THE EUENGAGE WORKING PAPER SERIES.

Assessing the complexity of European polity in time of crises.

Network analysis of Brexit discussion on social media

Kenneth Benoit
London School of Economics and Political Science, K.R.Benoit@lse.ac.uk

Akitaka Matsuo
London School of Economics and Political Science, A.Matsuo@lse.ac.uk

ABSTRACT:
We analyze the engagement of citizens, media and politicians on social media during the campaign period of Brexit referendum in the UK in June 2016. We focus on the social media conversation on Twitter, the largest micro-blogging service with more than 300 million active users and approximately 500 million new messages generated per day. We analyse the networks of Tweets focusing on both the topics users have discussed during the referendum campaign and the network of followerships. Using machine learning, we classify users as favouring Leave or Remain, and use this information to compare the estimates of the topics they discuss in the debate. The topic analysis reveals that the structure of topics by Leave users are more densely related than those by Remain users. Furthermore, by analyzing the network analysis of followership, we reveal that the important accounts in the Brexit debate such as media and politicians are categorically different between Leave and Remain, indicating a fundamentally divided messaging communication structure, reinforcing work done elsewhere on how social networks form ideological "echo chambers" reinforcing one's pre-existing attitudes.

KEYWORDS: Social Media, Brexit, Network Analysis, Text Analysis

This project has received funding from the *European Union's Horizon 2020 research and innovation programme* under grant agreement No 649281

1. Introduction

In this paper, we analyze the engagement of citizens, media and politicians on social media during the campaign period of the "Brexit" referendum held in the UK in June 2016. The particular focus is placed on the social media conversation on Twitter, the largest micro-blogging service with more than 300 million active users and over 500 million new messages generated per day. We analyse traits of the networks of Tweets focusing on both the network of topics users have discussed during the referendum campaign and the network of followership. The purpose of our analysis is to detect the differences between Leave and Remain for various aspects of network of Twitter communications, in particular the network of topics each side of the debate (Leave or Remain) used in Twitter conversations, and also the network of follower-followee relations. The former focuses on the structure of Tweet contents, and the latter focuses on the information environment of Twitter users who participated in conversations on the Brexit referendum on social networking platforms.

The analysis of topics reveals that the structure of topics by Leave supporters are more densely related to one another than those used by Remain supporters. The network analysis of followership reveals that the important accounts in the Brexit debate such as media and politicians are categorically different between Leave and Remain. This observed division indicates a fundamentally divided structure for conveying and receiving political communication, indicating that similar to results shown in other contexts (Flaxman, Goel and Rao, 2016; Del Vicario et al., 2016; Allcott and Gentzkow, 2017), social media in the Brexit context formed similar ideological “echo chambers” (Sunstein, 2001) in which users reinforced their pre-existing attitudes by following accounts with like-minded content.

2. Data

In order to capture the discussion on the Brexit referendum in Twitter in its entirety, we set up the Twitter downloader through an access to the Twitter “firehose”, the high-capacity live stream of all Tweets, which guaranteed the delivery of all posts matching the capture criteria. Another option to capture Tweets is to use Twitter’s Streaming API[1], which delivers a random sample of Tweets up to one percent of all Tweets generated at a given time.

Table 1. Capture Criteria

Search Criteria	Terms
Simple words	brexit
Hashtags	#betterdealforbritain #betteroffout #brexit #euref #eureferendum #eu-summit #getoutnow #leaveeu #no2eu #notoeu #strongerin #ukineu #voteleave #wewantout #yes2eu #yestoeu
User screen names	vote_leave brexitwatch eureferendum ukandeu notoeu leavehq ukineu leaveeuofficial ukleave_eu strongerin yesforeurope grassroots_out stronger_in

The use of the streaming API is preferred in many studies because any Twitter users have an

[1] https://dev.Twitter.com/streaming/overview

access to this public API. For our purposes, the capturing method also provided a more accurate picture of Twitter conversation for our research purposes, because the volume of Tweets in pronounced events such as Brexit might exceed the Streaming API's rate limit.[2] Through the trial and exploration, we selected a set of search criteria for capturing Tweets, based on a core set of unambiguously Brexit-related terms.

The search terms we used are presented in Table 1. The capture terms consists of three sets. The first is a general search term ("Brexit"), the second consists of prominent hashtags (hyperlinks on Twitter prefaced by the "" symbol) related to Brexit, and the third consists of Twitter usernames of groups and users clearly set up for the purposes of communicating about Brexit (e.g. @voteleave).

Any Tweets that contain one of these terms were captured in our data collection which we started on January 6, 2016 and ran through July 2016, the month following the referendum.

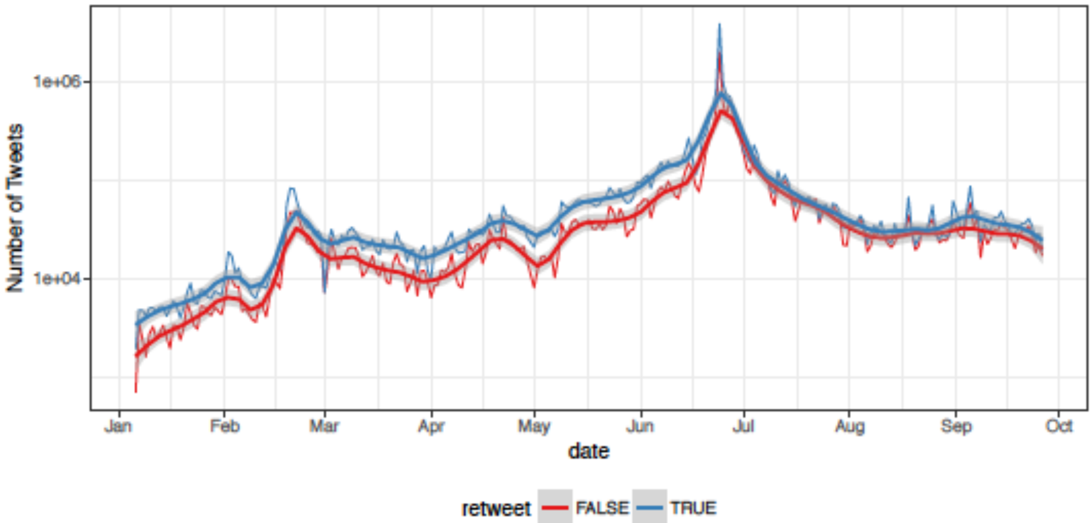
Our download efforts captured a large number of Tweets. From the start to the end of June 2016, 26 million Tweets were collected, where 10 million Tweets are original Tweets (as opposed to retweets). Our Twitter data has a very long tail. The data includes about 3.5 million Twitter users in which majority of accounts sent only a single Brexit-related Tweet. There are a small number of users who generated a very large number of Tweets. Some of the highest volume accounts, which generated tens of thousands of Tweets during this time, are obviously spam accounts, and we try to exclude these accounts.[3] Figure 1 is a time series of the volume

