Determining the Presence of Political Parties in Social Circles

abstract

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

Publication date
2015

Document Version
Final published version

DIR 2015

Amsterdam

Proceedings

Proceedings of the 14th Dutch-Belgian Information Retrieval Workshop

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. These connections and tweets referred to at least 10 million users in total, which corresponds to the total number of nodes $|V|$.
We then built a graph structure from the gathered data. More precisely we introduce a directed edge from user i to user j if user i follows user j or when user i retweets a message from user j. In an initial experiment we assign both these interactions a unit weight. Later we also considered weights for the inversed edges, such that if user i follows user j a directed edge is added from user j to user i and similarly for retweets. We also considered a weight that is added when both user i and user j follow each other, which indicates reciprocal following. These additional interactions are interesting as they give insight in how actions of other users influences one's position in the social graph.
We asked Twitter users to evaluate their individual classification a week before the Flemish elections of 2014. Users were shown a distribution over the eight political parties and could voluntarily provide feedback between 0 and 100, where a higher score indicates a better classification. Some users expressed concern as they observed a non-zero probability for right- or left-winged extremist parties. We believe that some feedback scores were purposely negative so as to deny association with these parties, even though these associations were negligible in a statistical setting. We received feedback from 2 258 users. The distribution is characterized by its population mean (51.57), standard deviation (35.97) and median (62.0) [7].

*The original four-page poster was published at the 9th International Conference on Web and Social Media (ICWSM 2015) in May 2015.

REFERENCES
[1] L. A. Adamic and N. Glance. The political blogosphere and the 2004 U.S. election: Divided they blog. In *LinkKDD '05*, pages 36–43. ACM, 2005.
[2] S. Baluja, R. Seth, D. Sivakumar, Y. Jing, J. Yagnik, S. Kumar, D. Ravichandran, and M. Aly. Video suggestion and discovery for Youtube: Taking random walks through the view graph. In *WWW 2008*, pages 895–904, 2008.
[3] S. Bhagat, G. Cormode, and S. Muthukrishnan. Node classification in social networks. *CoRR*, abs/1101.3291, 2011.
[4] M. Conover, B. Gonçalves, J. Ratkiewicz, A. Flammini, and F. Menczer. Predicting the political alignment of Twitter users. In *SocialCom 2011*, pages 192–199, 2011.
[5] M. McPherson, L. Smith-Lovin, and J. M. Cook. Birds of a feather: Homophily in social networks. *Annual Review of Sociology*, 27(1):415–444, 2001.
[6] L. Page, S. Brin, R. Motwani, and T. Winograd. The PageRank citation ranking: Bringing order to the web. Technical Report 1999-66, Stanford InfoLab, November 1999.
[7] C. Van Gysel, B. Goethals, and M. de Rijke. Determining the presence of political parties in social circles. In *ICWSM*. AAAI, 2015.