USER CONFINEMENT ON TWITTER: WHERE STRUCTURAL AND SEMANTIC COMMUNITIES INTERSECT

Jonas STEIN, Jérémie POIROUX and Camille ROTH

Aims & scope

Online communities typically feature clusters, be it social (interactional) or semantic (informational), which often significantly overlap [1,2,5]. Several studies show and use the existence of such overlap, mainly to semantically map the areas of a given social network for a specific case study. Yet, the variety of the configurations of this kind of overlap and the extent to which structural clusters match semantic ones and vice versa [4] remains understudied. Focusing on several Twitter topic-induced networks, we examine the variation of the socio-semantic overlap, both at a macro-level (overlap between clusters in terms of nodes) and at a micro-level (overlap between neighborhoods of individual nodes).

Data

Tweets from 4,631 German Twitter users during March 2020 mentioning "corona" or "covid". We choose nine distinct conversations by assigning each of the most popular hashtags in the sample to one conversation.
For each conversation:
- the structural network is a directed graph of unweighted follower-followee relationships between users
- the semantic network is an undirected graph of weighted tweet links (cosine similarity of word co-occurrences for a user's aggregated tweets)

Method

- *structural* and a *semantic partitions* through unsupervised community detection (Louvain [3])
- *spatialized structural graphs* and their communities with the ForceAtlas2 algorithm [7]
- *intersection of structural and semantic communities* using Sankey flow diagrams
- frequency distributions of user's *semantic neighborhood confinement* (*snc*)
  - users with structural neighbors of exclusively the same (resp. another) semantic community have an *snc* of 100 percent (resp. 0 percent)
- *Adjusted Rand index* (ARI) [6,8] as overall measure of intersection/overlap between structural and semantic partitions