[2] For the detailed comparison between Firehose and Streaming API, see Morstatter et al. (2013)
[3] These spam accounts typically exclusively sent re-tweets, and also despite very high volume of Tweets,

of captured Tweets on the logarithmic scale. Each line shows the volume of original posts and reTweets on each day.

The day after the Brexit referendum (24 June) recorded the largest volume of Twitter transactions. Over 13 million Tweets were recorded on the day of the referendum itself.

Figure 1. Volume of Brexit Tweets (Jan - Sep 2016)



Note: y-axis is on a logarithmic scale

3. Classifying Remain v. Leave Accounts

The first, crucial task from the large corpus of Tweets with millions of users is to predict the

their names are rarely mentioned by other Twitter accounts.

sides (Leave or Remain) of Twitter accounts expressed in their Tweets on Brexit conversation. We conduct this task by generating a training set consisting of human-classifed accounts to generate a core training set, and using supervised machine learning to predict the side of the remaining users. The work is conducted through the several steps. The first step is to identify a few hundred important accounts and manually classify these accounts into Leave or Remain. We selected these important accounts based on the number of mentions to a user account in Tweets on Brexit. A user mention is a part of Twitter text that contains a user's screen name followed by the "@" character. This feature of Twitter makes a Tweet with a reference to specific users searchable, and is used to aim a Tweet at a mentioned user.

For top 200 most mentioned accounts, our research collaborators visited in December 2016 the website of their Twitter accounts to check whether these accounts are still active and their position in the Brexit referendum. We found that fifty-five accounts are on Leave and twenty-five are Remain. From this list of known accounts, we extend the list of Twitter accounts on each side through the analysis of hashtag usage. A hashtag is an alpha-numeric string with a prepended hash ("#") in social media texts. Inclusion of hashtags will make Tweets easier to find and deliver messages to other Twitter users who share the interests with the user who posted the original Tweet with hashtags. We identify the hashtags disproportionally used by one of two sides by looking at the hashtag use of human-coded accounts,[4] and identify other

[4] Hashtags by the Remain side are #strongerin, #intogether, #infor, #votein, #libdems, #voting, #incrowd, #bremain, and #greenerin. Hashtags used by the Leave side are #voteleave, #inorout, #takecontrol, #voteout, #takecontrol, #borisjohnson, #projecthope, #independenceday, #ivotedleave, #projectfear, #britain, #boris, #lexit, #go, #takebackcontrol, #labourleave, #no2eu, #betteroffout, #june23, and

Twitter users active in the Brexit discussion (i.e. more than 50 Tweets in our data).

We use this list of hashtags to find a large number of Leave and Remain Twitter accounts from the accounts active in the Brexit referendum debate on Twitter, about 15,000 accounts with more than 50 Tweets in our Brexit Twitter corpus. We calculated the Leave hastag score of each account by counting the number of Leave hashtags subtracted by the number of Remain hashtag. We select the top (or bottom) ten percent user accounts as the Leave (or Remain) as the training data.

The next step is the classification of Twitter accounts into "Leave" and "Remain" using this training data. For the simplicity and interpretability of outcomes, we use a simple Naive Bayes classifer with a flat prior. For creating the feature matrix, we employ a "bag-of-words" methodology. We first combine all Tweets from each account and then extract features. Since the document feature matrix is extremely sparse (over 97 percent sparsity) even only using unigrams, we do not try n-grams. We apply lower-casing to all words and then stem them, except for Twitter specific entities (hashtags and mentions). Features with a document frequency less than 0.1 percent were also removed.

Table 2: Average Accuracy By Feature Selections

Feature Type	Accuracy	N
All Features	0.93	64206
Hashtags and Mentions	0.88	18705
Hashtags	0.85	7972
Mentions	0.87	10733

#democracy. Since these hashtags are used for training data generation, we removed them from the features in the classification model.

The next step is the selection of features. The concern is the overfitting of the model. Even after the cleaning of the data, there are over 60,000 features left for only 3,000 training documents.

To check whether the inclusion of all features is appropriate, we tested several different selections of features. The setting we have tested are: hashtag-only, mentions-only, hashtags and menstions, and all features. For each of these four settings, we run a ten-fold cross-validation ten times each, and calculated the average accuracy for out-of-sample predictions.[5] Table 2 report the outcomes.

All settings are reasonably accurate, but the use of all features outperfomed the other approaches, and we therefore decided to use that model for generating our Leave versus Remain labels on the full set of user accounts.

We estimate the model with full training data. The estimated model provide two sets of predictions: the first is the probablity of each feature belong to the Leave or Remain, and the second is the predicted probability of each account belongs to the Leave or Remain side. Figure 2 depicts the relative frequencies of the hastags used by each side, plotting the size of each hashtag proportional to its relative frequency, as a “word cloud”. The partitions distinguish the usage by the partitions of our classifer.

[5] Since both “Remain” and ”Leave” are important outcomes, we report the accuracy rather than other statistics that focus on positive predictability, such as precision or recall.

