



# Multi-dimensional Echo Chambers: Language and Sentiment Structure of Twitter Discussions on the *Charlie Hebdo* Case

Svetlana S. Bodrunova<sup>(✉)</sup>, Ivan S. Blekanov, and Mikhail Kukarkin

St. Petersburg State University, St. Petersburg 199004, Russia  
s.bodrunova@spbu.ru

**Abstract.** *Background.* Public discussions on social networks have trans-border and multilingual nature. This is especially true for conflictual discussions that reach global trending topics. Being part of the global public sphere, such discussions were expected by many observers to become horizontal, all-involving, and democratically efficient. But, with time, criticism towards the democratic quality of discussions in social media arose, with many works discovering the patterns of echo chambering in social networks. Even if so, there is still scarce knowledge on how affective hashtags work in terms of user clusterization, as well as on the differences between emotionally ‘positive’ and ‘negative’ hashtags. *Objectives.* We address this gap by analyzing the Twitter discussion on the *Charlie Hebdo* massacre of 2015. In this discussion, the Twittershpere has created #jesuischarlie and #jenesuispascharlie - two discussion clusters with, allegedly, opposite sentiments towards the journal’s ethics and freedom of speech. *Research design.* We were interested in whether echo chambers formed both on the hashtag level (based on language use) and within a language (based on user sentiment of French-speaking users). For data collection, we used vocabulary-based Twitter crawling. For data analysis, we employed network analytics, manual coding, web graph reconstruction, and automated sentiment analysis. *Results.* Our results show that #jesuischarlie and #jenesuispascharlie are alike in language distribution, with French and English being the dominant languages and the discussions remaining within the Euro-Atlantic zone. The language-based echo chambers formed in both cases. But if #jesuischarlie was a clear sentiment crossroads, #jenesuispascharlie was a negative echo chamber, thus allowing us to draw conclusions about multi-layer echo chambering.

**Keywords:** Twitter · *Charlie Hebdo* · Sentiment analysis · Echo chambers

## 1 Introduction

Public discussions on social networks potentially have trans-border and multilingual nature. This comes true in heated conflictual discussions that reach global trending topics. Such discussions are expected to demonstrate ‘civilizational clashes’ [1].

Being part of the global public sphere, since the 1990s, such discussions were expected by many observers to be more horizontal, all-involving, and democratically

efficient [2] than the traditional mass-mediated discussions [3]. But, with time, criticism towards the democratic quality of discussions in social media arose, with many works discovering the patterns of echo chambering and discourse polarization in social networks [4–9], which lowered the capacities of inter-group discussions and, thus, just formed an additional line of social segregation.

Object-oriented hashtagged discussions have been thoroughly studied in the 2010s, including those on political and social conflicts. But there is still scarce knowledge on whether affective hashtags [10] that convey emotions – either of solidarity with or of anger towards a particular social group – work in terms of user clusterization. Also, there is no clear understanding of comparative democratic quality of emotionally ‘positive’ and ‘negative’ hashtags in terms of echo chambering.

In this paper, we address these gaps by analyzing the Twitter discussion on the *Charlie Hebdo* massacre of 2015. In the discussion upon the mass killings, the Twittershpere has created #jesuischarlie and #jenesuispascharlie - two emotionally differing discussion clusters with, allegedly, opposite sentiments towards the journal’s ethics and freedom of speech; the hashtags soon became ‘role models’ for online solidarity towards the victims of terrorist attacks and anthropogenic disasters.

To analyze the echo chambering patterns in the two discussions, we have focused upon two levels of echo chambers. We were wondering whether echo chambers formed on the level of a hashtag (based on language use) and within a particular language (based on user sentiment of French-speaking users).

The remainder of the paper is organized as follows. Section 2 reviews the literature on echo chambering in social media. Section 3 presents our methodology and the conduct of the research. Section 4 presents our results and discusses them.

## 2 User Groupings on Twitter and the Efficacy of Public Sphere

### 2.1 Social Media and the Public Sphere: Echo Chambers vs. Opinion Crossroads

Public sphere as a spatial metaphor for a complex of discussions and procedures with a public status and decision-making goals [11] has been amplified by the appearance of social media in the 2000s. By 1990s, it had been established in the academic literature that mediatized public sphere with traditional media playing the role of information hubs was uneven and hardly efficient in terms of access to opinion expression, as whole social groups remained under-represented, and newsmakers privileged in comparison to *vox populi*. With the appearance of social media, hopes arose that the new communicative milieus would foster horizontalization of communication and provide for democratization and higher political participation [2]. Also, hopes for better understanding and resolution of non-political inter-group conflicts existed.

