Measuring agenda-setting effects on Twitter during the 2016 UK EU referendum

Tiago Santos
ISTAR–Instituto Universitário
de Lisboa (ISCTE-IUL)
Lisbon, Portugal
tmsss@iscte-iul.pt

Jorge Louçã
ISTAR–Instituto Universitário
de Lisboa (ISCTE-IUL)
Lisbon, Portugal
jorge.l@iscte-iul.pt

Helder Coelho
Universidade de Lisboa
Lisbon, Portugal
hcoelho@di.fc.ul.pt

Abstract— **This paper investigates first-level agenda-setting effects on Twitter during UK's EU referendum campaign. Using topic-modeling techniques, we investigate the dynamics of the interaction between users and the media content, and how the media outlets relate to each other. Results show that traditional media outlets dominated the debate, but alternative media played an important part. The media outlets that supported "Leave" stood closer to users' opinion who contributed to polarize the media's message, with pro-Leave side successfully framing some media's message in his own terms.**

Keywords: Agenda-setting; Brexit; Twitter; Public opinion; Topic modeling

I. INTRODUCTION

The current tidal wave of change in global politics seems to configure a historical phase dominated by the rise of populism. The use of social media in this cycle, from the UK EU referendum to the election of Donald Trump, or the recent election of Jair Bolsonaro in Brazil, has been deeply instrumental and controversial resulting in a widespread concern that social media may be undermining democracy [1] [2]. The relation of social media with traditional media is also complex and problematic, with recent research evidencing election coverages shrouded in partisanship, political polarization, and disinformation [3-5].

Agenda-setting theory is around for almost fifty years [6], and a wide range of literature is devoted to explaining the role of the media in society. Nevertheless, even if previous research shows that social media broke the dominance of traditional media [7] [8] [9], it also demonstrates that they continue to shape political campaigns by setting the agenda [10][11][12][13]. These pieces of evidences, apparently contradictory, remind us that we need further research to understand the complex dynamics between traditional media and social media.

In this paper, we will examine first-level agenda-setting effects. However, instead of figuring out the specific issues that were transferred to the public, we will work on an aggregated level, using each media's body of news. With this approach, we want to find out what were the media outlets that transferred their agenda successfully, to which extent, and how they relate to each other. We will use, as the case study, the Twitter debate over the United Kingdom's (UK) referendum on whether the UK should remain a member of the European Union (EU) that took place on 23 June 2016. Popularly known as the "Brexit" referendum, it divided the nation in the 'Remain' and 'Leave' camps that comprehended those who were in favor or against the UK to remain a member of the EU, respectively. At the end of the campaign, the outcome turned in favor of the withdrawal from the EU with 52% of the votes. According to Moore and Ramsay [4], it was the "UK's most divisive, hostile, negative and fear-provoking" campaign of the century, characterised by a hyper-partisan media coverage where national news outlets public endorsed their side and selected and framed the news, editorials and leader columns in accordance with it, except for the public service broadcasters that are bound to neutrality by law. The rest of this paper is organized as follows. In the next section, we summarise the literature on agenda-setting. Section III comprehends the methodology, where we describe the analyzed data and its modeling. Section IV presents the results and their discussion. Section V concerns the conclusions, and we finish with regards to limitations and future work.

II. RELATED WORK

The agenda-setting theory is one of the main perspectives in mass communication effects research and provides an explanation of the role played by the media in public opinion processes [14], it assumes that the media is the privileged medium by which voters have access to the candidates. Therefore, the media must play an important role in shaping political reality by choosing the information available to voters. Consequently, the agenda-setting effect is determined by the correlation between the media and public agendas and measures the success of the media in transferring the salience of issues to the public agenda.

In order to understand how audiences pick and choose among different media agendas in an active way, McCombs, Shaw, & Weaver (2014) identified two types of media: vertical media and horizontal media. The authors classify as "vertical media" the network and local broadcast stations that reach for the diversity of a large community and usually reflect the society's basic institutions. In opposition, the magazines, cable television, blogs, and social network sites, corresponding to the horizontal media that connect to the audience via personal interest communities, as if they stood on the same level. Consequently, the public agenda is built upon a mix of vertical, institutional and valued personal information, in a social process defined as agenda melding. Political systems that have high correlations with vertical media suggest a relatively stable social system, with reasonable alternative horizontal agenda providing challenging views to the dominant vertical agenda. McCombs, Shaw, & Weaver (2014) argue that the media agenda is community, and the lack of consensus around a civic agenda means that there is a decentralization of authority, from the core toward the periphery, a worldwide phenomenon that does not seem to come directly from changing media technology, but from the way emergent media agenda communities are melding.