Figure 2. Word Cloud of High Frequency Hashtags

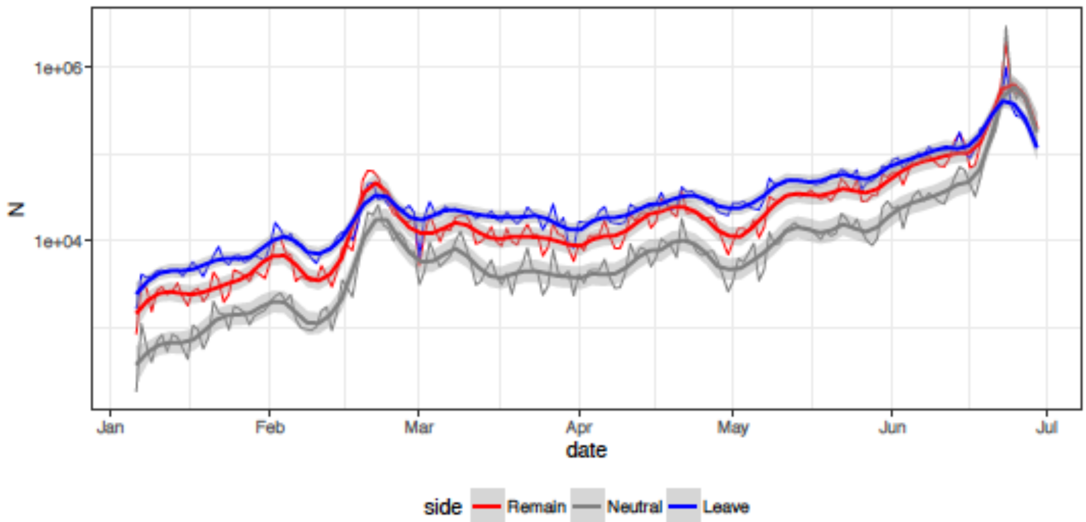
Note: Top seventy hashtags in each category. The categories of hash tags are generated from the predicted probability of a hashtag to belong to Leave (0-0.25: Remain, 0.25-0.75: Neutral, 0.75-1: Leave). Size of hashtag is proportional to the frequency. Color indicates the category.

After training the NB classification machine, we apply it to all three million accounts. The classification results at Tweet level are shown in the table and figure below. As Figure 2 indicates, the Tweets by Leave accounts exceed the volume of Tweets by Remain accounts in most of the time approaching to the referendum. The greater number of total Tweets by Remain (Table 4) is the result of the massive volume of Tweets after the referendum. The day after the referendum (24 June, 2016) recorded the highest volume in our data and the majority Tweets on the day were made by the accounts our model classified as Remain.

Table 3: Classification Results of Tweets

Side	Count	Oct.
Remain	9780223	36.93%
Neutral	7786297	29.40%
Leave	8914207	33.66%

Figure 3 Time-series of Leave and Remain Tweets (Jan June 2016)



Note: y-axis is on a logarithmic scale

4. Network of topics

Applying the Structural Topic Model (STM) by Roberts et al. (2014, 2015), we attempt to identify the topics in Twitter conversations and their dynamic across time and by side of the debate. STM is a methodology that is well-suited to identifying the topics in the debate on Twitter, their associations with each side, and their chronological evolution. The strength of STM, which is built upon other models to detect the topics, such as the Latent Dirichlet Allocation model, is that it allows the incorporation of additional covariates to estimate the effect of these factors on the use of topics.

There are other methods for measuring topics that can be used for this purpose, such as Twitter-LDA (Zhao et al., 2011) or Author-Topic-model (Xu et al., 2011). However, these models are not particularly suitable for our research purposes, because these models are more focused on the classification of individual documents, and do so by setting restrictions on the topic distributions of each document. A simple LDA model (Blei, Ng and Jordan, 2003) and STM, which is based on the LDA, assume that each document is a mixture of topics. This is more suitable for our data aggregation strategy detailed below. Also, the feature of STM to allow the use of additional covariates as the determinants of prevalence of each topic is particularly useful for this research, because one of the main interests for us to fit a topic model is to explore the temporal dynamics of evolution of topics by each side of Leave and Remain. The unit of observation in the model is a text of combined Tweets created by each user in a given week.[6] The Twitter users' positions in Brexit are the prediction from the Naive Bayes model described in the previous section. There are two reasons why we use this aggregation strategy. First, by combining Tweets from each user, we can provide more information to the model to improve the performance of the classifier (Hong and Davison, 2010). Second, we combine Tweets for each week so that we can capture the dynamics of topics. The side is a binary variable that takes a value of either Leave or Remain. Neutral users are removed from the data because many users are the low frequency Tweeters (on Brexit, at least) who might not provide the enough information to our Naive Bayes classifier to clearly determine their side. In the model specification, we used two covariates and their interaction. One is Side and another is Week. The Week variable takes an integer value that indicates a week number in the year of 2016, and converted to a b-spline to prevent over fitting. An interaction between Side and Week is included as we expect that the topics discussed do not move in parallel across sides. In STM, the number of topics (k) has to be fed, and we chose k = 40 after the exploration of k through 10, 20, 30 and 40.

STM, like other topic models, is a method of unsupervised machine learning that does not require pre-coded training data. This means that it is up to researchers' qualitative judgment to interpret the estimated topics means and give them substantively understandable labels (Chang et al., 2009; Wallach et al., 2009). The judgment is based on the high-probability or exclusive words for each topic. Out of forty topics we estimated, thirty-five seem to have consistent themes while five do not provide consistent, interpretable contents. Depending on the

[6] A week starts at 12 am GMT on Monday.

effect of Side on the prevalence of topics, we categorize topics into Leave, Remain, and Neutral. Table 4 provides the labels we applied to our topics (see the appendix for a more detailed presentation of the proportion of topics by Leave versus Remain side, used to determine the association in the table).

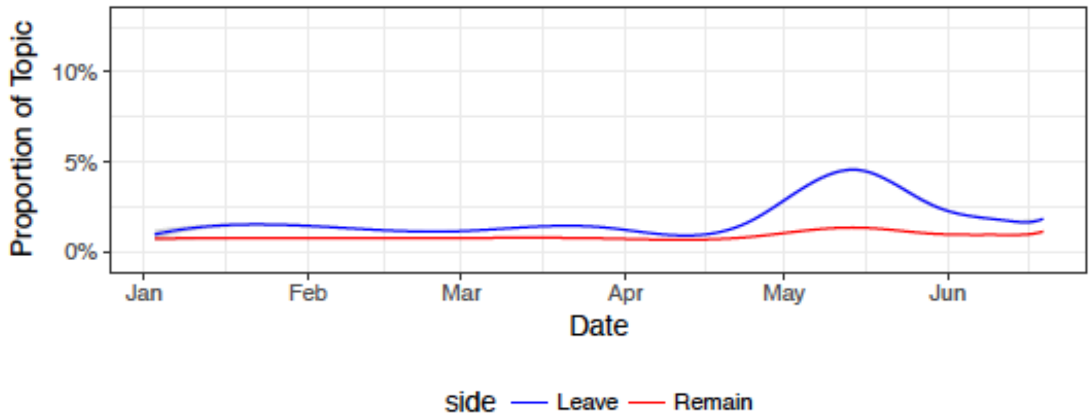
The time variable, another covariate of the model, allows us to check the dynamics of topics by each side. The following two figures are examples of the topic evolutions. The horizontal axis indicates the week of the year in 2016 (the referendum was held on Week 26) and the vertical axis is the proportion of topic in the entire Twitter conversation in the week. The first figure is on the topic "Brexit: The Movie", which is a documentary film created by a film director who advocates the UK concession from EU, had a significant spike on the week of its release date (11 May 2017) from Leave (blue line), while there is no response to this topic from Remain side Twitter accounts (red line).