But with time, these hopes faded, as offline disparities seemed to reproduce online, including political interests, race, gender, and other inequalities [12]; emotion and affect proved to rule the discourse [10], with publics even in most democratically developed countries moving from diverse in opinion to dissonant and disconnected

[13]. With the development of social network analysis (SNA) and its application to social media research, the question of the efficacy of the public discussions in social media [14] became linked to network and structural features of the discussions, such as the influencer status [15, 16] and clusterization of users also known as user polarization [6, 7] and echo chambering [8, 9]. Some evidence was also gathered on discussion spherules forming on the global scale just as well as nationally [17].

## 2.2 Why the Twitter Discussions Fragment: Linguistic Properties of Speech as Catalyzers of Echo Chambering

In early studies of social networks and its users, authors interested in testing the ability of networks to pull together users from distant locations and weak ties linked geographical distance with factors like residence of users and their language profile [18]. This is why Twitter that enabled the (arguably) quickest possible information spread across locations and languages became a major attractor of scholarly attention [19, 20]. But despite the global reach of the platform, several studies have found that people were still connected locally on Twitter [17, 21].

Along with locality and residence, linguistic factors, arguably, play a major role in user grouping on the global scale. Thus, the language(s) used by the discussion participants is the first natural barrier that is expected to make users group together and communicate within their language-based echo chambers [22], both on Twitter on the whole and within particular hashtags [23].

Other factors have also been discussed as the catalyzers of user grouping on Twitter. Among those, political attitudes lead the research agenda [6, 7]. Here, several ways to detect user clusterization exist. Of them, use of network or semantic proxies like friendship ties [24], patterns of following [25] and retweeting [26], content sharing [8, 27] etc. is till today the most prominent; another is automated analysis of user sentiment, either in general or toward an issue/actor in question (object-oriented) [28].

But all these studies depict user groupings within a single dimension; our idea is to try and trace user groupings multi-dimensionally – both on the level of a hashtag (based on language use) and within a language nebula (based on user sentiment).

## 2.3 The *Charlie Hebdo* Case: Emotional Hashtags and User Groupings

To search for multi-dimensional echo chambers, we have chosen the case of the *Charlie Hebdo* massacre of 2015. Here, we could hypothesize that the existence of emotionally opposite hashtags (#jesuischarlie and #jenesuispascharlie) already creates enclaves within the general discussion on the case. Then, within the hashtags, several clusters based on language structure may exist. Then, on the third level, we will look whether within the language clusters sub-clusters of sentiment form. In general, our idea is to see how exactly the language clusters correspond to the sentiment clusters in each language and whether ‘positive’ users within one language are linked to such in another language, while ‘negative’ users also group across languages in a similar way. But here we present only preliminary results that check if the user clusters may at all be detected based on language and on sentiment within a language.

## 2.4 Research Hypotheses

Thus, our hypotheses are the following:

**H1a.** Non-random user groups will be detected for both hashtags, as based on language use.

**H1b.** #jesuischarlie will not differ from #jenesuispascharlie in their language structure, as both hashtags have reached global trending topics and are expected to show ‘civilizational clashes’.

**H2a.** Non-random user groups will be detected within one language (French), as based on positive and negative sentiment.

**H2b.** #jesuischarlie will differ from #jenesuispascharlie in the grouping based on user sentiment, due to the emotional opposition of the hashtags themselves.

**H3.** Multi-layer echo chambering may be detected in trans-border hashtagged discussions of global reach.

## 3 Data Collection and Conduct of Research

### 3.1 Data Collection and the Datasets

Using a web crawler developed especially for the Twitter data collection, we gathered all the tweets published openly under the hashtags #jesuischarlie and #jenesuispascharlie (by separate crawls) within January 7 to 9, 2015, as these three days covered the active conflict (from the killings in the editorial office to the assailants’ death) when the users provided virtually millions of tweets for collection.

The collected datasets included: for #jesuischarlie: 420,080 tweets; 266,904 tweeters; 719,503 users who interacted with the posted tweets (by likes, retweets, or comments); for #jenesuispascharlie: 7,698 tweets; 5,466 tweeters; 17,872 users who posted and interacted with the posted tweets (by likes, retweets, or comments).

These full datasets were later used to reconstruct the overall web graphs for the two hashtags. But the datasets were very different in size, and be able to color them with language markers, we needed to sample the users for coding having in mind the volume difference of the datasets.