TABLE II. MEDIA SUBSET STATISTICS

Media	Type	Side [4]	Tweets	%(1)	Retweets	Favorites	Unique links	%(2)	Articles	%(3)
The Guardian	Traditional	Remain	41100	3,5	82017	56215	11988	3,3	10898	90,9
Financial Times (FT)	Traditional		16980	1,5	27401	17196	9758	2,7	4487	46
The Independent	Traditional		9463	0,8	27539	17339	2494	0,7	2386	95,7
The Economist	Traditional		3821	0,3	25018	19415	845	0,2	687	81,3
Daily Mirror	Traditional		3295	0,3	16482	8927	1377	0,4	867	63
Sub-total			**74659**	**6,4**	**178457**	**119092**	**26462**	**7,3**	**19325**	
The Telegraph	Traditional	Leave	26587	2,3	73258	48820	8151	2,3	6913	84,8
Express	Traditional		19412	1,7	100051	60629	8538	2,4	8053	94,3
The Daily Mail	Traditional		15227	1,3	45525	28755	7816	2,2	7552	96,6
Breitbart	Online partisan		6561	0,6	35286	23689	2499	0,7	1802	72,1
Sub-total			**67787**	**5,9**	**254120**	**161893**	**27004**	**7,6**	**24320**	
BBC	Traditional	Neutral	22175	1,9	37602	27165	10139	2,8	3622	35,7
Sky	Traditional		5061	0,4	11589	8761	1193	0,3	834	69,9
Sub-total			**27236**	**2,3**	**49191**	**35926**	**11332**	**3,1**	**4456**	
Bloomberg	Emerging	-	10951	0,9	30226	16729	1884	0,5	1006	53,4
Reuters	News agency		10406	0,9	15738	10480	9836	2,7	9284	94,4
NY Times	Traditional		4146	0,4	6590	6971	982	0,3	918	93,5
LinkedIn	Social media		4126	0,4	2278	1910	712	0,2	590	82,9
Yahoo	Emerging		3921	0,3	2326	1725	3110	0,9	2604	83,7
Russia Today (RT)	News agency		3642	0,3	10937	7211	1908	0,5	1431	75
Politico	Emerging		3299	0,3	9137	5593	714	0,2	656	91,9
CNBC	Traditional		3099	0,3	4436	2943	880	0,2	769	87,4
Wall Street Journal (WSJ)	Traditional		2986	0,3	9560	6979	1811	0,5	1543	85,2
Sub-total			**46576**	**4,1**	**91228**	**60541**	**21837**	**6**	**18801**	
Total			**216258**	**18,6**			**86635**	**24,1**	**66902**	

a. Percentages marked with (1) and (2) refer, respectively, to the total of tweets and unique links from Table 1. Percentage (3) is the relation of number of articles collected / total of unique links.

From this perspective, if we measure the correlation between vertical and horizontal media, we will be able to define the stability of a political system. Intermedia agenda-setting focuses on the transference of issue salience across media [16], and with the rise of social network sites (SNS), there is a growing literature dedicated to the research of the relationship between traditional and social media. Results have shown that despite the transformations in media environments, the traditional media are still successful in influencing issue salience in SNS [10], and at a faster rate [12][13][17]. This is the first-level in agenda-setting, the power to dictate what the public thinks about. Nonetheless, this does not imply that they have the ability to define how the public thinks about the issues, which is the second-level in agenda-setting. In this sense, studies report that SNS provide an environment where complex and dynamic interactions defy the traditional media's power to set the interpretative agenda [7][18][9][11][19]. Nevertheless, research on election coverages reports evidence of partisanship, political polarization, and disinformation, supporting to the "echo chambers" view that Twitter and other social media create networks prone to homophily and bot manipulation [17] .

III. METHODOLOGY

A. Data collection and sample description

We used a python script based on the TwitterScraper [20] algorithm to collect the data from Twitter containing the "Brexit" keyword. This process yielded a dataset with 1.163.311 tweets, from 236.196 users, corresponding to the time interval of the year before the referendum (June 2015 – July 2016).

