

Finding Interesting Persons in the Dutch Political Twitter Landscape

Erik Tjong Kim Sang
Netherlands eScience Center
e.tjongkimsang@esciencecenter.nl

April 13, 2017

1 Introduction

The project *Automated Analysis of Online Behaviour on Social Media* of the University of Groningen studies the discursive practices of politicians and journalists on Twitter. While it is easy to obtain a list of active politicians, for example by selecting the members of the government, the parliament or politicians on election ballots, it is more difficult to find relevant journalists. In this paper we examine a method for finding journalists and other people of interest by checking who interacts with Dutch politicians on Twitter. We collect the asymmetric friend connections on Twitter (also known as follower lists) of Dutch politicians and use these to find other persons of interest¹.

In the next section, we discuss some related work. Next we describe our data, our approach and the results. In the final section we conclude.

2 Related work

Our goal is to extract a network of interesting people people on a social media network site based on a seed set and the connections of the people in the seed set to other users. The identification of groups of people based on their interactions is a familiar problem in social sciences. The specific task of automatically inferring a social network from a seed list, is called *community search*, a topic which has been studied since 2010 [4]. This task has been studied for various datasets, among which tweets [3].

Brems et al. [1] collected accounts of 40 journalists from Twitter. For this purpose they asked their own users who the most important Dutch-writing journalists are and selected from the candidates a list of 40 journalists balanced with respect to country (The Netherlands vs Belgium), gender (male vs female) and professional status (employed vs freelance).

¹The software associated with this study is available at <https://github.com/online-behaviour/find-journalists>

Source	Count	On Twitter	Percentage
Cabinet 2012-2017	26	14	54%
Parliament 2016	150	139	93%
Senate 2017	75	46	61%
Ballot 2017	28	24	86%
Parliament 2017	150	143	95%
duplicates	-102	-96	
total	326	270	83%

Table 1: Our seed set of Dutch politicians contained 326 persons from five sources: Dutch cabinet 2012-2017, Dutch parliaments of 2016 and 2017, Dutch Senate of 2017 and the ballot of the Dutch elections of 15 March 2017. We found 270 Twitter accounts associated with this group, which corresponds with a coverage of 83%.

3 Data , methods and results

We are looking for interesting Twitter users related to the topic of Dutch politics. As a starting set for this collection, we use a list of Dutch politicians. The list contains five parts: members of the Dutch cabinet in 2012-2017 (26 people), members of the Dutch parliament in 2016 (150 people at 15 March 2017), members of the Senate (75 people at 27 March 2017), party leaders on the parliament election ballot of 15 March 2017 (28 people) and members of the new 2017 parliament (150 people at 27 March 2017, see Table 1). About 83% of these individual have an account on Twitter. In total, we found 270 accounts for 326 people.

We collected the accounts followed by these 270 political accounts². This resulted is a list of 87,700 different accounts. Next, we looked up the followers of these accounts. We did not perform a full search here: we only looked up the accounts that were followed by two or more of the politicians (5,026 accounts).

We define interesting Twitter users as those that follow at least 10% of our group of 270 politicians while at least 10% of these politicians follows them (an approach similar to that of Hanson [2]). We found 377 of such users. We inspected this group and found that most of them (134) were (political) journalists (including talk show hosts and bloggers). We also found politicians (representing counties, provinces and Europe), former politicians and various types of organizations (like news organizations, political parties and NGOs). The remaining group of users has been left unclassified and included among others PR people, script writers and people of unknown professions (see Table 2).

Our group of 134 journalists is considerably larger than the 20 (Dutch) journalists used in the study of Brems et al. [1]. However, that study also restricted the selection of the journalists with respect to gender, occupation

²The lookup for followers was performed on 29 and 30 March 2017. Since then the distribution of the follower relations may have changed.

Group	Count	Notes
Journalists	134	including bloggers and talk show hosts
Organizations	70	related to news, politics and NGOs
Former politicians	45	left the cabinet, parliament or Senate
Politicians	43	county-level and EU politicians
Other	85	script writers, PR people, unknowns
Total	377	

Table 2: Occupations associated with the 377 accounts on Dutch Twitter that follow at least 10% of our seed set of politicians and are followed by at least 10% of them. Most of these are (political) journalists.

type and press media type. We have no information on the latter two categories but we found that gender is not evenly distributed in our journalist set (74% men).

More journalists can be found by searching beyond the threshold of 10% of followed politicians. However, the looser this restriction is made, the fewer relevant users will be found. At Figure 1 shows, approximately 70 % of the discovered users are journalists around a threshold value of 40% but that number quickly goes down when the threshold value is decreased. At our selected threshold value of 10% less than 20% of the new users prove to be journalists.

Classifying users according to their profession, is a labor-intensive manual process. We created several data visualizations to see if we could divide users in different occupation classes based on the follower data. With principal component analysis we obtained the best results when using the connections with the 33% least-connected users were used as classification features. In the resulting map the politicians were positioned at different positions than the journalists (Figure 2). However even on this best map there was no clear difference between the journalists and the users that were not politicians. If automatic identification of journalists is possible, it is a difficult task.

