Introduction

The arising popularity of social media has changed the political communication process nowadays, gradually replacing the role of conventional media as a primary source of information transmission channel.

2.Data

We construct a panel data including social media activities and hate crime information in the United States. The data consists of the following sources: 1) Facebook data including posts, user likes, user comments and page information on US politician pages and other most popular 1000 pages which had mentioned Trump and Clinton during the presidential campaign. 2) Hate crime reporting data from FBI. Our panel dataset covers 50 states in the US and range from 2015/05~2016/11, primarily focusing on the period of US presidential campaign. The summary statistics of primary variables are shown in table .

2.1 Facebook Data

To measure the salience of immigration and race-related information on Facebook, the most popular social media in the United States, we collect posts and reaction data from all fan pages of current national politicians, including members and

candidates of the Senate, the House, and the past and present Governors. Our data also includes the most popular 1000 pages that have ever mentioned the two major presidential candidates: Donald J. Trump and Hillary Clinton, in August 2016, ranked by the total number of likes, comments, and shares of candidate-related posts in these pages.

We also narrow down our user sample to US potential users by considering those having reacted on a national politician's fan page. The basic assumption behind this criterion is that users from other countries have a significantly less interest toward US politician. In addition, this criterion also limited our sample to those that are highly engaged in political-related issues.

We first identify posts that are related to immigration and race by supervised text classification model. To begin with, we select a number of keywords (table 1) to filter out posts that are possibly related to immigration and race. This step is to guarantee our pool of posts contains a sufficient proportion of related posts, therefore our training data sampled from this pool will less likely to encounter imbalanced problem.

Moreover, we randomly sample 500 posts each month for both issues to be labeled from the testing data. The sampling process guarantee our training data to cover different topics popularly discussed in each month. Afterwards, we labeled each posts to identify whether it is immigration/race related by 7 trained undergraduate assistants. Table 2 shows some examples of posts that are labeled as immigration/race related.

We transform these labeled posts into tf-idf matrices to offsets the counts of the words by the number of articles a word appears in. Later, we trained several text classification model including logistic regression and support vector machine. In addition, we also tried transfer learning of BERT and convolutional neural network in case to avoid the shortcomings of statistical learning model in understanding the semantic structure within text data. The performance of each model based on cross-validation process is shown in table 3. Based on this result, we select logistic regression as our final model. Figure 1 shows the trend of immigration and race-related posts amount on Facebook from 2015/05~2016/11.

We measure the salience of immigration and race-related information on Facebook by the amount of reactions toward these posts. Arguably, the amount of reactions not only reflect how many Facebook users are concerned about these issues, but also positively correlate with the range of propagation of these posts. To be more specific, the algorithm on Facebook allow users to see the posts their friends have reacted to recently, implying that the more reaction a post gains, the larger propagating effect it might have. In order to exploit the cross-sectional variation among different states, we