

Appendix

This document is accompanying the submission *Linking User Opinion Dynamics and Online Discussions*. The information in this document complements the submission, and it is presented here for completeness reasons. It is not required for understanding the main paper, nor for reproducing the results.

A Posting analysis and dataset profiling

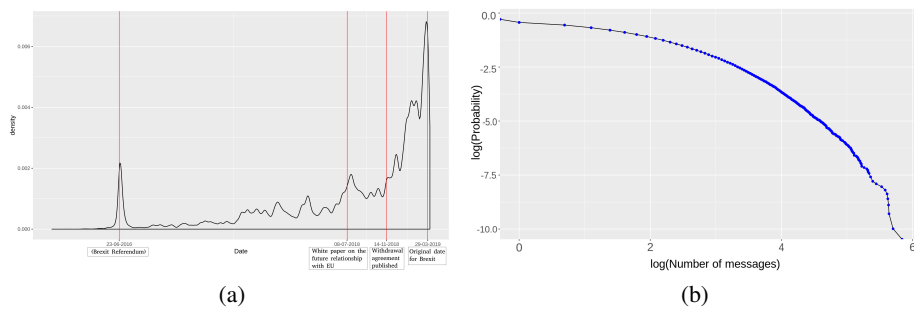


Fig. 1: **(a)** Time distribution of the submissions collected from Reddit (subreddit brexit), between November, 2015 and April, 2019. **(b)** Complementary Cumulative Distribution Function of the number of messages sent by each user.

As far as the statistical structure of the submitted messages is concerned, the very vast majority of them are comments as opposed to initial thread starting messages, representing respectively 91 % and 9 % of the entries. Moreover, 20% of all the unique authors are **only** thread initiators. This means they only send a single message, starting a discussion thread, in which they never post again. On the other hand, 19% of the authors, are both thread starters and commenters, meaning that they start threads and take part actively in the discussions, posting answers in their own started thread or getting involved in other discussions. The majority of the unique users (61 %) are **only** commenters, meaning that they never start discussions, but usually engage in them.

Figure 1b presents the Complementary Cumulative Distribution Function of the number of messages per user (in log log scale). It shows that a very large number of users send only a few messages, whereas there are a few users sending many messages in the observed interval. These users can either be opinion formers or simply paid social bots. One of the tricky tasks of this work was to identify the bots and remove them, as they do not bring new information, but most of the time reshare news and posts.

B Prediction for Neutral

CL: remove paragraph ci dessous ou a verifier -> Predicting the following stance

Table 1: Events and time periods used for splitting the Reddit dataset.

Int.	Start date	Posts	Users	Important events in the Brexit chronology
T1	2015-11-16	3367	1268	23 June 2016 The UK holds a referendum on whether to leave the European Union (EU). 51.9% of voters vote to leave. 24 June 2016 David Cameron announces his resignation as Prime Minister.
T2	2016-06-25	6265	1623	13 July 2016 Theresa May accepts the Queen’s invitation to form a government
T3	2016-07-14	3084	810	27 July 2016 The European Commission nominates French politician Michel Barnier as European Chief Negotiator for the United Kingdom Exiting the European Union. 7 December 2016 The UK House of Commons votes 461 to 89 in favour of May’s plan to trigger Article 50 by the end of March 2017
T4	2016-12-08	1466	320	24 January 2017 UK Supreme Court rules that Parl. must pass legislation to authorize the trigger of Art. 50. 26 January 2017 The UK Gov. introduces a 137-word bill in Parl. to empower May to initiate Brexit by triggering Art 50.
T5	2017-01-27	2300	431	16 March 2017 The bill receives Royal Assent. 29 March 2017 A letter from May is handed to President of the European Council Donald Tusk to invoke Article 50, starting a two year process with the UK due to leave the EU on 29 March 2019.
T6	2017-03-30	4102	557	19 June 2017 Brexit negotiations commence.
T7	2017-06-20	54505	2339	6 July 2018 A UK White paper on The future relationship between the United Kingdom and the European Union is finalized. 8 July 2018 Davis resigns as Secretary of State for Exiting the EU. Dominic Raab is appointed as his successor the following day.
T8	2018-07-09	23067	1479	21 September 2018 EU rejects the UK white paper.
T9	2018-09-22	15385	1480	14 November 2018 Brexit withdrawal agreement published. 15 November 2018 Raab resigns as Secretary of State for Exiting the EU. Stephen Barclay is appointed as his successor the following day.
T10	2018-11-16	3718	732	25 November 2018 Other 27 EU Member States endorse the Withdrawal Agreement.
T11	2018-11-26	25468	2485	15 January 2019 First meaningful vote held on the Withdrawal Agreement in the UK House of Commons. The UK Gov. is defeated by 432 votes to 202
T12	2019-01-16	54850	4489	12 March 2019 Second meaningful vote on the Withdrawal Agreement with the UK Government defeated again by 391 votes to 242. 14 March 2019 UK Gov. motion passes 412 to 202 to extend the Article 50 period.
T13	2019-03-15	9119	1836	20 March 2019 May requests the EU extend the Article 50 period until 30 June 2019. 21 March 2019 The European Council offers to extend the Article 50 period until 22 May 2019 if the Withdrawal Agreement is passed by 29 March 2019 but, if it does not, then the UK has until 12 April 2019 to indicate a way forward. The extension is formally agreed the following day.
T14	2019-03-22	13414	2444	29 March 2019 The original end of the Article 50 period and the original planned date for Brexit. Third vote on the Withdrawal Agreement after being separated from the Political Declaration. UK Government defeated again by 344 votes to 286.
T15	2019-03-30	9509	1840	5 April 2019 May requests for a second time that the EU extend the Article 50 period until 30 June 2019.
	2019-04-05			–dataset end–

Furthermore, in order to explain these results, we compute volume of users corresponding to each of the 9 transitions. The results are shown in Table 2. This table explains the values from ?? . Firstly, the number of users having a Neutral - Neutral trajectory is considerably small, because of the applied filtering, namely the removal of entries having translation Neutral - Neutral which appear only once, for a given set of input features. Next, we obtain very low scores for the transitions from a pro Brexit stance to an against Brexit stance and vice-versa because the number of examples of users in these situations is very small compared to the other categories, 35 and 33 respectively.