### 3.2 Language Analysis: Sampling, Coding, and Graph Reconstruction

To answer H1a and H1b, we have coded the users for their language use and then applied these data to the datasets for web graph reconstruction.

After eliminating bots and bot-like users (those who posted over 60% of doubled tweets) as well as hashtag-only tweets, we have followed the strategy developed by the research group for previous Twitter studies [29–31], namely uniting random sampling with detection of influential users (influencers) for taking them into account. Then, we have coded all the influencers (disregarding the number of tweets they posted; for #jesuischarlie, 402 users, for #jenesuispascharlie, 85 users) and ‘ordinary users’ sampled in the feasible and comparable way. For #jenesuispascharlie that was substantially smaller, all the users with 3 and more tweets were coded (339 users); for

#jesuischarlie, the ‘ordinary users’ with 5 tweets or more were taken into account (9,090 users), and of them, each second was coded (4500 users).

All the sampled underwent expert reading and were coded manually marking the number of tweets in language 1, language 2, and other languages; thus, users posting on one, two, and three or more languages were defined. The languages were identified for each user; in case of rare languages, Yandex language identifier was employed.

To reconstruct the graphs, we use Gephi API algorithms openly available online. Of the available algorithms, two were chosen: Hu [32] and OpenOrd [33]; here, YifanHu-based graphs are presented, as the OpenOrd graphs require more space for presentation. We colored both the nodes (users) and the edges (connections between users). To prove that the visual nebulae are not artifacts of subjective viewership, we calculated the percentage of edges between and inside language groups, eliminating the ‘loops’ of self-commenting/liking by the users.

### 3.3 Sentiment Analysis: Sampling, Vocabulary Building, and Graph Reconstruction

After we have proved that the French-speaking users show non-random grouping in both cases (see below), we have taken them for sentiment analysis. The number of users for #jesuischarlie included 1291 user; for #jenesuispascharlie, 117 users.

Our strategy for French-language sentiment analysis was the following. We have united three sources in our vocabulary: the existing French dictionary with sentiment marking, machine translation from an additional Wordnet vocabulary, and the case-based vocabulary created from the collected tweets and manually marked for positive, negative, and neutral sentiment regardless of the case-specific meanings.

This vocabulary was applied to each tweet of the abovementioned French-speaking users; for each tweet, the sentiment was calculated. Then, the thresholds were defined: positive and negative were the users with positive(+neutral) and negative(+neutral) tweets, respectively; neutral were those with neutral tweets only; mixed were those with positive + negative(+neutral) tweets.

Then, we also checked the groupings with by calculating the percentage of edges both between and inside language groups, eliminating the ‘loops’ of self-commenting/liking by the users.

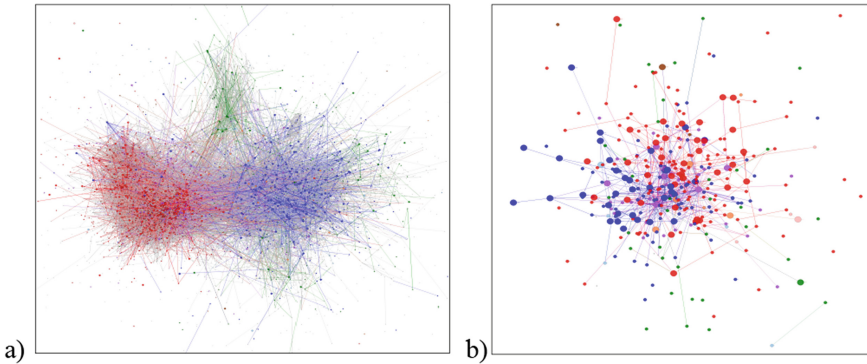
## 4 Results and Discussion

Our results are described below with regard to the hypotheses stated above.

*H1a/H1b.* To assess the user groupings in both hashtagged discussions, we have reconstructed the web graphs for the coded users (see Fig. 1a for #jesuischarlie and Fig. 1b for #jenesuispascharlie, respectively). What we see on the graphs are three nebulae for (a): French, English, and other European, and two for (b): French and English. But the results of calculations of percentage of edges between and inside groups tell that the actual grouping is slightly different from what we see with unaided eyes. For #jesuischarlie, the nebulae with density higher than the inter-group ones are French, English, and French/English (52.1%, 16.7%, and 16.9%, respectively, against