While the follower information did not distinguish the journalists from the other nonpoliticians, it does contain some interesting information. On the bottom right of the map we found a separate cluster of politicians with among them a few journalists (Figure 3). All politicians of that cluster turned out to be members of the same party, the right-wing PVV. It is a surprise that clustering program found a party-specific cluster without any party information input or tweet text input. The presence of some journalists close to this group of politicians also surprised us.

4 Concluding remarks

We have presented a method for finding interesting users of the social media network Twitter related to Dutch politics. The method started with a seed set of 270 accounts of Dutch politicians and found 377 other interesting users

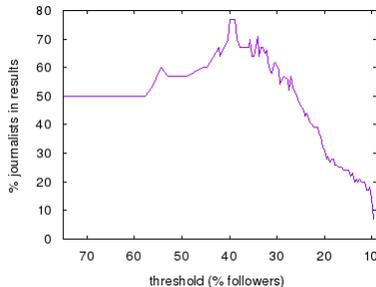


Figure 1: The relation between the requirement that new users should follow at least some percentage of the politicians (threshold) and the percentage of journalists in the suggestions. At a threshold value of 40% 70% of the suggestions are journalists but at the threshold value of 10% this number has dropped below 20%.

based on the accounts that followed and are followed by the politicians. We are primarily interested in finding accounts of (political) journalists on Twitter and of the 377 new accounts 134 (36%) belonged to this group. We looked at two methods for expanding this group from the data we had collected, lower selection criteria and data visualization, but found that further expansion of the group is difficult. In one of our data visualizations we found an interesting division based on party politics without putting any of this information in the data.

In this study, we explicitly avoided to use tweet text. For future work, this is an obvious next step. Using tweet text and information about replies and retweets may produce fewer links but the results will be more up-to-date than that of follower lists which may be outdated.

References

- [1] Cara Brems, Martina Temmerman, Todd Graham, and Marcel Broersma. Personal branding on twitter: How employed and freelance journalists stage themselves on social media. *Digital Journalism*, 2016.
- [2] Carl Lee Hanson, Ben Cannon, Scott Burton, and Christophe Giraud-Carrier. An exploration of social circles and prescription drug abuse through twitter. *Journal of Medical Internet Research*, 15(9), 2013.
- [3] Rong-Hua Li, Lu Qin, Jeffrey Xu Yu, and Rui Mao. Influential community search in large networks. In *Proceedings of the VLDB Endowment*, volume 8, pages 509–520, 2015.
- [4] Mauro Sozio and Aristides Gionis. The community-search problem and how to plan a successful cocktail party. In *KDD'10*. ACM, Washington DC, 2010.

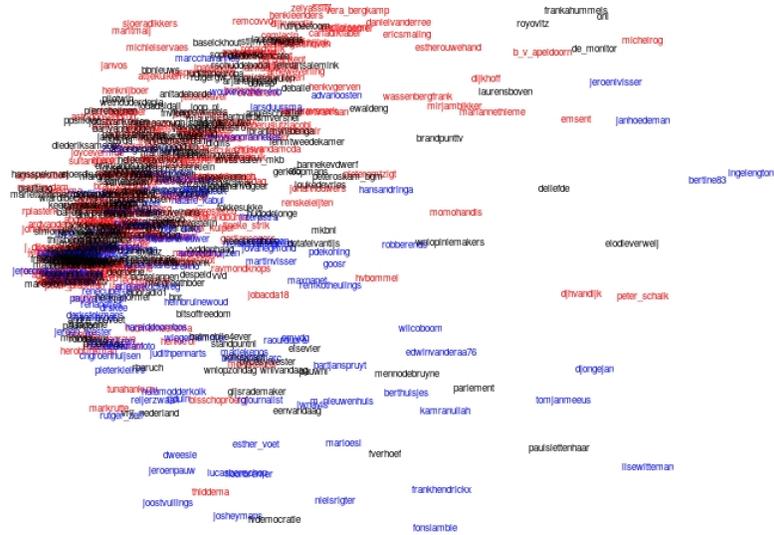


Figure 2: Part of a principal component analysis map with three groups of Dutch Twitter users: politicians (red), journalists (blue) and other (black). There is little overlap between the journalists and the politicians but in this map it was not possible to distinguish the journalists from users that were not politicians.

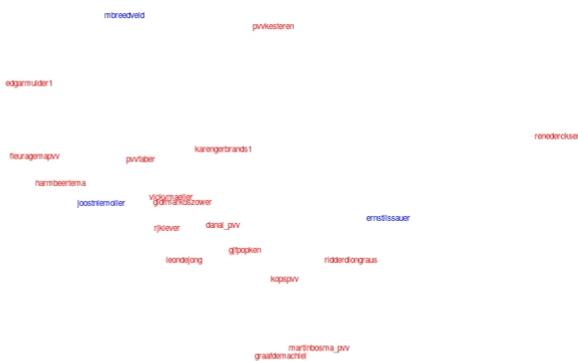


Figure 3: Another part of a principal component analysis map with a group of politicians (red) disjoint of the main group, which all turn out to be member of the extreme right PVV party. Three journalists are clustered close to this group.