References

[1] Adamic, L. A., and Natalie G. "The political blogosphere and the 2004 US election: divided they blog." *Proceedings of the 3rd international workshop on Link discovery* (2005).
[2] Baltzer, A., et al.. "Interactional and Informational Attention on Twitter." *Information* 10.8 (2019): 250.
[3] Blondel, V. D., et al. "Fast unfolding of communities in large networks." *Journal of statistical mechanics: theory and experiment* 2008.10 (2008): P10008.
[4] Ding, Y. "Community detection: Topological vs. topical." *Journal of Informetrics* 5.4 (2011): 498-514.
[5] Garimella, K., et al. "Quantifying controversy on social media." *ACM Transactions on Social Computing* 1.1 (2018): 1-27.
[6] Hubert, L., and Phipps A. "Comparing partitions." *Journal of classification* 2.1 (1985): 193-218.
[7] Jacomy, M., et al. "ForceAtlas2, a continuous graph layout algorithm for handy network visualization designed for the Gephi software." *PloS one* 9.6 (2014): e98679.
[8] Rand, W. M. "Objective criteria for the evaluation of clustering methods." *Journal of the American Statistical association* 66.336 (1971): 846-850.
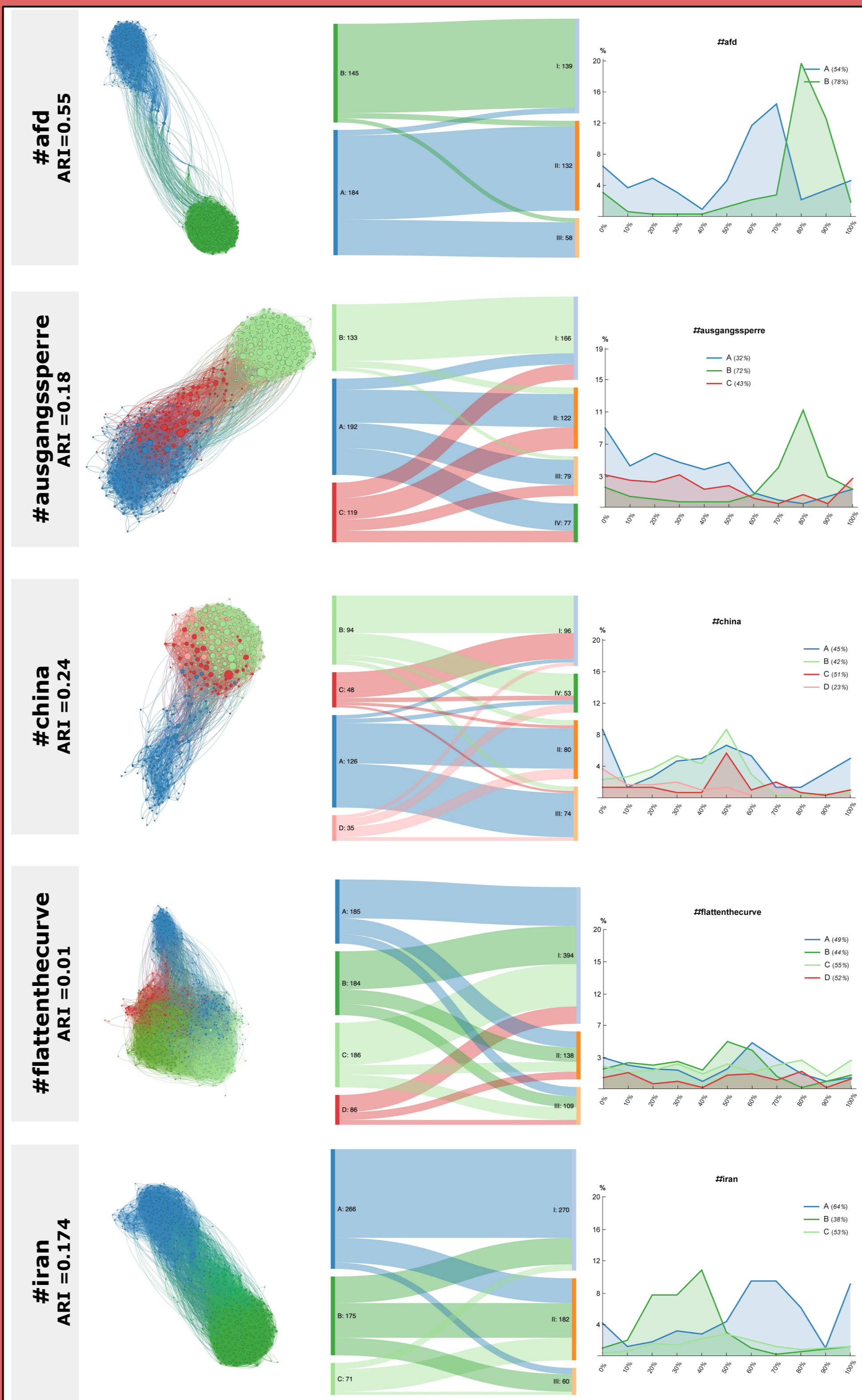
Discussion