TABLE I. DATASET STATISTICS

Tweets	Links shared	Unique links	Domains	Users
1163311	655709	360105	57390	236196

From the original dataset, we selected the tweets that shared links, but we noticed that URL shortening services had generated a significant part of the links. Consequently, we had to check each URL for its original version to retrieve the domain of the website. This yielded a subset with 360.106 unique links from 57.390 domains, from which we chose, for the sake of feasibility, the top 20 media in terms of total count (Table 2). Amongst this top list of domains, there were social networks sites that we deliberately removed. We excluded Twitter itself, due to the fact that concerns inbound sharing of images and links to other tweets, and YouTube because it is out of the scope of this research to analyze video. We made an exception regarding LinkedIn because it mostly comprises professional insights about the referendum.

After this procedure, we collected the web pages linked in the tweets, and extracted their titles and text. Some of these links were not accessible, either because they were already removed by the publishing media, or due to errors in the URL. Therefore, we collected past versions of the articles via the Internet Archive's Wayback Machine API [21], or, if this failed, gathered the correct URL through the Google CustomSearch API [22].

B. Data modeling

From the collected data, for each of the 20 top media, we created three subsets concerning the text of the tweets, the titles of the articles shared in the tweets, and the text of these articles. We used unsupervised machine learning method - topic modeling - to discover the latent topics in the collections. Our model relied on Latent Dirichlet Allocation (LDA) [23], a generative probabilistic model that presents topics as multinomial distributions over words, where each document in a collection is described as a mixture of topics. Previous research shows that short documents like tweets may not contain sufficient data to build robust machine-learning models, recommending topic model training on aggregated messages in order to obtain better performance [24]. This approach suits our research needs since we want to discover the latent topics for the 20 top media present in the dataset. Therefore, we aggregated the three previously enunciated subsets for each media and generated the bag-of-words necessary for the LDA model, applying stop-word removal, lemmatization, POS tagging and, finally, TF-IDF weighting to filter out common words.

In terms of the parameters used to tune the LDA model, our experiments revealed meaningful topics when training models with 10 topics. These LDA models provided us with an output of 10 topics, from each we chose the top 15 keywords that are representative of the topic, composing a matrix for the three media subsets.

To discover agenda-setting effects, we performed three sets of analysis where we evaluated the similarity between subsets. Assuming that subsets with similar topic distributions should have smaller distances between them, we used the standard Euclidean distance as a metric of the difference between subsets.

Firstly, we computed the distance of the titles and text of the articles and the tweets that contained links of each media in order to understand the users' level of endorsement of the media's message. Smaller distances here represent the user tacit agreement with the message that the media is conveying in the article. There are some concerns regarding the comparison between the text of the tweets and the text body of the articles. As they are structurally different in semantic terms, the LDA model may generate topic distributions that would not lead to safe conclusions. To overcome this hurdle, we compared also the media titles, which normally follow the 140-character structure of the tweets, and

