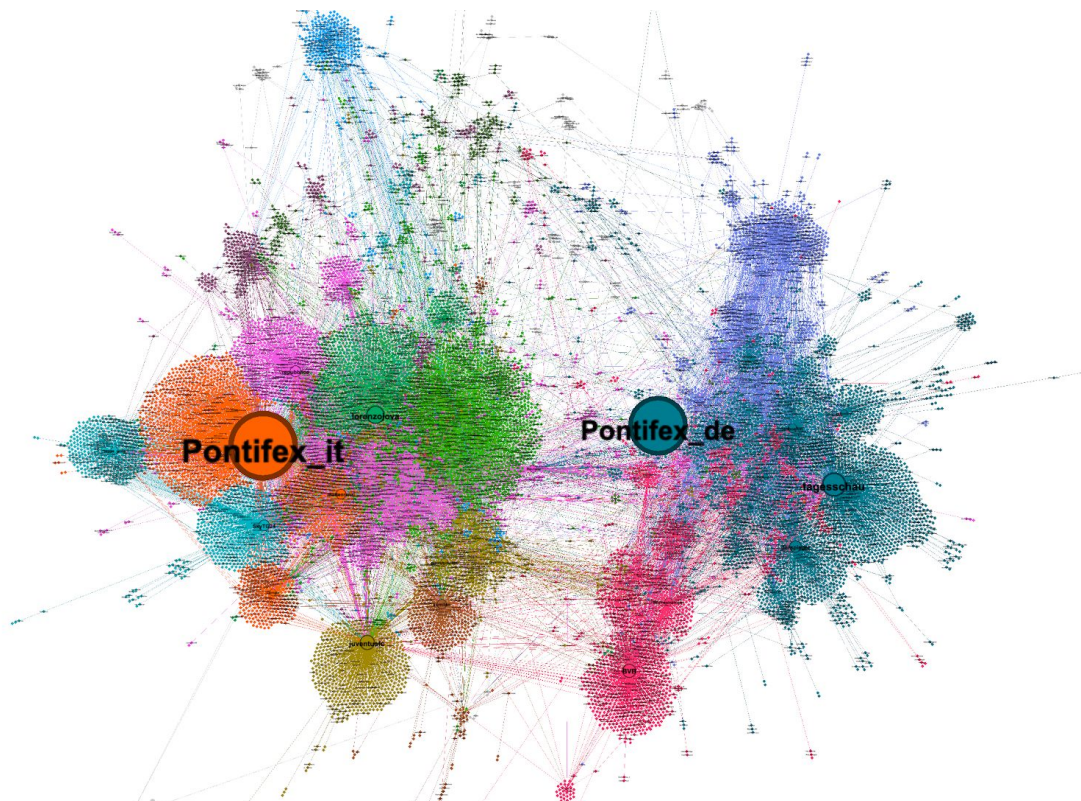
Figure 2: *Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (left) - German (right) follow network sample. Nodes sized by Page Rank (Brin & Page, 1998). Coloured by modularity maximising clusters.*

Each account is assigned to a language based on its tweets. The result, visible in ***Figure 1,*** demonstrates the usefulness of the method as a powerful exploratory tool. The German part of the network clearly reflects earlier results on the German Twitter