There are also multiple topics from Remain sides. The following figure is one of such Remain topics. We consider this is a topic on the "Financial Risk" or financial concerns about the Brexit because of its high probability words (e.g. brexit, fears, vote, stocks, markets, global). The topic had been always discussed by Remain Twitter users throughout the time-period until the time of the referendum, and there was a huge spike of this topic in a couple of weeks before the referendum.

Table 4 Detected Topics

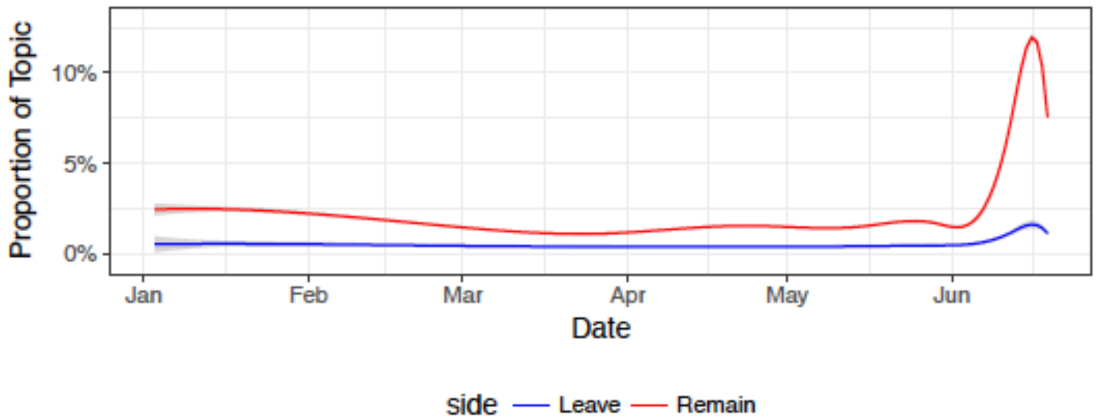
Side	Topic Title
Remain	Brexit ideologues
	Economic consequence of Brexit
	Encourage participation
	Exchange rate
	Financial risk
	Globalization and migration
	Ireland
	Obama in London
	Stock market risk
	StrongerIn
	Tabloid (Trump, Queen, Jo Cox)
	Talking about articles on Brexit
	The city, and big business
	Voter registration
Leave	Brexit Movie
	Brexit, UKIP, Leave EU
	David Cameron and Brussels
	David Camerons lies
	Debate and discussions
	Economic/Financial Sovereignty
	Free trade and migrants
	General leave argument
	Jobs and social security
	Leave campaign
	News discussion
	Reasons to Leave EU
	Take back control
	Undemocratic EU
	Vote Leave
Neutral	Fiscal burden of EU membership
	Football, nationalism, racism
	John Oliver
	Opinion polls
	Parties and leaders (Scotland)
	Uncertainty

Figure 4: Topic Evolution "Brexit: The Movie"



Note: y-axis indicates the proportion of topic

Figure 5 Topic Evolution "Financial Risk"

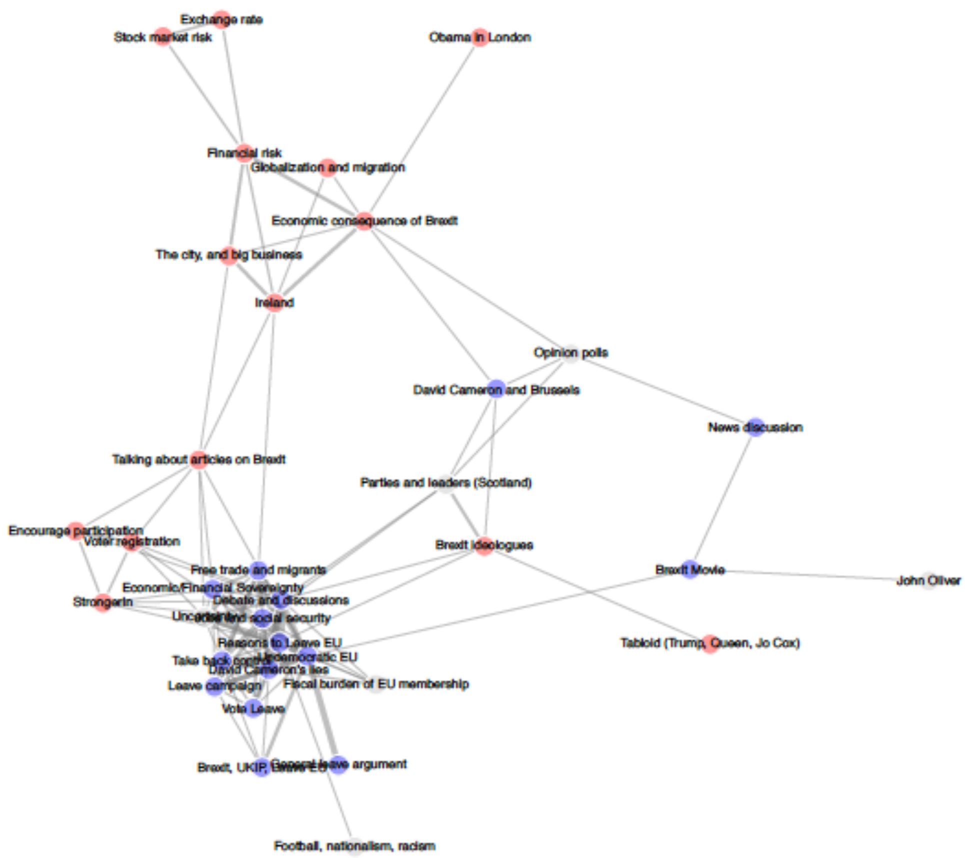


Note: y-axis indicates the proportion of topic

5. Structure of topics

STM allows the estimated topics, meaning that it can correctly identify topics that co-occur in the Tweets by a user. By checking the topic correlations, we can also map the network structure of topics. The following figure shows the network of topics (Figure 6).