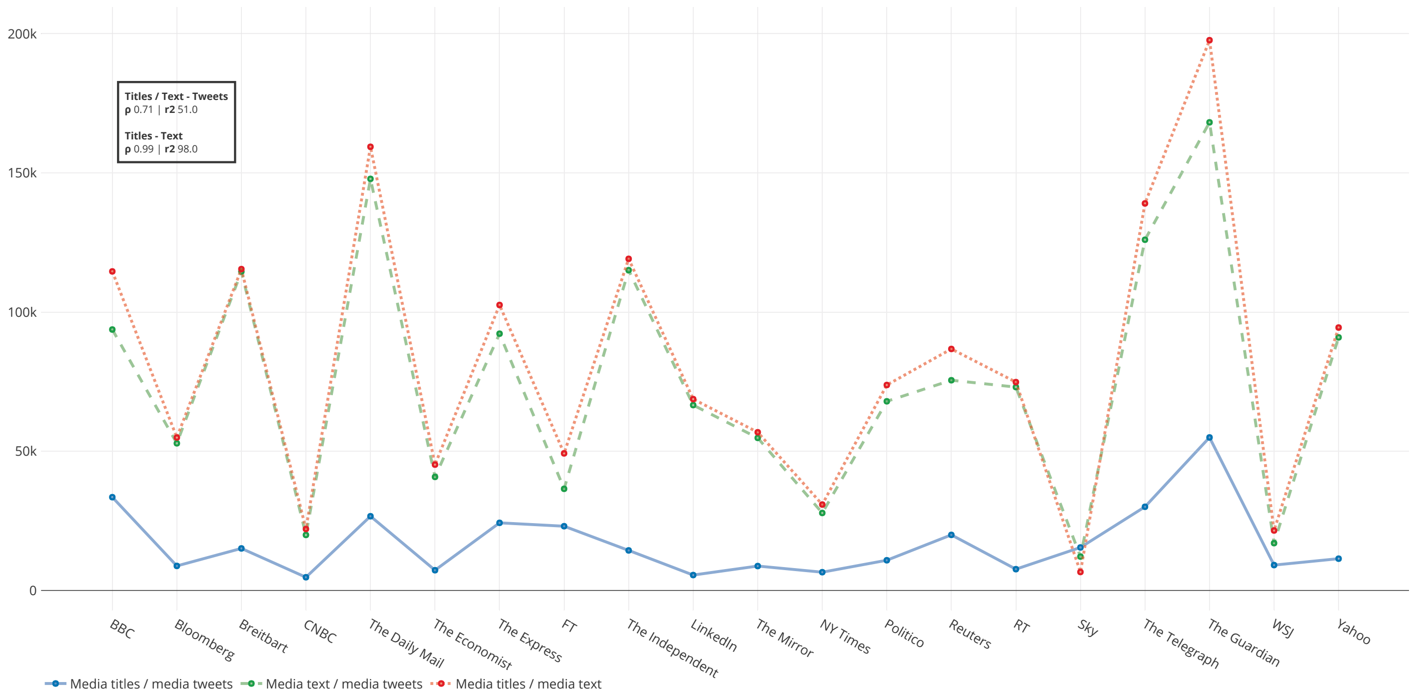

Figure 1. The Euclidean distance between the article titles (solid line) and text (punctuated line), and the tweets that contained links of each media (Pearson: 0.71 | r2: 51.0)). The correlation of the distance between the article text and tweets (dashed line), and the distance between article titles and text is Pearson: 0.99 | r2: 98.0.

the articles' text body to infer if the results derive from structural differences, or if they have semantic meaning indeed. To reinforce the confidence in these measures, we computed Pearson correlations between them.

Secondly, we calculated the similarity between the media's subsets and the corpus comprising all the tweets in the dataset (with and without shared links). This approach enabled to identify to which extent the media message represents the general opinion in the dataset.

Finally, we computed the similarity between media subsets to perceive how the media's discourse and user-generated content relate to each other. To visualize these relations, we created distance matrices from which we produced dendrograms through a method of hierarchical cluster analysis.

IV. RESULTS AND DISCUSSION

The media subset comprises almost 20% of the total tweets, from which was generated approximately 25% of the linked content. This is quite expressive if we think that these 20 domains are 0,03% of a 'long tail' of domains. In terms of media types, we have 13 traditional media outlets, a clear signal that the established news media still dominates, which is in line with previous research [25]. Nevertheless, the presence in the top 10 of Breitbart, a hyper-partisan right-wing media outlet that was instrumental in the election of Donald Trump [3], reveals that, in social media sites, alternative media can stand alongside traditional media. Regarding campaign endorsement, five traditional media outlets publicly supported the "Remain" campaign, while three chose the "Leave" side [4]. Grouping the media accordingly, findings reveal that the "Remain" side dominates in terms of tweets counts, but under-performs when it comes to retweets and favorites. This is especially visible when comparing The Guardian with the Ex-