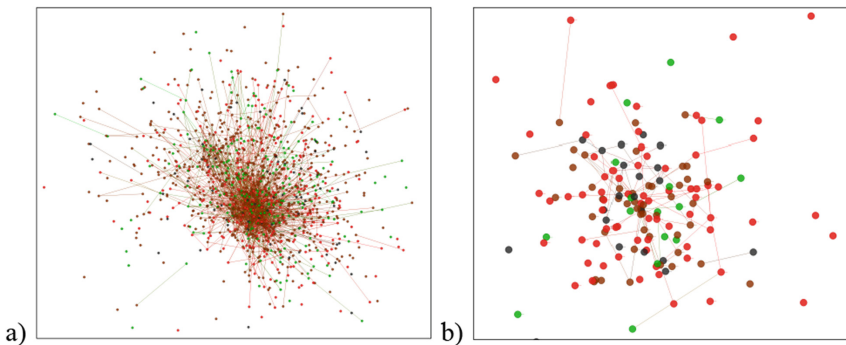
6.17% for inter-group edges) and not other European. For #jenesuispascharlie, the graph is much denser (26%), but still the same three clusters show up, with 26.2%, 22.4%, and 18.8%, respectively; in both cases, other language clusters are virtually non-existent and do not mount to 3%. As we stated in our earlier investigations [31], we have not seen a sign of ‘civilizational clashes’ in any of the hashtags.

Thus, H1a is proven; H1b is proven too but not due to ‘civilizational clashes’.



**Fig. 1.** The YifanHu graphs (fragments) for language distribution in (a) #jesuischarlie and (b) #jenesuispascharlie. Red: French; blue: English; lilac: French/English; green: other European. (Color figure online)

*H2a/H2b.* To see the user grouping and sentiment cleavages within the French-speaking parts of the discussions, we have reconstructed the web graphs for them (see Fig. 2a for #jesuischarlie and Fig. 2b for #jenesuispascharlie, respectively).



**Fig. 2.** The YifanHu graphs (fragments) for language distribution in (a) #jesuischarlie and (b) #jenesuispascharlie. Red: French; blue: English; lilac: French/English; green: other European. (Color figure online)

Here, H2a should be rejected for #jesuischarlie and partly supported for the second hashtag. Both in the graph and in the edge percentage calculations, it is only users with mixed sentiment who form a group (38.33% against 56.05% for the inter-group connections) in #jesuischarlie. But for #jenesuispascharlie, both mixed and negative user groups seem to have a potential for grouping (19.5% and 15% against 62.1% for inter-group connections). Thus, H2b is supported: the cases do differ.

Even with the cases of such different sizes, H3 is supported: thanks to the negative nebula in #jenesuispascharlie, we can state that echo chambers are able to form on at least two levels of the trans-border conflictual discussions of global (or, more precisely, macro-regional) reach. This adds to our understanding of the nature of public discussions in social media, even if lowers hopes for all-encompassing public spheres.

**Acknowledgements.** This research has been supported in full by Presidential Grants of the Russian Federation for young scholars with Doctoral degrees (grant МД-6259.2018.6).

## References

1. An, J., Kwak, H., Mejova, Y., De Oger, S.A.S., Fortes, B.G.: Are you Charlie or Ahmed? Cultural pluralism in Charlie Hebdo response on Twitter. In: Proceedings of ICWSM, pp. 2–11 (2016)
2. Fuchs, C.: *Social Media: A Critical Introduction*. Sage, London (2017)
3. McQuail, D.: *Media Performance*. Sage, Thousand Oaks (1992)
4. Sunstein, C.: *Republic.com*. Princeton University Press, Princeton (2001)
5. Sunstein, C.: The law of group polarization. *J. Polit. Philos.* **10**(2), 175–195 (2002)
6. Barberá, P., Jost, J.T., Nagler, J., Tucker, J.A., Bonneau, R.: Tweeting from left to right: is online political communication more than an echo chamber? *Psychol. Sci.* **26**(10), 1531–1542 (2015)
7. Bastos, M.T., Mercea, D., Baronchelli, A.: The spatial dimension of online echo chambers. arXiv preprint [arXiv:1709.05233](https://arxiv.org/abs/1709.05233) (2017)
8. Colleoni, E., Rozza, A., Arvidsson, A.: Echo chamber or public sphere? Predicting political orientation and measuring political homophily in Twitter using big data. *J. Commun.* **64**(2), 317–332 (2014)
9. Conover, M., Ratkiewicz, J., Francisco, M.R., Gonçalves, B., Menczer, F., Flammini, A.: Political polarization on Twitter. In: Proceedings of ICWSM, vol. 133, pp. 89–96 (2011)
10. Papacharissi, Z.: *Affective Publics: Sentiment, Technology, and Politics*. Oxford University Press, New York (2015)
11. Kleinstüber, H.J.: Habermas and the public sphere: from a german to a European perspective. *Javnost Publ.* **8**(1), 95–108 (2001)
12. Daniels, J.: Race and racism in Internet studies: a review and critique. *New Med. Soc.* **15**(5), 695–719 (2013)
13. Pfetsch, B.: Dissonant and disconnected public spheres as challenge for political communication research. *Javnost Publ.* **25**, 1–8 (2018)
14. Bruns, A., Highfield, T.: Social media and the public sphere. In: *The Routledge Companion to Social Media and Politics*, vol. 56 (2015)
15. Bodrunova, S.S., Blekanov, I.S., Maksimov, A.: Measuring influencers in Twitter ad-hoc discussions: active users vs. internal networks in the discourse on Biryuliovo bashings in 2013. In: Proceedings of AINL, pp. 1–10. IEEE (2016)