Table 2: The volume of users in the testing set for each category of transitions.

Current Stance	Following Stance	Number of Users
Against	Against	158
Against	Brexit	33
Against	Neutral	371
Brexit	Against	35
Brexit	Brexit	60
Brexit	Neutral	332
Neutral	Against	387
Neutral	Brexit	350
Neutral	Neutral	27

Indeed, the chances that a user will change his position in two consecutive time-frames from totally Brexit supporting to totally against Brexit are low and this is revealed by the distribution presented in Table 2. In most situations, there will be a transition through the intermediate Neutral state, also shown by Table 2, the number of users going from Against and Brexit to Neutral being 371 and respectively 350. This allows our classifier to understand the underlying structure of the dataset, leading to high F1-scores for these translations.

The XGBoost classifier obtains an overall F1-score of 0.57, using the FS4 feature set. We can identify two main sub-components of this score: the first one is represented by the low scores obtained when predicting transitions Brexit - Against or Brexit - Brexit, which lowers the overall score. The second sub-component is defined by the predictions involving the Neutral state, namely from Neutral to the other three states.

Predicting the following stance a user will arrive into turns out to be a difficult task, notably when the user changes of opinion. In Fig. 2 we show the volume of users translating from neutral state to a Brexit supporting state or Against Brexit state, in consecutive time-frames. The left figure shows the percentage of users, while the right figure shows the exact number of users going to the two polarized states. We can observe that except the starting time-frame, where there was a strong campaign for Brexit which is reflected in the ratio of users who migrated to a Brexit stance in the second time-frame, usually the trend is in favor of the Against side. In translations from

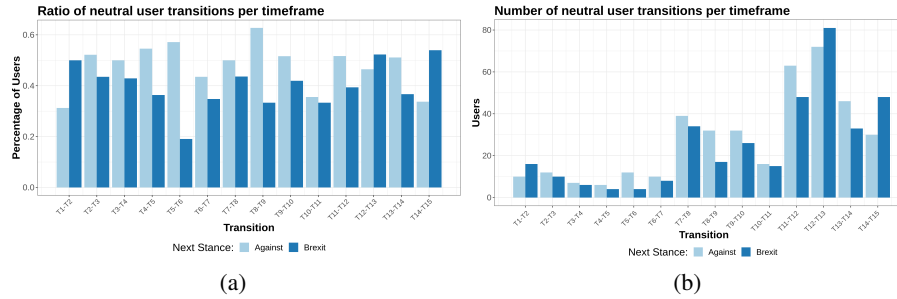


Fig. 2: **MAR: PDF (vectorial) version needed!** a) Ratio of users transitioning to the other stances in consecutive time-frames, ratio. b) Number of users transitioning to the other stances in consecutive time-frames.

T2 to T12, neutral users tended to move to Against Brexit. However, in T12, there is a change in the trend.

We performed some analysis on this period and found out that the main event in this period of time was the second negative vote in the Parliament for the Withdrawal Agreement negotiated by Theresa May. UK would have had to pay the European Union 39 billion pounds, which was upsetting and disappointing for people, as revealed by some of the utterances we checked (we checked utterances with high leave probability): *"It's taken far too much time, we should leave hard and deal with the consequences. It will be tough for a time but there's no price not worth paying for freedom from the globalist overlords..."* *"It's **better** to **die free**, **than live** as a slave."* *F. Douglass* or *"Democracy? We roam the world dismantling dictatorships to install democracy, leaving failed states, but we cannot deliver the democratic will of our own population!"*. We assume people tended to be disappointed by the incapability of the politicians to deliver the Brexit, thus the raise in Neutral people from time-frame 12 transitioning to Brexit in time-frame 13.

We tested our best predictor, the XGBoost using FS4, the feature set comprising all the other sets, on users on this period of time. Thus, we considered time-frame 12 and tried to predict how many users will have a trajectory towards the Brexit side and how many will migrate towards the Against Brexit side. The results are shown in Table 3.

Table 3: Number of neutral users in time-frame 12 predicted to migrate to each of the other stances in time-frame 13. **MAR: Bring observed volumes here, as a new line.**

Against	Brexit	Neutral
71	83	4

Even though the task of predicting the future stance of groups of users is difficult, we managed to correctly predict towards which group will neutral users migrate in a

consequent period of time. This shows that our predictor correctly captured the trend and the underlying dynamics of the online population, despite the sudden change of tendency, after a long series of similar inter-period transitions. Such kind of results can be very interesting for polling agencies trying to discover before-hand the outcome of important events debated online. What is more surprising, these results are obtained without the involvement of textual information at all in the feature sets, but just with taking advantage of the interaction between people and the diffusion of information.