Sphere (Münch et al., 2019). In *Figure 2*, the calculation of Page Rank reveals a striking prominence of the Catholic Pope. A visualisation of betweenness centrality suggests the bridging quality of soccer, namely teams playing in the Champions League, e.g., AC Milan, Borussia Dortmund, and Juventus Turin (*Figure 3*, bottom), as well as connections between political clusters of both languages by (former) members of the European Parliament, such as Martin Schulz (*Figure 3*, centre). While betweenness centrality already provides information about the bridging character of single accounts, we also will use bridging centrality (Hwang et al., 2008) to identify bridging accounts more effectively.
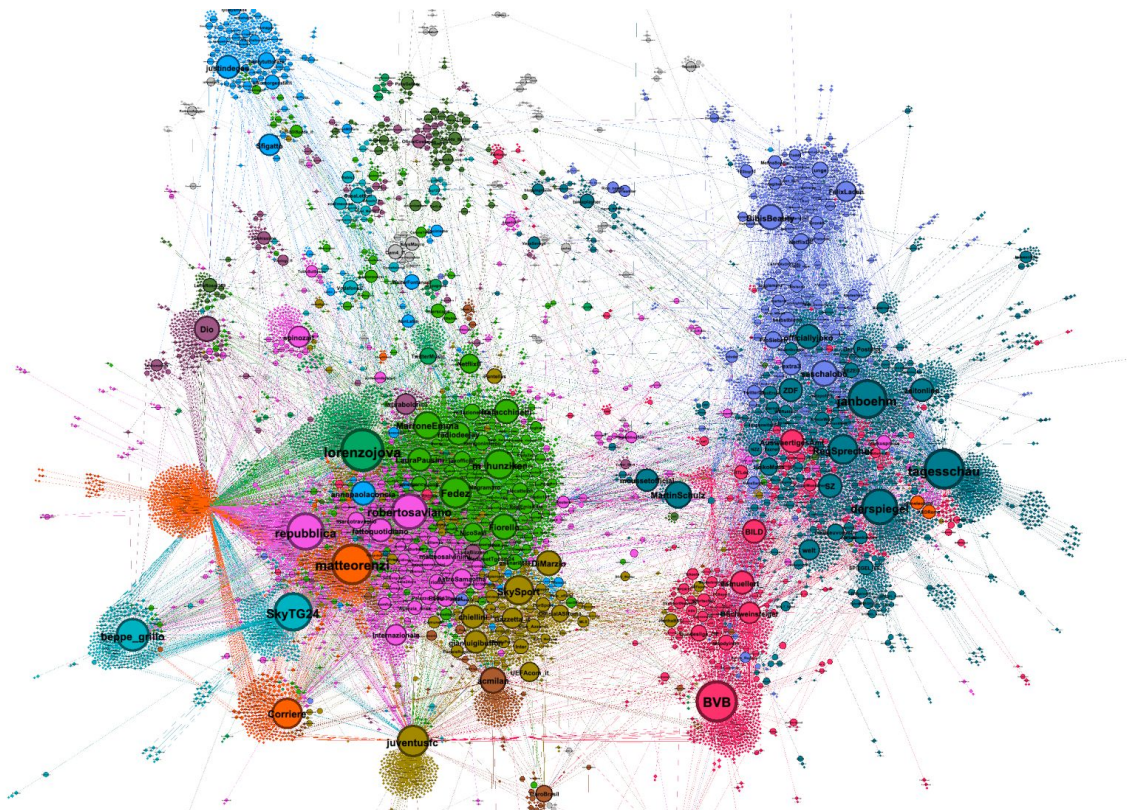

Figure 3: Force directed layout (Force Atlas 2 (Bastian et al., 2009)) of the Italian (left) - German (right) follow network sample. Nodes sized by betweenness centrality (Brandes, 2001). Coloured by modularity maximising clusters.

At a meso level, the data allow to analyse how different clusters in the network are connected across both linguistic domains. Starting from 49 clusters, identified through Louvain modularity optimization (Blondel et al., 2008), we have analysed how many links leaving the clusters connect to German or Italian speaking accounts. *Figure 4* shows the percentage of outgoing connections directed to Italian or German accounts for the non-isolated clusters (N= 41) in relation to the number of members. The largest clusters seem to be connected with one linguistic domain while smaller communities show a different behaviour connecting both with Italian and German speaking accounts.

