Information diffusion in online communities

Andrei Mardale, Christine Largeron & Marian-Andrei Rizoiu
Universitè Jean Monnet & University of Technology Sydney

1 The problem

- Do discussed topics change in time in online social networks? **Yes, however some subjects remain actual for more than one period.**
- Can we visualize the dynamics of discussion topics and the trajectory of users in terms of addressed topics? **Yes!**
- Are the future political opinions of online users influenced by online discussions they are exposed to? **We believe so, especially when the users do not have a strong preference to begin with.**

2 Case study - Brexit in Reddit

Reddit dataset – ideal for longitudinal study, having structured content in subreddits and tree like threads.

- 207,894 comments
- 21,725 initial threads
- 14,362 users
- Time extent: Nov. 2015 - Apr. 2019.

Elements of Reddit platform

Tree structure

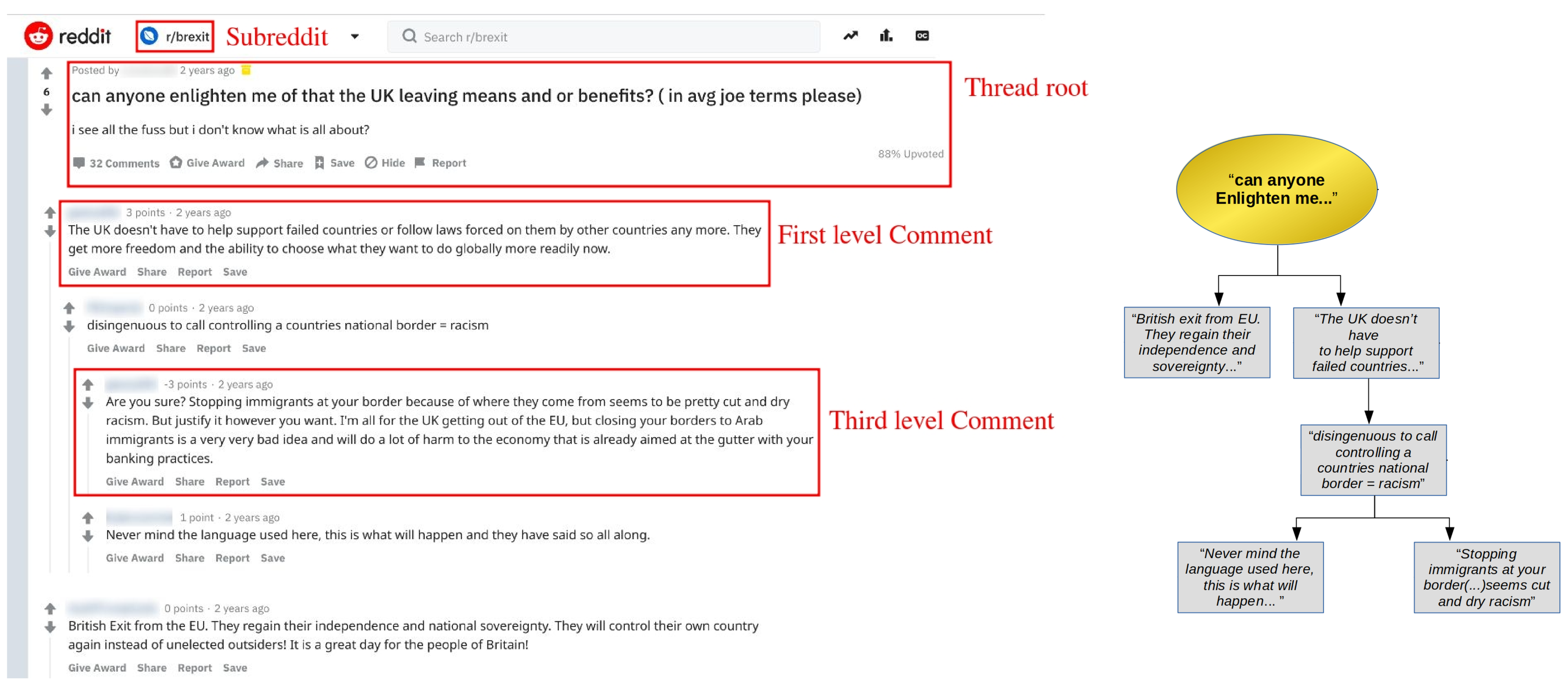


Figure 1: Left: example of a discussion thread, with multi-level comments, inside a subreddit. **Right:** example of the logical structure of a thread used for analyzing data and predicting future political stance.

