



**Selected Papers of AoIR 2016:
The 17th Annual Conference of the
Association of Internet Researchers**
Berlin, Germany / 5-8 October 2016

MINING ONE WEEK OF TWITTER. MAPPING NETWORKED PUBLICS IN THE DUTCH TWITTERSPHERE

Daniela van Geenen
University of Applied Sciences Utrecht

Mirko Tobias Schäfer
Utrecht University

Thomas Boeschoten
Utrecht University

Erik Hekman
University of Applied Sciences Utrecht

Piet Bakker
University of Applied Sciences Utrecht

Jonas Moons
University of Applied Sciences Utrecht

Suggested Citation (APA): Geenen, D. van, Schäfer, M. T., Boeschoten, T., Hekman, E., Bakker, P. and Moons, J. (2016, October 5-8). *Mining One Week of Twitter. Mapping Networked Publics in the Dutch Twittersphere*. Paper presented at AoIR 2016: The 17th Annual Conference of the Association of Internet Researchers. Berlin, Germany: AoIR. Retrieved from <http://spir.aoir.org>.

Introduction

How do “networked publics” (boyd, 2010) evolve and manifest themselves through everyday communication practices in the social media ecosystem Twitter? Twitter is often analyzed in relation to events, high-impact incidents (specific issues or hashtags), or with focus on particular groups represented on the platform (see e.g. case studies in Weller et al., 2014). Others examine the geographical dimension of follow relationships and hashtag use as a national indicator on Twitter (Bruns et al., 2014). However, the investigation of Twitter as infrastructure for everyday communication and social interaction is underrepresented. Taking the name of the medium for the message, this paper casts a view upon the daily chatter of participants. In an effort to map all conversations that constitute the Dutch-speaking Twittersphere over a period of one week, we identify the many conversations and the diverse interrelated publics constituting the ‘static’ of Twitter communication.

Data collection and sampling

The starting point for our analysis is a dataset of more than 4.4 million Dutch tweets distributed between 4 and 12 September 2016¹ gathered in collaboration with media monitoring company Buzzcapture.² With approximately 2.8 million registered accounts (Newcom, 2016), Twitter is exceptionally popular in the Netherlands. To give an impression of its popularity, the country counts a population of around 17 million inhabitants (CBS, 2016). This observation is underscored by a count of 763,255 active users that tweeted during the selected period.

We employed digital methods (Rogers, 2013) to collect and sample the data querying Twitter’s application programming interfaces (APIs). The data extraction was based on language detection using (A) a predefined list of frequently occurring Dutch words and (B) an empirically defined extended list of distinctive Dutch words, in terms of integrity retrieving data through both Twitter’s Streaming and Search API (see Figure 1 for detailed information). Beside the tweet content and the account name, the set consists of metadata such as date and time. Moreover, this study leaves aside geolocation to identify Dutch tweets, as only 0,5 percent of accounts provided this information in tweets (also discussed by Wilken, 2014, p. 161).

¹ We chose to expand the timeframe of one week by including two additional days in order to create the opportunity to follow conversations that started on 10 September 2016.

² For more information visit Buzzcapture’s website: <http://www.buzzcapture.com/>.

