Determining the Presence of Political Parties in Social Circles

abstract

Van Gysel, C.; Goethals, B.; de Rijke, M.

Citation for published version (APA):
Determining the Presence of Political Parties in Social Circles

(abstract)

Christophe Van Gysel
University of Amsterdam
cvangysel@uva.nl

Bart Goethals
University of Antwerp
bart.goethals@uantwerp.be

Maarten de Rijke
University of Amsterdam
derijke@uva.nl

ABSTRACT

We derive the political climate of the social circles of Twitter users using a weakly-supervised approach. By applying random walks over a sample of Twitter’s social graph we infer a distribution indicating the presence of eight Flemish political parties in users’ social circles in the months before the 2014 elections. The graph structure is induced through a combination of connection and retweet features and combines information of over a million tweets and 14 million follower connections. We solely exploit the social graph structure and do not rely on tweet content. For validation we compare the affiliation of politically active Twitter users with the most-influential party in their network. On a validation set of 700 politically active individuals we achieve $F_1$ scores of 0.85 and greater.

INTRODUCTION

Blogs and social networks play an important role in the diffusion of political ideas [1]. We investigate the spread of influence of political parties in Twitter. We look at the contributions of eight political parties in the months before the 2014 elections in Flanders. We postulate that political influences travel through social graphs similarly to how ideas spread viva voce. Many social networks exhibit homophilic properties [4], implying that personal networks are grounded in sociodemographic, behavioral and intrapersonal characteristics [5]. Based on these properties we extract link-based features from interaction data on Twitter and simulate a random walk over an induced graph structure.

METHODOLOGY

Classification of nodes is performed through an iterative algorithm [2, 3] equivalent to performing random walks over an absorbing Markov chain. Their formulation is similar to the iterative formulation of the PageRank algorithm [6].

EXPERIMENTS

We consider a node classification problem over eight political parties active in the Flemish region of Belgium. While some of these parties share common views, they are relatively spread out over the political spectrum. These parties collectively published lists of 780 Twitter users with whom they associate themselves. We consider these lists as ground truth data and use these to measure classification performance and to determine the interaction weights. We targeted 12,254 Twitter users who followed at least two of these eight parties and consequently retrieved their information as described in the previous section. We gathered 1,249,091 tweets (of which 273,213 were retweets) and 14,849,213 follower connections.

REFERENCES