Analysing the structure of the dataset

The interest for the subject of Brexit has been growing considerably and has been constantly echoed the political decisions made in this case.

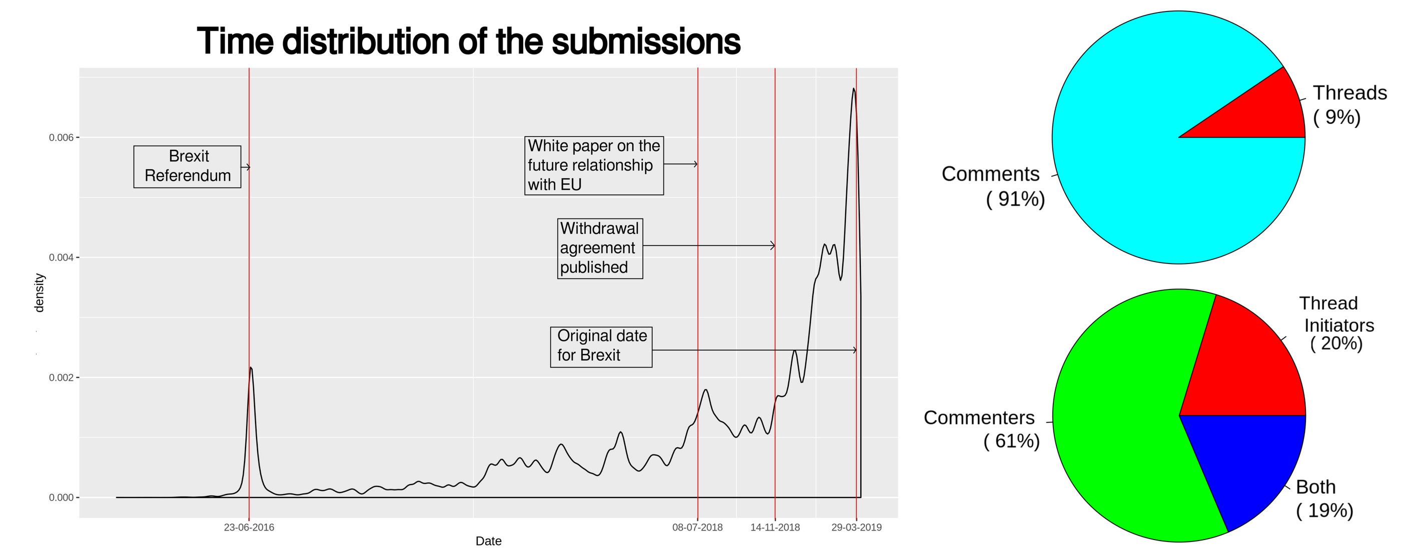


Figure 2: Left: Temporal distribution of the posts. The ascending trend is only perturbed by the initial peak, on the 23 June 2016, when Brexit Referendum was held. **Right:** Distribution of submitted messages and distribution of users in terms of roles. People tend to comment more on already commenced threads.

3 Longitudinal analysis of topic evolution

The posts were split temporally in 15 time-periods. For each time-period, each user's messages were aggregated and using LDA, we determined the probability distribution of the 10 most present topics in their discourse.