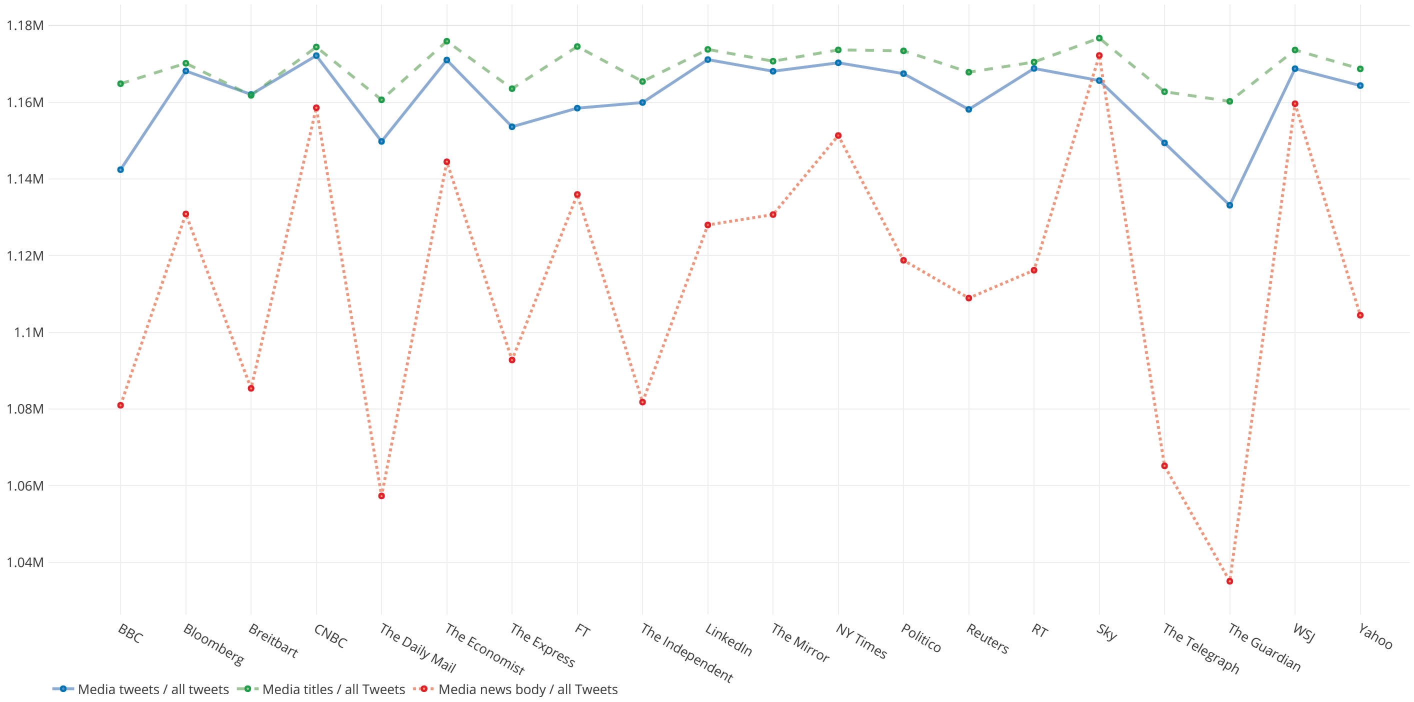

Figure 2. The Euclidean distance between media's tweets (solid line), article titles (dashed line) and text (punctuated line), and all tweets in the dataset.

press. The latter has almost half of the tweet count but scores more retweets and favorites.

Figure 1 summarizes the first set of experiments where we computed the similarity between the article titles and text, and the text of the tweets for each media. Results show that the article titles are more similar to the text of the tweets than the text of news article. This could derive from the structural similarities between the titles and 140-character anatomy of the tweet. Or from the fact that social media management software and widgets, like Hootsuite, have the title as text default when sharing web pages in social media. However, when we analyze the similarity of titles and text of the articles, we see that values almost overlap with the distance of the article text to the tweets, with a Pearson correlation of 0.99. For this reason, it is safe to conclude that these distances are semantically representative of the similarity between corpora. From this premise, we can affirm that, in a general way, the titles of the articles reflect the message in the article, and that the users distanced themselves from this message according to each media. In this way, higher distances indicate a lower level of endorsement of the media's message. This does not necessarily mean that the users disagreed with the media's message, but that they added information to the original message, whether to reinforce it or to subvert it. Media like The Guardian, BBC, The Telegraph or The Daily Mail show less similarity. On the other hand, users that shared links from CNBC, WSJ, LinkedIn, The NY Times, Bloomberg, The Economist, or RT tend to agree more with the messages present in the titles of these media. It seems that in a campaign where the economy was the most-covered political issue [4], the business-related media outlets (with exception of the Financial Times), tended to be less disputed than more generalist news media. A tendency that may be explained by the fact that Britons tend to regard the latter media as biased [26].

In the second experience (Figure 2), regarding the similarity between the media subsets and the tweet corpus as a whole, we observe that the distances in titles and tweets change according to the trend described in the previous experiment. But, surprisingly, the text of the articles is closer to the overall opinion. Given what we've previously seen, we would expect that that should be the case of the articles titles, but results indicate otherwise. A possible explanation is that the readers that share the articles tend to follow the title of the article, but when engaging in debate end up gravitating around what they read in the article. The media closer to the general opinion are The Guardian, The Telegraph, The Daily Mail, BBC, and The Independent. These were precisely the same media that were less endorsement by the users. The Guardian example is paradigmatic, it has an article text corpus more similar to the general opinion, but at the same time shows a lower message acceptance by the users sharing its articles. In the same logic, but in less extent, follow Breitbart and the Express. In the opposite direction, media that previously revealed higher approval, like CNBC, Sky, WSJ, The Economist, or The NY Times, reveal now less similarity with the overall opinion. In fact, when correlating the distances of the titles and text to each media's tweets and all tweets, we obtain a Pearson value of -0.71, in the case of the titles, and a value -0.99 for the text, which are the exact opposite correlation values obtained in the prior experiment. These results are in line with prior research [7][19], in regard that traditional news media have the power to dictate what the public thinks about (first-level agenda setting), by influencing the salience of the issues that Twitter users discuss. Nonetheless, this does not necessarily imply that they have the ability to define how the public should think about that issue (second-level agenda setting). In this case, and knowing the outcome of the referendum, this especially applies to the media that supported the "Remain" campaign.