- Controversial conversations spark strong structural-semantic intersection: #trump, #afd
  - ➤ Test hypothesis with larger graphs and different conversations for better generalizability
  - ➤ Enhance NLP tools for semantic community detection
  - ➤ Explore qualitative notion of semantic communities for improved understanding
- *snc* as a nouvel approach to measure semantic confinement and polarization in social networks
  - ➤ Outline and analyze *snc* for extended neighborhoods
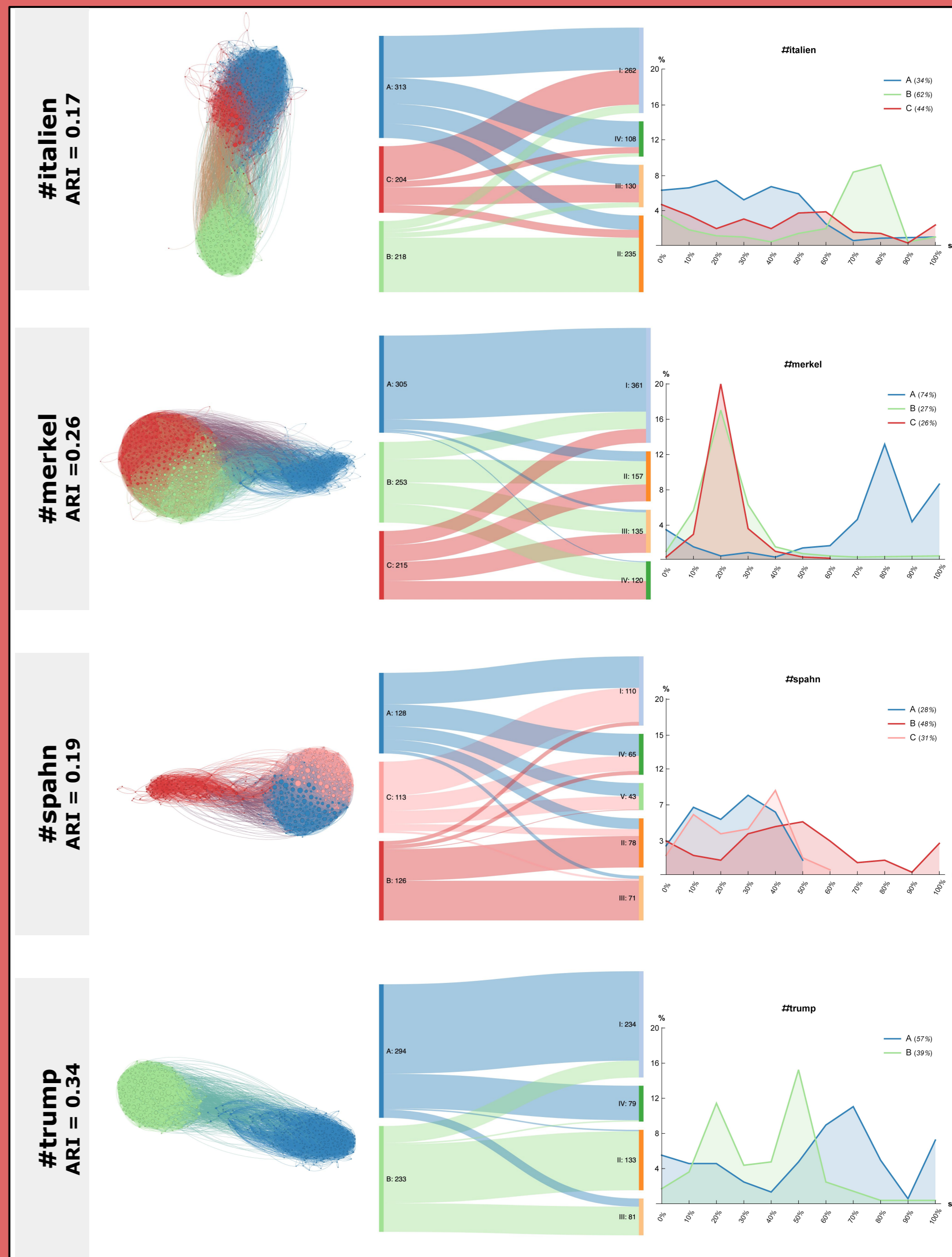
On the left: Spatialization of structural communities using *ForceAtlas2* in *Gephi*.
In the middle: Sankey graph for intersections of structural (*left*, capital letters) and semantic communities (*right*, roman numerals). Numbers indicate users per community.
On the right: Frequency distributions (absolute) for semantic neighborhood confinement (*snc*) of users in structural communities.

Results

- Visual drawing of graphs suggest structurally bi-polarized conversations, independently of the number of structural communities (#china, #italien)
- Conversations with fewer structural communities tend to have stronger intersections of structural and semantic communities (#afd, #trump)
- Users in structural communities with many intersecting semantic communities tend to have lower *snc* (communities B, C in #spahn and #merkel)
- The ARI overlap is greater in conversations where intersections between structural and semantic communities are visually apparent (#afd, #trump)