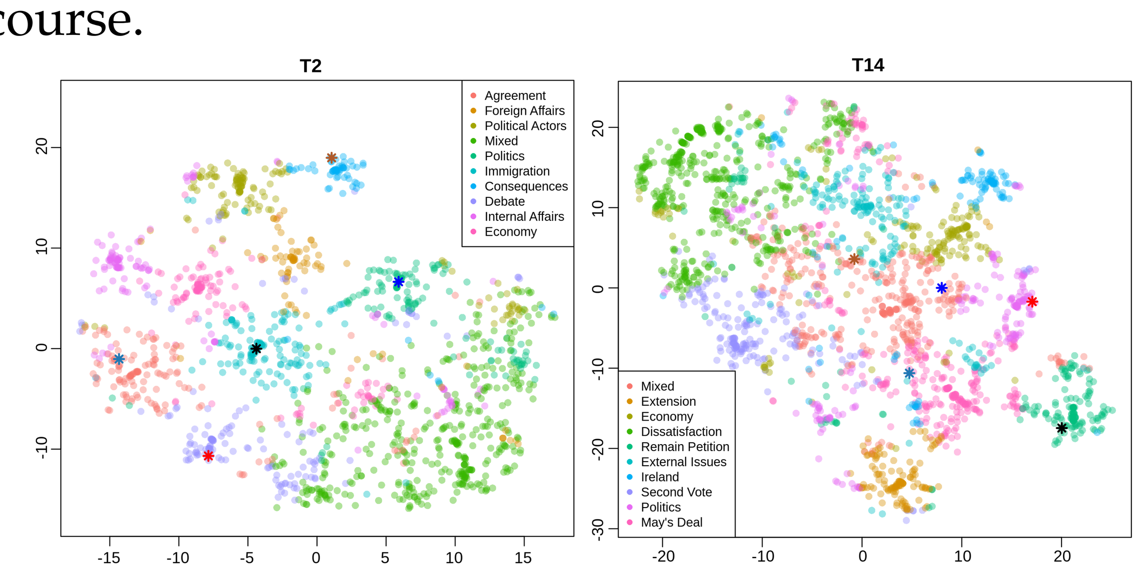


Figure 3: K-Medoids clustering and t-SNE representation of the vectorial embedding of the users in the topic space. Between the incipient and the ending periods (*eg.* T2 and T14 in figure), topics shift from the agreement on Brexit, immigration, the consequences to remaining petitions, external affairs, Ireland or second vote.

[1] Benoit, K. and Matsuo, A. (2018). Network analysis of Brexit discussion on social media.

4 Political stance predictor

Aim: predict users' political stance (**A**gainst, **N**eutral, **B**rexit) changes in consecutive time-frames by analysing the interactions with other users and the structure of the conversations.

- A classifier trained on Twitter [1] is used to label users' discourse as pro, against or neutral to Brexit.
- Current stance and features extracted from the structure of the diffusions are used to predict the stance at the next time-period.

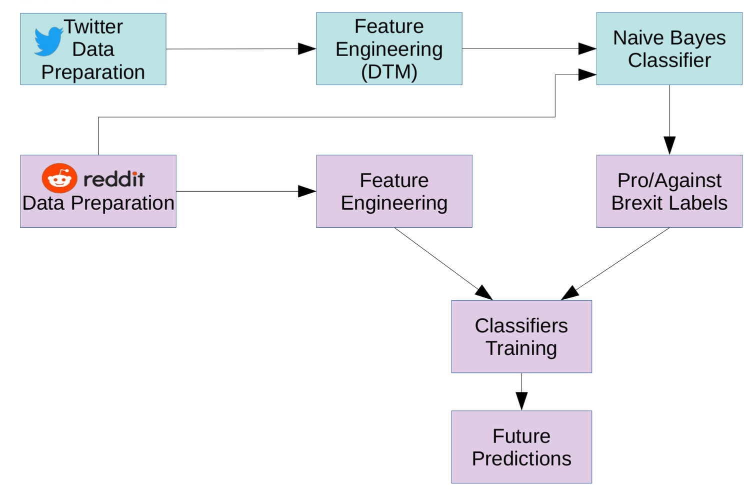


Figure 4: Proposed strategy for predicting political stance changes.