Figure 6 Network of topics



Note: The red circles are Remain topics (judged from the estimated proportions of topics, described above), and blue circles are Remain topics.The widths of the lines that connect two topics indicate proximity between two topics measured by the strength of correlations. The distance of two topics in the plot also reflects the proximity.

The figure indicates that the Leave topics are strongly correlated. The dense network of a number of Leave topics on the bottom left in the figure indicates that the Leave Tweets had well organized topic structure centered around “Vote Leave”, the core topic by Leave side.

In contrast, the Remain topics are rather sporadic. There are correlated topics of mobilization attempts (middle-left), but these are not much related to topics of substantial issues. For remain side, topics on the substantial issues are mostly on economy (middle-top), but these are rather separated from the core topic of remain side (“Stronger In” topic on the bottom left). In terms of the dissemination of topics on the social media, the Leave side seemed to have the advantage against the Remain side. Because a user is more likely to Tweet both the simple message of taking action (e.g. “Vote Leave”) and the substantial issues that explain why such actions are necessary (e.g. “Jobs and social security”), followers of such account would have received the strong message for Leave.

6. Analysis of followership network

In this section, we estimate the positions of Twitter accounts through a network analysis of followership during the Brexit referendum campaign. The data we use are the network of followers for 400 accounts that are the most frequently mentioned in the Brexit referendum Tweets. We use a method of network-based ideology scaling developed in Barberà (2015). The methodology recovers an issue/ideology space from the network of followership, and locates followers and followers in a single dimensional space. The key assumption of this methodology is that Twitter users prefer to follow accounts that have opinions similar to their own opinions. In this specific application, we think that by focusing on the important accounts in Brexit debate selected by the frequency of mentions to the accounts in Tweets related to Brexit, we can reproduce the ideological space in Brexit (i.e. Leave or Remain).

The estimated results from this model consist of two parts.

1. Estimated positions of important Twitter accounts
2. Estimated positions of Twitter accounts who follow these important accounts.

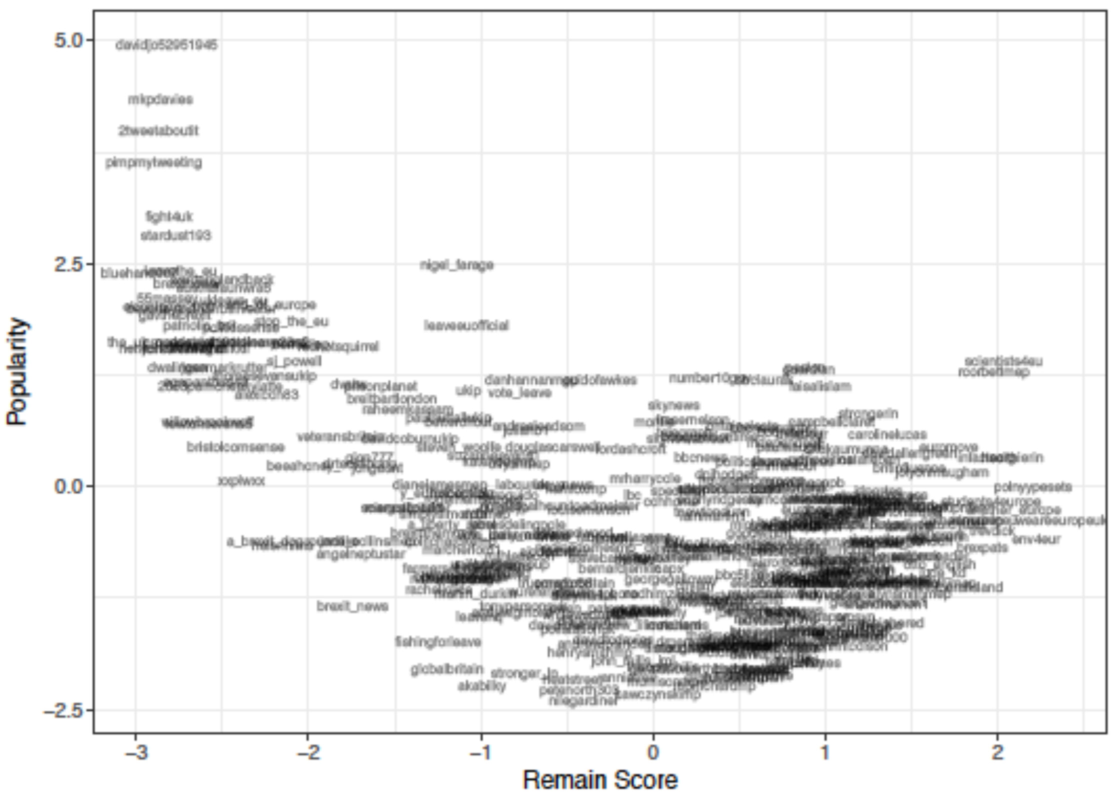
We will explore the results in this order.

7. Ideology distribution of followees

The following is the estimate of all frequently mentioned accounts (Figure 7). The horizontal dimension is the position (Right = Remain). The vertical dimension is the popularity (high = popular).

Then we highlight what we consider the politically influential accounts, namely Twitter accounts of MP and media outlets.