Outlook

The final paper will present the result of several months of data collection focusing specifically on the relation between topics discussed within the clusters and their connectivity. While the paper will be limited to the study of the German-Italian Twittersphere, the methods can easily be extended to more countries opening up new avenues of enquiry regarding multi-language public spheres as well as the connections and flows of information between national public spheres.
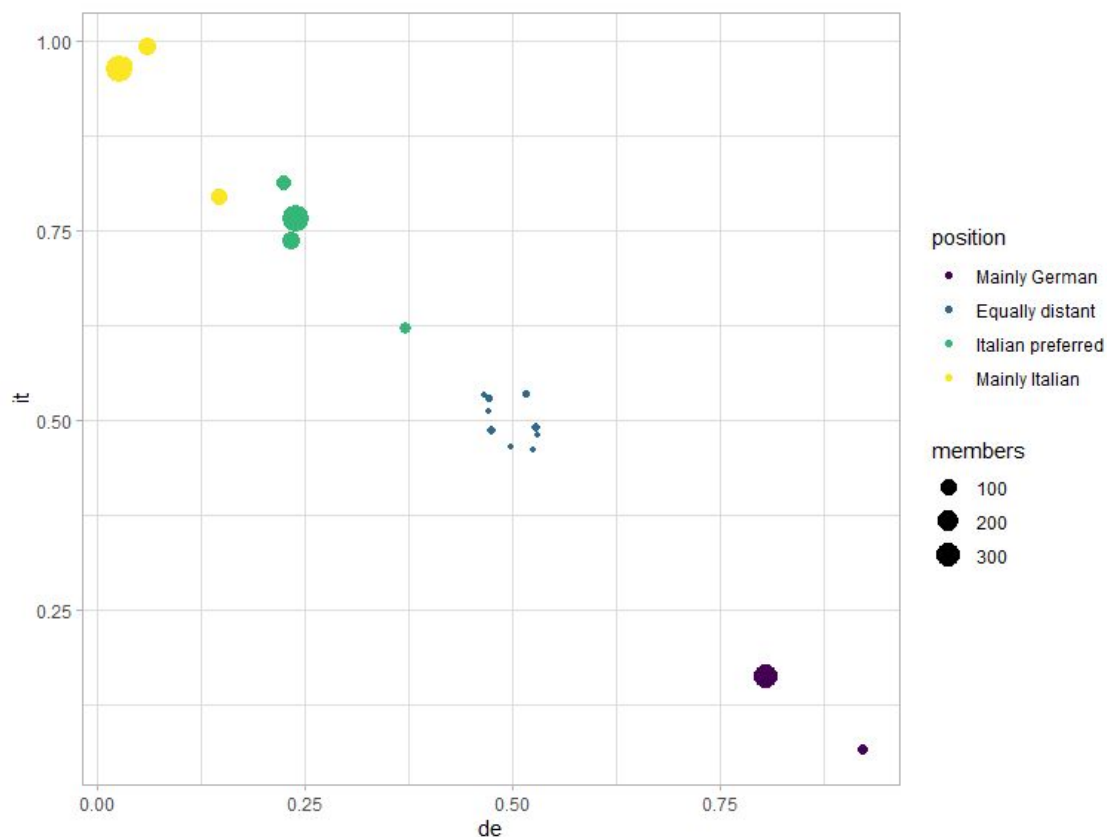
Figure 4: *Fraction of each cluster's outgoing connections towards German or Italian accounts. Size represents the population of the cluster. Jitter has been added to the plot to make overlapping nodes visible.*

References

Bastian, M., Heymann, S., & Jacomy, M. (2009). Gephi: An open source software for exploring and manipulating networks. *ICWSM*. http://www.aaai.org/ocs/index.php/ICWSM/09/paper/viewPDFInterstitial/154Forum/1009

Bastos, M. T., & Mercea, D. (2019). The Brexit Botnet and User-Generated Hyperpartisan News. *Social Science Computer Review*, *37*(1), 38–54. https://doi.org/10.1177/0894439317734157

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

Brandes, U. (2001). A faster algorithm for betweenness centrality. *Journal of Mathematical Sociology*. http://www.tandfonline.com/doi/abs/10.1080/0022250x.2001.9990249

Brin, S., & Page, L. (1998). The anatomy of a large-scale hypertextual Web search engine. *Computer Networks and ISDN Systems*, *30*(1), 107–117. https://doi.org/10.1016/S0169-7552(98)00110-X

Goodman, L. A. (1961). Snowball Sampling. *The Annals of Mathematical Statistics*, *32*(1), 148–170. JSTOR. https://www.jstor.org/stable/2237615

Hänska, M., & Bauchowitz, S. (2019). Can social media facilitate a European public sphere? Transnational communication and the Europeanization of Twitter during the Eurozone crisis. *Social Media + Society*, *5*(3), 2056305119854686. https://doi.org/10.1177/2056305119854686

Hwang, W., Kim, T., Ramanathan, M., & Zhang, A. (2008). Bridging centrality: Graph mining from element level to group level. *Proceeding of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD 08*, 336. https://doi.org/10.1145/1401890.1401934

Leskovec, J., & Faloutsos, C. (2006). Sampling from large graphs. *Proceedings of the 12th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 631–636. https://doi.org/10.1145/1150402.1150479

Münch, F. V., Thies, B., Puschmann, C., & Bruns, A. (2019). Walking Through Twitter: Sampling a Language-Based Follow Network of Influential Twitter Accounts. *ArXiv:1908.07788 [Cs]*. http://arxiv.org/abs/1908.07788

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. https://doi.org/10.1007/s41109-017-0047-y