Features based on the structure of the diffusions

FS1. User activity

- # initiated threads;
- # submitted comments;
- # received replies per comment;
- stance at current time-frame;

FS2. User activity per group

- # initiated threads;
- # submitted comments;
- # received replies per comment from each group (**A**, **B**, **N**);
- # submitted comments to users from each group (**A**, **B**, **N**)
- stance at current time-frame;

FS3. Structure of diffusion

- % comments from each group (**A**, **B**, **N**), in the diffusions the user takes part in;
- stance at current time-frame;

FS4. All

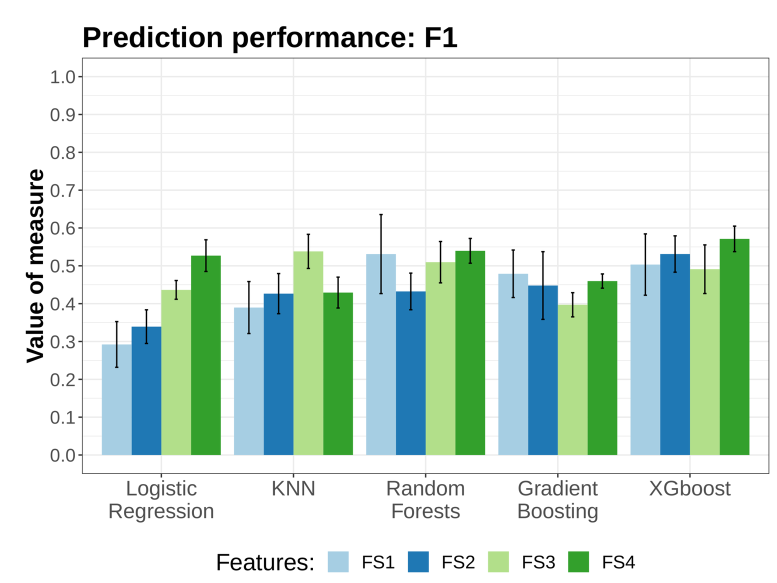
- FS1 + FS2 + FS3

Classifiers

Logistic regression, k-NN, Random Forest, Gradient Boosting, XGBoost, hyperparameters tuning and evaluation using 10-Cross Validation, reporting mean and standard deviation of F1 Score.

5 Results

- Good F1 scores obtained by most predictors (3-class prediction problem);
- The best predictor is *XGBoost* trained with *FS4* (F1 = 0.57, sdev = 0.03);
- Promising results when predicting future stance for Neutral users.
- The quality of the predictions is proportional with the number of users performing these transitions.



Following / Currently	Against	Brexit	Neutral
Against	0.26	0	0.98
Brexit	0	0	1
Neutral	0.73	0.63	0.45

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

Figure 5: Top: Overall prediction evaluation (F1 Score). XGBoost trained with FS4 performs best in terms of F1 score, especially on transitions from and to Neutral state, due to the larger training set with these examples. **(bottom row, left and right)**.

Predicting *Neutral* users behaviour

We managed to predict correctly the sudden alternation from the last 3 transitions. This allows us to anticipate the sizes of the new groups in consequent periods, a useful result especially for polling agencies.

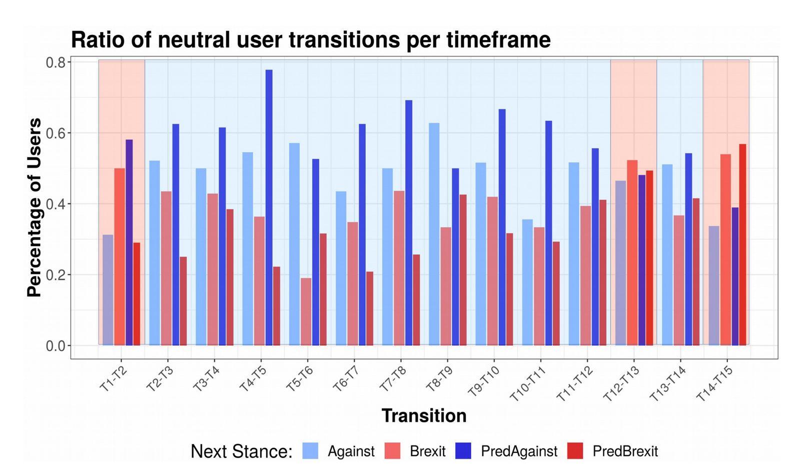


Figure 6: Neutral users transition ratios: in transition 1-2 more users changed to Against, in the next 10 transitions users migrated more to Brexit. A sudden change appears in transition 12 - 13.