Figure 7. Network based scaling of Brexit issue positions frequently mentioned user accounts

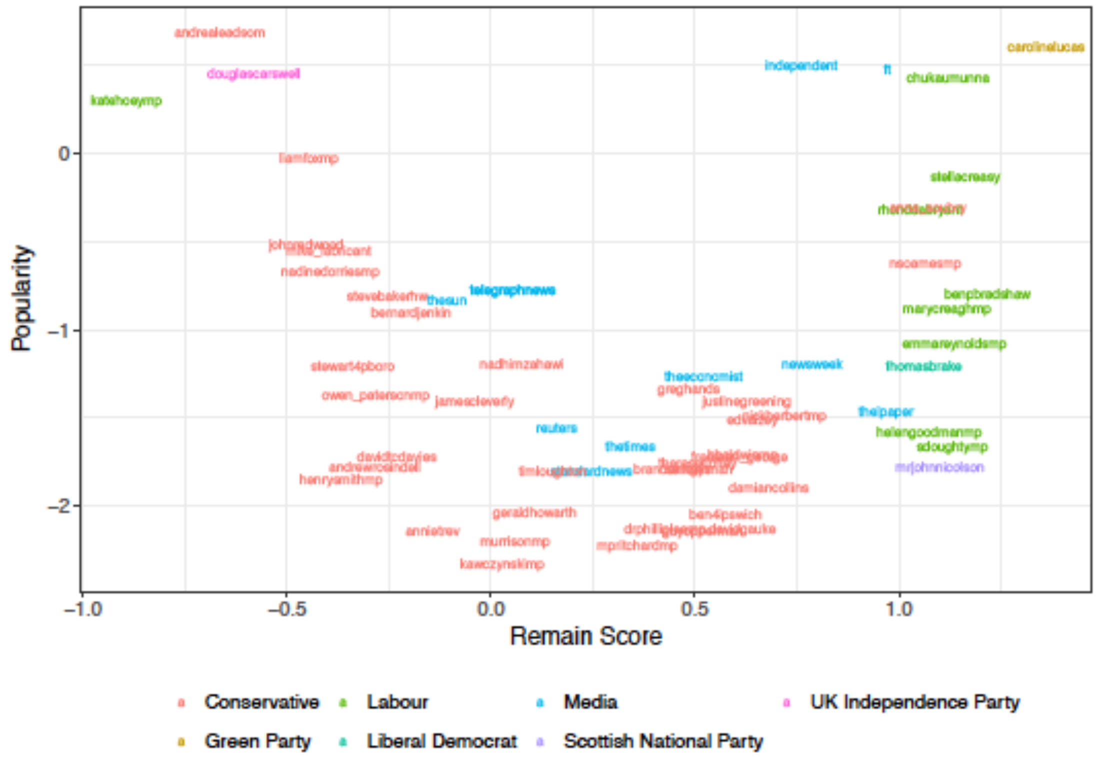


The results are depicted in Figure 8. The messages we can get from this picture are the following:

- Conservatives MPs positions are very diverse. Their positions are distributed from moderately Leave to moderately Remain.
- Labour MPs positions are relatively coherent and placed as the firm Remain. An exception was Kate Hoey, who was one of the most active figures in the Grassroots Out group.
- Other parties MP positions are in the location we expect (SNP, Greens, UKIP).

- Media account positions ranged from from moderate Leave (The Sun) to firm Remain (Financial Times).

Figure 8: Network based scaling of Brexit issue positions of politically important Twitter accounts



8. Distribution of followers

We now move on to the analysis of the distribution of followers, depicted in Figure 9. There are two peaks in the distribution of follower accounts. The plot of the distributions of Remain and Leave users (based on the Nave Bayes estimation) explains where these two peaks come from. The peak on the left corresponds to the Leave users, and peak on the right corresponds to the Remain users. This result indicates that there are distinct patterns in the choice of whom to follow by the users who participated in the Brexit debate on Twitter. Leave users followed Leave accounts more, and therefore they are more likely to be exposed to the information broadcasted by Leave accounts on their Twitter feed, then as a result mentioned these accounts frequently in their Tweets to disseminate the information further. A similar pattern is found among the users classified as pro-Remain.