In relation to the distances amongst media, from the articles' titles (Figure 3), we can observe two main groups. The first with The Guardian standing alone. The second group comprises two sub-groups, The Daily Mail appears isolated and the rest of the media is grouped in the other. The latter group subdivides in a

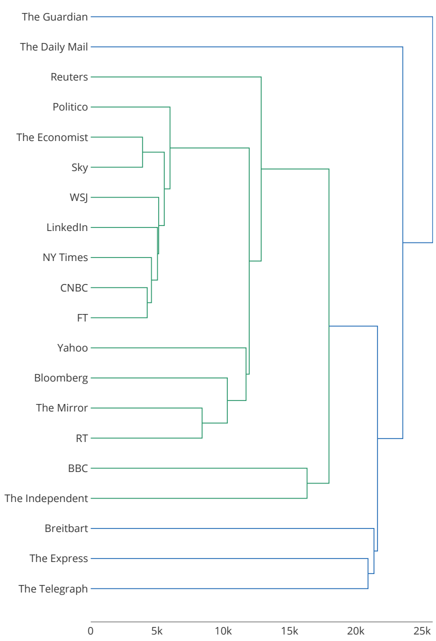

Figure 3. The Euclidean distance dendogram between media's article titles.

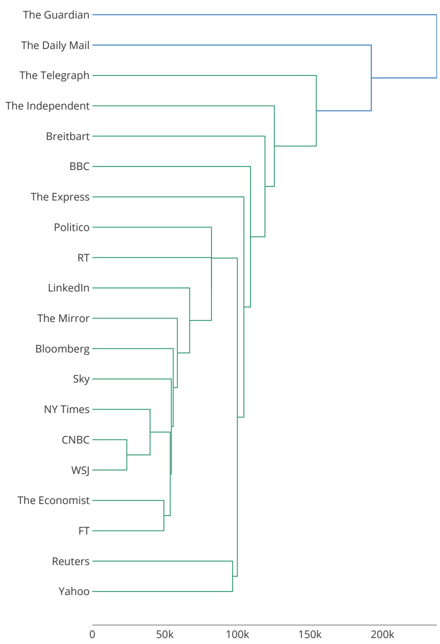

Figure 4. The Euclidean distance dendogram between media's article text.

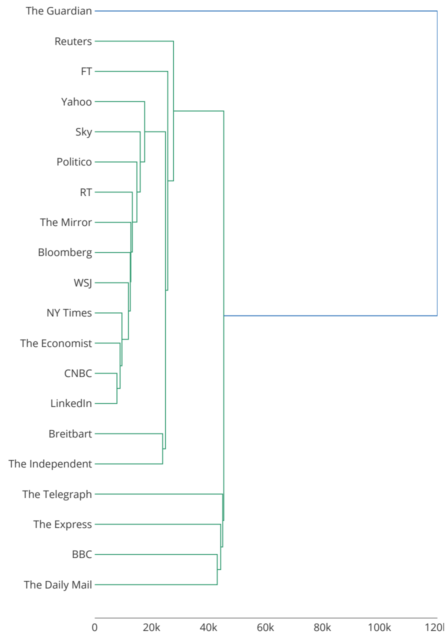

Figure 5. The Euclidean distance dendogram between media's article tweets.

cluster with the media that supported the "Leave" campaign, and the remainder media outlets in the other. Partisanship dynamics are evident here, especially on the "Leave" side. The proximity of traditional media like The Telegraph and the Express to Breitbart seems to follow the same pattern of inflection of the center-right to a more extreme form of right-wing politics, reported in the 2016 U.S. Presidential elections [3]. On the opposite campaign, the media ecosystem was closer to the business and financial media outlets, and other generalist media, which it is also not surprising due to the importance of the economic issues in the campaign. The Guardian relative distance to "Leave" supporters, like The Daily Mail, may be justified by the dynamics of an acrimonious campaign [4], where contenders end up debating the same issues.