16. Bodrunova, S.S., Litvinenko, A.A., Blekanov, I.S.: Influencers on the Russian Twitter: institutions vs. people in the discussion on migrants. In: Proceedings of EGOSE, pp. 212–222. ACM (2016)
17. Cammaerts, B., Audenhove, L.V.: Online political debate, unbounded citizenship, and the problematic nature of a transnational public sphere. *Polit. Commun.* **22**(2), 179–196 (2005)
18. Takhteyev, Y., Gruzd, A., Wellman, B.: Geography of Twitter networks. *Soc. Netw.* **34**(1), 73–81 (2012)
19. Lotan, G., et al.: The revolutions were tweeted: information flows during the 2011 Tunisian and Egyptian revolutions. *Int. J. Commun.* **5**, 1375–1405 (2011)
20. Hong, L., Convertino, G., Chi, E.H.: Language matters in Twitter: a large-scale study. In: Proceedings of ICWSM, pp. 518–521 (2011)
21. Yardi, S., Boyd, D.: Dynamic debates: an analysis of group polarization over time on Twitter. *Bullet. Sci. Technol. Soc.* **30**(5), 316–327 (2010)
22. Chen, W., Tu, F., Zheng, P.: A transnational networked public sphere of air pollution: analysis of a Twitter network of PM<sub>2.5</sub> from the risk society perspective. *Inf. Commun. Soc.* **20**(7), 1005–1023 (2017)
23. Bastos, M.T., Puschmann, C., Travitzki, R.: Tweeting across hashtags: overlapping users and the importance of language, topics, and politics. In: Proceedings of the 24th ACM Conference on Hypertext and Social Media, pp. 164–168. ACM (2013)
24. Barberá, P., Rivero, G.: Understanding the political representativeness of Twitter users. *Soc. Sci. Comput. Rev.* **33**(6), 712–729 (2015)
25. Rivero, G.: Preaching to the choir: ideology and following behaviour in social media. *Contemporary Soc. Sci.* 1–17 (2017)
26. Calais Guerra, P.H., Meira-jr, W., Cardie, C., Kleinberg, R.: A measure of polarization on social media networks based on community boundaries. In: ICWSM Proceedings (2013)
27. Bakshy, E., Messing, S., Adamic, L.A.: Exposure to ideologically diverse news and opinion on Facebook. *Science* **348**(6239), 1130–1132 (2015)
28. Conover, M.D., Gonçalves, B., Ratkiewicz, J., Flammini, A., Menczer, F.: Predicting the political alignment of twitter users. In: Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), pp. 192–199. IEEE (2011)
29. Authors, 2016a
30. Authors, 2016b
31. Authors, 2018
32. Hu, Y.F.: Efficient and high-quality force-directed graph drawing. *Math. J.* **10**, 37–71 (2005)
33. Martin, S., Brown, W.M., Klavans, R., Boyack, K.: OpenOrd: an open-source toolbox for large graph layout. In: Proceedings of the SPIE Conference on Visualization and Data Analysis (VDA), vol. 7868. [https://www.researchgate.net/profile/Kevin\\_Boyack/publication/253087985\\_OpenOrd\\_An\\_Open-Source\\_Toolbox\\_for\\_Large\\_Graph\\_Layout/links/0deec5205279e8c66a000000.pdf](https://www.researchgate.net/profile/Kevin_Boyack/publication/253087985_OpenOrd_An_Open-Source_Toolbox_for_Large_Graph_Layout/links/0deec5205279e8c66a000000.pdf)