Figure 9: Distribution of non-elite accounts based on the network scaling

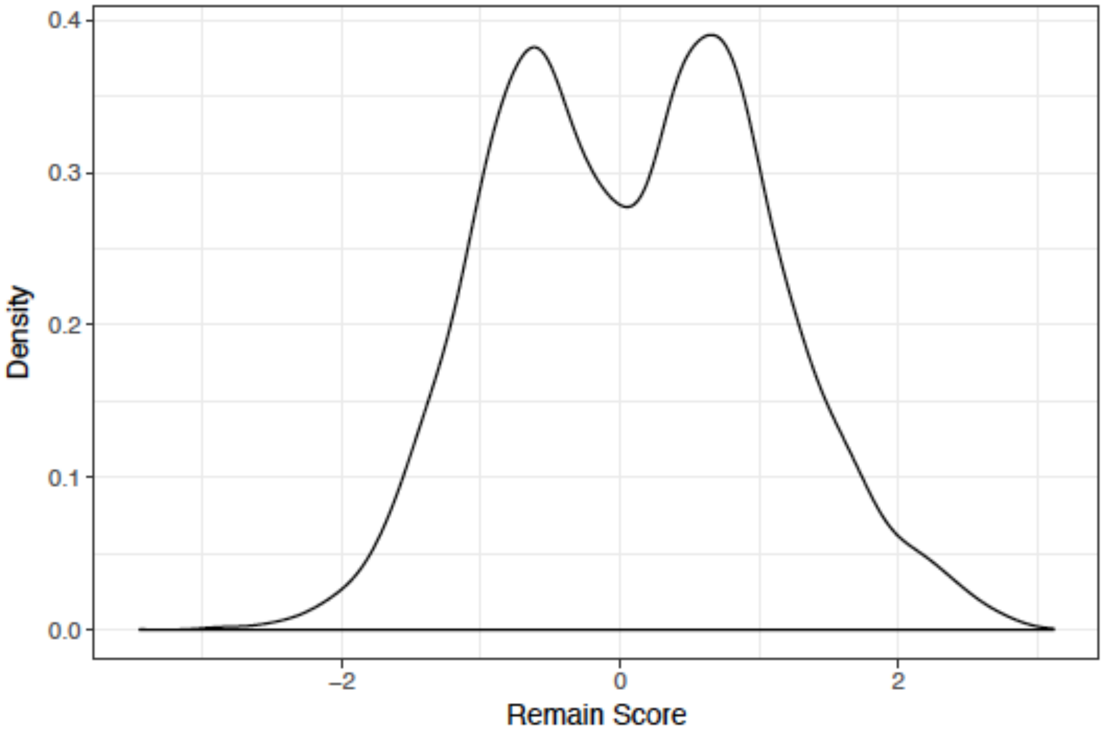
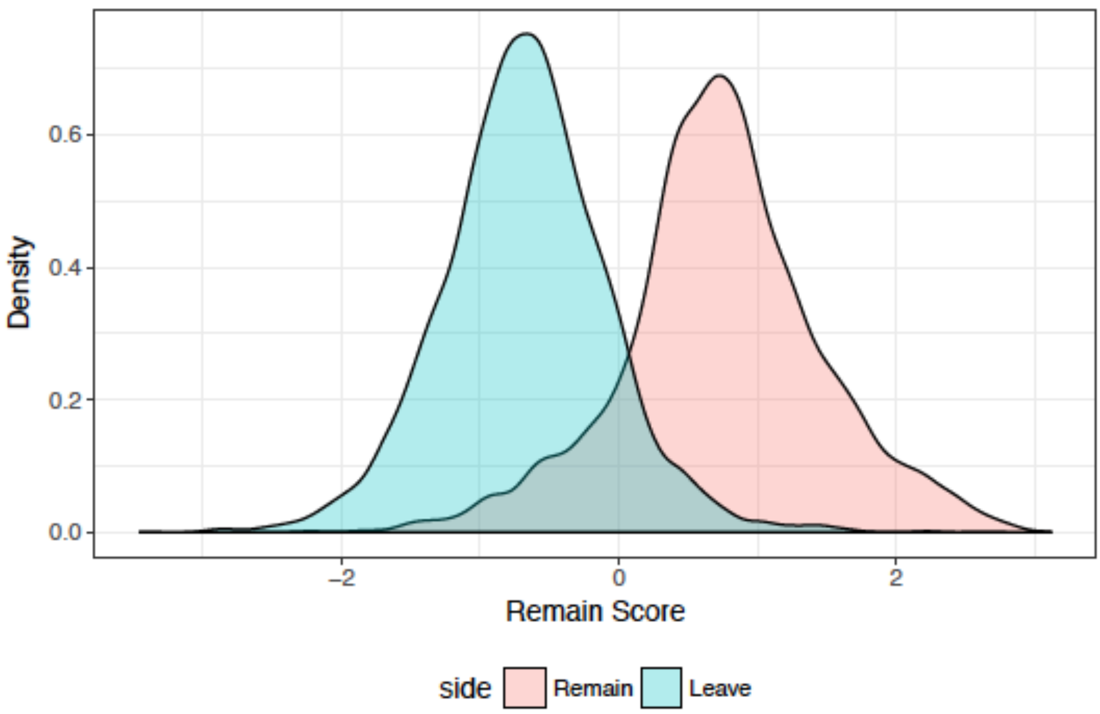


Figure 10: Distribution of non-elite accounts based on the network scaling by side



9. Comparison between followership network on Brexit and all MPs and media

In the previous section we found that there are distinct patterns of followership by Leave and Remain Twitter users. In this subsection, we attempt to address the origin of this pattern, in particular whether this pattern has any relations with the normal left-right ideological spectrum, through the comparison of two ideological scaling, one from the Leave and Remain positions in the Brexit Tweets network and another from the positions. Following the strategy of Barbera (2015), we obtained the followership network data of all MPs during the Brexit campaign period (i.e. before the 2017 General Election) and major media outlets, and estimated the positions of Twitter users in this network.

The next figure shows the differences and similarities between the ideological positions estimated from these two networks 11. The horizontal axis indicates the positioning of Twitter users in the network of Brexit Tweets (where larger values indicate being more pro-Remain) and the vertical axis indicates the positioning of Twitter users in the followership of MPs and media accounts.

The larger value indicates a more conservative position. There is a clear correlation between two estimates. The conservative Twitter users are more likely to take the Leave position. However, the relations are not strong. The correlation between these two variables are only 0.396. This result suggests that although the Leave-Remain dimension is partially the reflection of liberal conservative dimension, but other concerns, extracted as the topic of either Leave or Remain in our topic model but not well-represented by the existing parties, would have played the key role for some citizens in the UK to decide their side on the debate.