Concerning the articles' text (Figure 4), the tendencies in the articles' titles are also present, with The Guardian detached from the rest, followed by The Daily Mail. The rest of the media are in a separate group according to with Figure 4. There is a higher range of distances (200k against 25k in the first experiment), which is expectable due to the fact that we are dealing with more information. Nonetheless, in the overall and comparatively, the media are closer. In relation to each media's tweets (Figure 5), The Guardian appears further isolated from the remainder of the media, that appear closely tied with each other. It is visible a compression most of the distances within media, at the same time that an extensive gap is created, a clear sign of the divisive and polarized character of the campaign. These results seem to resonate with "echo chamber" effect described in previous research [17], [25]. One explanation to these findings is that the "Leave" campaign successfully set the interpretative agenda, framing some media's message in his own terms, while literally creating distance from others. This explains the proximity of media like Breitbart and The Independent and the distance of The Guardian. In relation to each media's tweets (Figure 5), The Guardian appears further isolated from the remainder of the media, that appear closely tied with each other. It is visible a compression most of the distances within media, at the same time that a extensive gap is created, a clear sign of the divise and polarized character of the campaign. One feasible explanation to these findings is that the "Leave" campaign successfully set the interpretative agenda, framing some media's message in his own terms, while literally creating distance from others. This explains the proximity of media like Breitbart and The Independent, and the distance of The Guardian.

V. CONCLUSIONS

This study explores first-level agenda-setting effects during the UK's EU Referendum campaign. We examined the dynamics of the interaction between users and the media content that users shared on Twitter using topic modeling. Results indicate that the traditional media outlets dominated the debate, but not unchallenged, with alternative media playing an important part in the campaign. Regarding the interaction dynamics between the users and media content, our findings suggest that the titles of the news articles are closer to the users' opinion that the text of these articles. But, when comparing with the overall debate, the text of the articles is closer to the general opinion. In terms of the relationships between media messages, our method was able to identify media partisanship dynamics in campaign coverage, in the same terms as reported in previous research [3-4]. Finally, we observed that the user-generated content contributed to polarizing the media's message, with pro-Leave side successfully framing some media's message in his own terms, while literally creating distance from others.

VI. FUTURE WORK

This study has some limitations. First, as the case study is focused on a small dataset of the political debate over the EU Referendum on Twitter, we cannot infer generalizations from the results or presume that they are fully representative of the dynamics of public opinion. Secondly, the unsupervised machine-learning techniques used in this paper need further exploration. Therefore, future research could apply the method used in this paper to other political events and datasets, to verify the reproducibility of the present results. In addition, further study could also concern model improvement with word embeddings [27], or the application of Non-negative Matrix Factorization instead of LDA [28].

VII. ACKNOWLEDGMENT

This work is supported by the Fundação para a Ciência e Tecnologia (FCT), Portugal project grant UID/Multi/04466/2019.

REFERENCES

[1] J. A. Tucker, Y. Theocharis, M. E. Roberts, and P. Barberá, "From Liberation to Turmoil: Social Media And Democracy," J. Democr., vol. 28, no. 4, pp. 46–59, 2017.

[2] P. Gerbaudo, "Social media and populism: an elective affinity?," Media, Cult. Soc., vol. 40, no. 5, pp. 745–753, 2018.

[3] R. Faris, H. Robert, B. Etling, N. Bourassa, E. Zuckerman, and Y. Benkler, "Partisanship, Propaganda, and Disinformation: Online Media and the 2016 U.S. Presidential Election," vol. 7641, 2017.

[4] M. Moore and G. Ramsay, "UK media coverage of the 2016 EU Referendum campaign," Policy Inst. King's, no. May, pp. 1–188, 2017.

[5] J. A. Tucker, A. Guess, P. Barbera, C. Vaccari, A. Siegel, S. Sanovich, D. Stukal, and B. Nyhan, "Social Media, Political Polarization, and Political Disinformation: A Review of the Scientific Literature," Ssrn, no. March, pp. 1–95, 2018.

[6] G. Bolsover and P. Howard, "Computational Propaganda and Political Big Data: Moving Toward a More Critical Research Agenda."

[7] S. Meraz, "The fight for 'how to think': Traditional media, social networks, and issue interpretation," Journalism, vol. 12, no. 1, pp. 107–127, 2011.

[8] G. R. Boynton and G. W. Richardson, "Agenda setting in the twenty-first century," New Media Soc., no. December 2014, p. 1461444815616226, 2015.

