Language bubbles in online social networks

Alessandro Bellina[1,2,3], Donald Ruggiero Lo Sardo[3,2,1], Emanuele Brugnoli[3,2,1],
Fabio Saracco[2,4,5], Pietro Gravino[6,2], Vittorio Loreto[1,3,2,7], Gabriele Di Bona[8,3,2]

1. Sapienza University of Rome, Department of Physics, Rome, Italy
2. Centro Studi e Ricerche Enrico Fermi (CREF), Rome, Italy
3. Sony Computer Science Laboratories – Rome (Sony CSL-Rome), Joint Initiative CREF-SONY, Rome, Italy
4. CNR, Institute for Applied Computing "Mauro Picone" (IAC), Sesto Fiorentino, Italy
5. IMT School for Advanced Studies Lucca, Lucca, Italy
6. Sony Computer Science Laboratories – Paris (Sony CSL-Paris), Paris, France
7. Complexity Science Hub Vienna, Wien, Austria
8. CNRS, GEMASS, Paris, France

Social media platforms have become crucial arenas for public discourse and political debate. While structural polarisation and echo chambers in online networks are well documented [2, 3], less attention has been devoted to their implications for the language used by different communities and the quality of online communication [4]. We address this gap by investigating the interplay between social fragmentation and linguistic patterns within Twitter/X [1].

We analyse more than 14 million tweets and retweets from Italian politicians and media outlets (2018–2022). By constructing retweet networks across several discursive domains (e.g., immigration, climate change, vaccines, sport), we identify structurally segregated communities [5] and evaluate their linguistic characteristics using multiple metrics of lexical diversity and complexity (e.g., vocabulary size, Shannon entropy, and Jensen–Shannon divergence).

Our analysis reveals two systematic patterns that together define what we call *language bubbles*. First, communities that are closer in the retweet network tend to adopt more similar vocabularies, whereas those that are distant exhibit stronger lexical divergence, consistent with semantic fragmentation [6]. Second, the most isolated and segregated communities consistently show lower lexical richness and complexity, manifesting in reduced vocabulary size, lower entropy, and greater divergence from the overall discourse. These effects are especially pronounced in highly polarised topics [7], while they are less evident in socially cohesive domains such as sports or entertainment.

The emergence of language bubbles highlights a structural link between online fragmentation and linguistic simplification: as communities retreat into segregated communication spaces, they not only diverge in content but also impoverish their linguistic repertoire. This underscores the broader implications of polarisation for the shared linguistic space required for constructive public dialogue.

References

[1] A. Bellina, D. R. Lo Sardo, E. Brugnoli, F. Saracco, P. Gravino, V. Loreto, G. Di Bona, "Language bubbles in online social networks", arXiv preprint arXiv:2507.13068 (2025).

[2] M. Conover et al., "Political polarization on Twitter", Proc. AAAI Conf. Web Soc. Media (2011).

[3] M. Cinelli, G. De Francisci Morales, A. Galeazzi, W. Quattrociocchi, M. Starnini, "The echo chamber effect on social media", Proc. Natl. Acad. Sci. 118, e2023301118 (2021).

[4] S. Flaxman, S. Goel, J. M. Rao, "Filter bubbles, echo chambers, and online news consumption", Public Opin. Q. (2016).

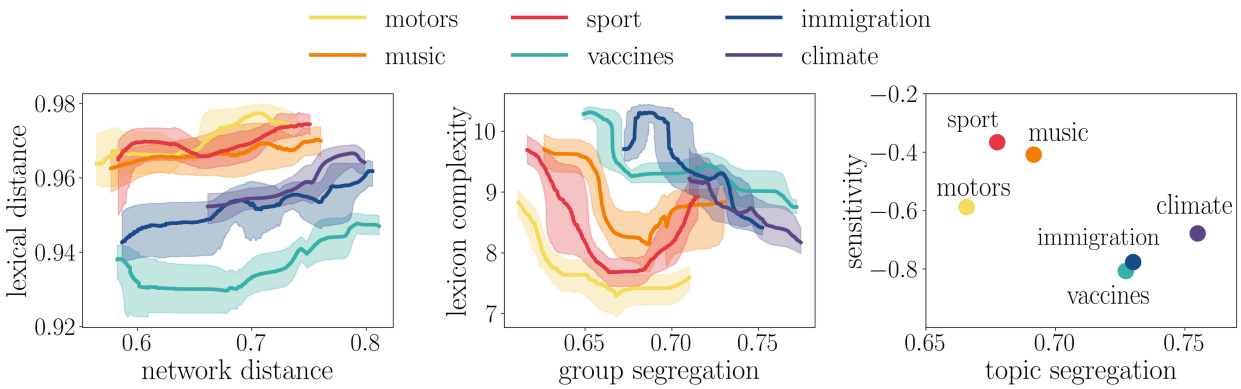

Figure 1: **Language bubbles in online networks. Left:** Lexical distance between communities increases with retweet-network distance. **Center:** Lexical complexity decreases with community segregation. **Right:** The association between segregation and linguistic simplification is stronger in polarised topics (e.g., climate change, immigration, vaccines) than in less fragmented ones (e.g., sports, music). All results are validated against null models.

[5] T. Radicioni, F. Saracco, E. Pavan, T. Squartini, "Analysing Twitter semantic networks: the case of 2018 Italian elections", Sci. Rep. 11, 4381 (2021).

[6] X. Ding, M. A. Horning, E. H. Rho, "Same words, different meanings: Semantic polarization in broadcast media language forecasts polarity in online public discourse", Proc. AAAI Conf. Web Soc. Media 17, 161–172 (2023).

[7] M. Del Vicario, W. Quattrociocchi, A. Scala, F. Zollo, "Polarization and Fake News: Early Warning of Potential Misinformation Targets", ACM Trans. Web 11, 1–22 (2017).