10. Discussion

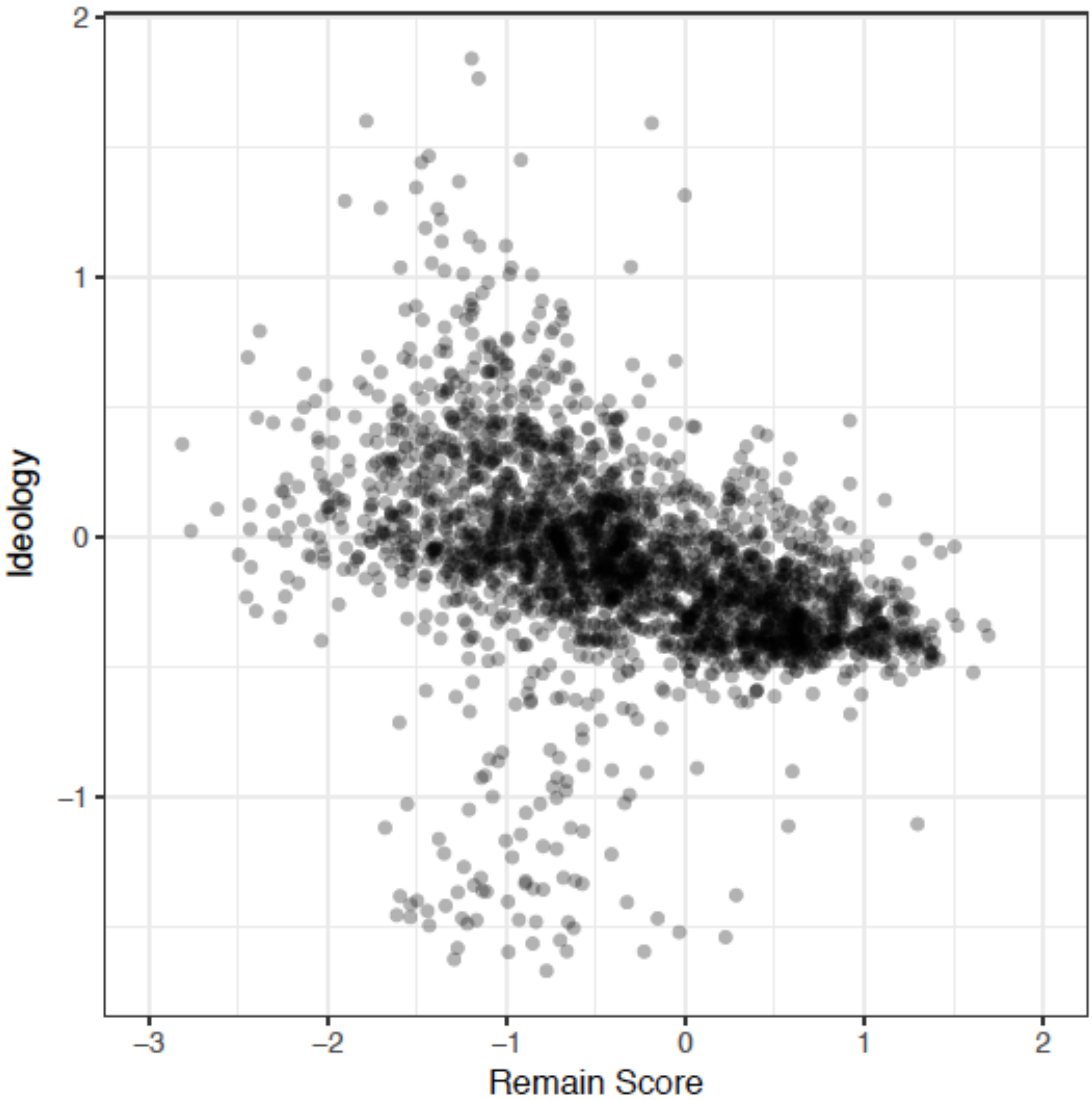
Our investigation of a large set of Twitter data yields important insights into the engagement of citizens, media and politicians using Twitter for political communication during the campaign period of Brexit referendum in the UK in June 2016. In particular, we have shown that both the content and the network of followerships reveal important differences among the sides in the debate, both about how individuals and organizations self-select their content and what messages they communicate, depending on whether they supported the Leave or Remain sides in the debate over Brexit.

Our analysis of the network of topics that users have discussed during the referendum campaign and the network of followership was aimed at detecting the differences between Leave and Remain. Our analysis shows that the structure of followership networks is both strongly divided and categorically different between the Leave and Remain sides.

Our analysis of the topics uses in the debate reveals that the structure of topics by Leave users are more densely related to one another those used that by Remain users. In other words, the messaging and discourse of the pro-Leave side was more dense and more coherent than the network of pro-Remain supporters, who advanced more diverse arguments in favor of their position. This mirrors earlier work on the sentiment of the messages that reveals more positive, less tentative, and more future-looking rhetoric among Leave supporters compared to Remain.

Overall, our analysis of political discussion on Twitter has revealed that this platform does contain informative content in the form of political communication that meaningfully distinguishes the sides in the Brexit debate. It is both possible to predict the orientation of a user from the user's Tweets, and possible to distinguish distinct topics and distinct positioning from the content and followership structures of the user's social media usage.

Figure 11: Comparison of Brexit ideology and normal Left-Right ideology based on network-based scaling



References

Allcott, H. and Gentzkow, M. (2017). Social Media and Fake News in the 2016 Election. *Journal of Economic Perspectives,* 31(2):211–236.

Barberá, P. (2014). Birds of the same feather tweet together: Bayesian ideal point estimation using Twitter data. *Political Analysis*, *23*(1), 76-91.

Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent dirichlet allocation. *Journal of machine Learning research*, *3*(Jan), 993-1022.

Chang, J., Gerrish, S., Wang, C., Boyd-Graber, J. L., & Blei, D. M. (2009). Reading tea leaves: How humans interpret topic models. In *Advances in neural information processing systems* (pp. 288-296).

Del Vicario, Michela, Alessandro Bessi, Fabiana Zollo, Fabio Petroni, Antonio Scala, Guido Caldarelli,

Del Vicario, M., Bessi, A., Zollo, F., Petroni, F., Scala, A., Caldarelli, G., ... & Quattrociocchi, W. (2016). The spreading of misinformation online. *Proceedings of the National Academy of Sciences*, *113*(3), 554-559.

Flaxman, S., Goel, S., & Rao, J. M. (2016). Filter bubbles, echo chambers, and online news consumption. *Public Opinion Quarterly*, *80*(S1), 298-320.

Hong, L., & Davison, B. D. (2010, July). Empirical study of topic modeling in twitter. In *Proceedings of the first workshop on social media analytics* (pp. 80-88). ACM.

Morstatter, F., Pfeffer, J., Liu, H., & Carley, K. M. (2013, July). Is the Sample Good Enough? Comparing Data from Twitter's Streaming API with Twitter's Firehose. In *ICWSM*.

Roberts, M. E., Brandon M. S. and Tingley, D. (2015). stm: R Package for Structural Topic Models. *Journal of Statistical Software* VV.

Roberts, M. E., Stewart, B. M., Tingley, D., Lucas, C., Leder-Luis, J., Gadarian, S. K., ... & Rand, D. G. (2014). Structural Topic Models for Open-Ended Survey Responses. *American Journal of Political Science*, *58*(4), 1064-1082.

Sunstein, C. (2001). *Echo Chambers*. Princeton University Press.

Wallach, H. M., Murray, I., Salakhutdinov, R., & Mimno, D. (2009, June). Evaluation methods for topic models. In *Proceedings of the 26th annual international conference on machine learning* (pp. 1105-1112). ACM.

Xu, Z., Ru, L., Xiang, L., & Yang, Q. (2011, August). Discovering user interest on twitter with a modified author-topic model. In *Proceedings of the 2011 IEEE/WIC/ACM International Conferences on Web Intelligence and Intelligent Agent Technology-Volume 01* (pp. 422-429). IEEE Computer Society.

Zhao, W. X., Jiang, J., Weng, J., He, J., Lim, E. P., Yan, H., & Li, X. (2011, April). Comparing twitter and traditional media using topic models. In *European Conference on Information Retrieval* (pp. 338-349). Springer, Berlin, Heidelberg.