[9] W. R. Neuman, L. Guggenheim, S. Mo Jang, and S. Y. Bae, "The Dynamics of Public Attention: Agenda-Setting Theory Meets Big Data," J. Commun., vol. 64, no. 2, pp. 193–214, Apr. 2014.

[10] A. Shehata and J. Stromback, "Not (Yet) a New Era of Minimal Effects: A Study of Agenda Setting at the Aggregate and Individual Levels," Int. J. Press., vol. 18, no. 2, pp. 234–255, Apr. 2013.

[11] L. Guo and C. J. Vargo, "The power of message networks: A big-data analysis of the network agenda setting model and issue ownership," Mass Commun. Soc., vol. 18, no. 5, pp. 557–576, 2015.

[12] B. A. Conway, K. Kenski, and D. Wang, "The Rise of Twitter in the Political Campaign: Searching for Intermedia Agenda-Setting Effects in the Presidential Primary," J. Comput. Commun., vol. 20, no. 4, pp. 363–380, 2015.

[13] B. A. Conway-Silva, C. R. Filer, K. Kenski, and E. Tsetsi, "Reassessing Twitter's Agenda-Building Power: An Analysis of Intermedia Agenda-Setting Effects During the 2016 Presidential Primary Season," Soc. Sci. Comput. Rev., vol. 36, no. 4, pp. 469–483, 2018.

[14] T. Takeshita, "Current critical problems in agenda-setting research," Int. J. Public Opin. Res., vol. 18, no. 3, pp. 275–296, Sep. 2006.

[15] M. E. McCombs, D. L. Shaw, and D. H. Weaver, "New directions in agenda-setting theory and research," Mass Commun. Soc., vol. 17, no. 6, pp. 781–802, 2014.

[16] M. McCombs, "A Look at Agenda-setting: past, present and future," Journal. Stud., vol. 6, no. 4, pp. 543–557, 2005.

[17] Y. Gorodnichenko, T. Pham, and O. Talavera, "Social media, sentiment and public opinions: Evidence from #Brexit and #USElection." 2018.

[18] C. J. Vargo, L. Guo, M. Mccombs, and D. L. Shaw, "Network Issue Agendas on Twitter During the 2012 U.S. Presidential Election," J. Commun., no. 64, pp. 296–316, 2014.

[19] A. Ceron, L. Curini, and S. M. Iacus, "First- and second-level agenda setting in the Twittersphere: An application to the Italian political debate," J. Inf. Technol. Polit., vol. 13, no. 2, pp. 159–174, 2016.

[20] S. Lindgren, "TwitterScraper." [Online]. Available: https://github.com/simonlindgren/TwitterScraper. [Accessed: 17-Mar-2017].

[21] "Wayback Machine APIs | Internet Archive." [Online]. Available: https://archive.org/help/wayback_api.php. [Accessed: 01-Oct-2018].

[22] "Google APIs Explorer." [Online]. Available: https://developers.google.com/apis-explorer/#p/customsearch/v1/search.cse.list. [Accessed: 01-Oct-2018].

[23] D. M. Blei, B. B. Edu, A. Y. Ng, A. S. Edu, M. I. Jordan, and J. B. Edu, "Latent Dirichlet Allocation," J. Mach. Learn. Res., vol. 3, pp. 993–1022, 2003.

[24] A. O. Steinskog, J. F. Therkelsen, and B. Gambäck, "Twitter Topic Modeling by Tweet Aggregation," Proc. 21st Nord. Conf. Comput. Linguist., no. May, pp. 77–86, 2017.

[25] G. Gorrell, I. Roberts, M. A. Greenwood, M. E. Bakir, B. Iavarone, and K. Bontcheva, "Quantifying Media Influence and Partisan Attention on Twitter During the UK EU Referendum," 2018.

[26] YouGov, "Is BBC News pro-Brexit or anti-Brexit?" 2018.

[27] C. Li, Y. Duan, H. Wang, Z. Zhang, A. Sun, and Z. Ma, "Enhancing Topic Modeling for Short Texts with Auxiliary Word Embeddings," ACM Trans. Inf. Syst., vol. 36, no. 2, pp. 1–30, Aug. 2017.

[28] T. Shi, K. Kang, J. Choo, and C. K. Reddy, "Short-Text Topic Modeling via Non-negative Matrix Factorization Enriched with Local Word-Context Correlations," Proc. 2018 World Wide Web Conf. World Wide Web - WWW '18, pp. 1105–1114, 2018.