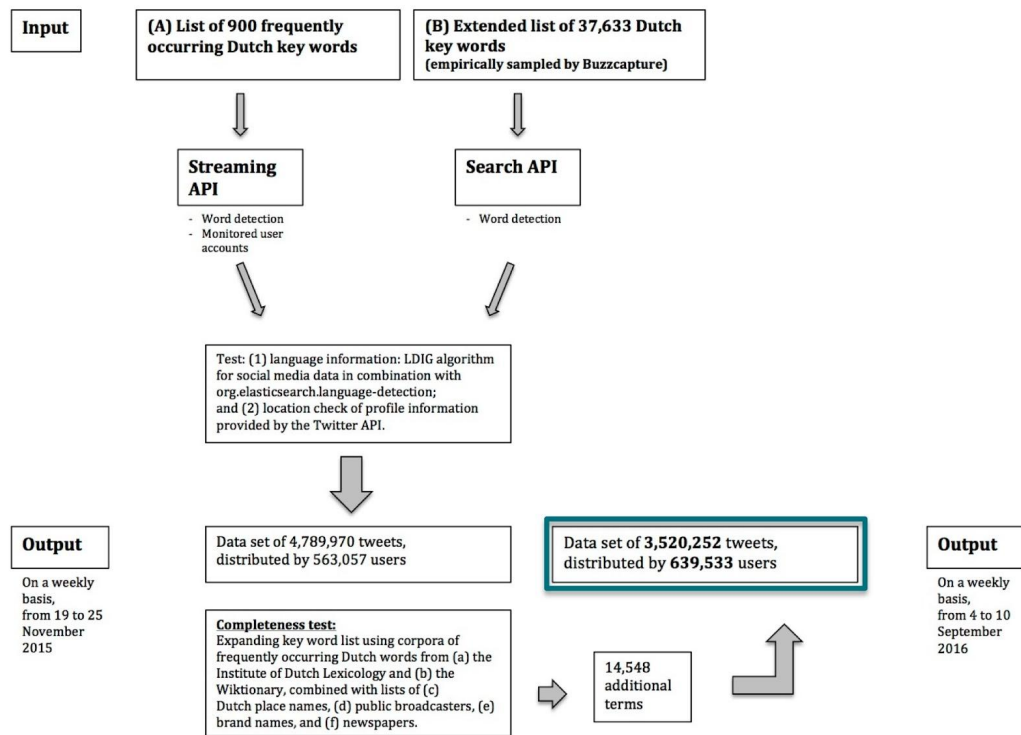


Figure 1. Corpus collection scheme

Methodology

The methodological strength of our investigation is the combination of quantitative and qualitative approaches to collecting, processing, and exploring the data. For the analysis we identified and focused on conversations enabled by Twitter's *@reply* functionality which affords interaction and discussion between (virtually) all Twitter users addressing a tweet's author directly (Bruns & Moe, 2014, pp. 19-23). In this sense, interaction is not necessarily limited to 'friends', users that follow each other, or users that constitute like-minded publics. Whoever takes a stand in a polarized debate, may count on responses by users with contradictory opinions. In that manner replies are suitable for mapping the infrastructure of public debate, as well as networked publics, since people tend to communicate with users similar to themselves, sharing the same interests or profession (cf. Boeschoten, 2015).

We performed social network analysis by the application of Gephi (Bastian et al., 2009) to map and statistically detect communities in the communication infrastructure. These "distant reading" strategies (Moretti, 2013) were combined with interpretative tactics of 'close reading' supported by participatory observations of the Dutch Twittersphere and its publics.

Different dimensions of Twitter use

We singled out two case studies focused on Twitter networks and conversations in two different cities in the Netherlands: 1) Almere and 2) Wijk bij Duurstede. We further investigated the relation of locality and media practice pertaining to these cases. The former is a fairly young metropolitan area, newly founded in the 1970s on reclaimed polder land, located near the capital city, Amsterdam. With 198,823 inhabitants Almere is the eighth most densely populated Dutch city (1 April 2016, CBS), though since 2003 it has been without a daily newspaper (cf. Van Kerkhoven, 2016). In which ways is Twitter used as a news medium and to distribute information to fill the gap of official news coverage? Both cases were chosen in an endeavor to examine if locality and thus, distinct local media practice are traceable in municipalities of diverse dimensions, including size, infrastructure, and demographics. Accordingly, the 23,392 inhabitants of Wijk bij Duurstede, a comparatively small municipality (1 April 2016, CBS), are on average older than the population of Almere. Furthermore, Wijk bij Duurstede is a rather rural commune, though it is located near Utrecht, a university town and the fourth most populous Dutch city.

Local dynamics were explored using a set of tweets distributed between 19 and 25 November 2015 that was collected in an earlier stage of the research (see Figure 1). In this data set of more than 4.78 tweets, we found 1,333 tweets sent by 199 unique users that indicated local engagement with Wijk bij Duurstede, either by means of the location information offered by the Twitter profile settings or the users' profile description. Adopting the same criteria for the identification of tweets and related accounts from Almere, we traced 20,486 messages sent by 2,069 unique accounts.

Findings

In our effort to provide an exhaustive mapping of national Twitter communication, we detected accounts of politicians, media organizations, and journalists, which form the highly connected core of the most active part of the Dutch reply network (see Figure 2). To be specific, accounts in this network sent at least ten responses within the selected timeframe. These clusters of networked publics are often studied posing questions about Twitter's quality as medium for deliberative communication and public debate (for related examples see Weller et al., 2014). Our approach traced and placed individual clusters in the overall context of sent replies during one week in the Dutch-speaking Twittersphere. By doing so, we revealed less visible publics and their everyday communication practices as well. The latter clusters include several well-defined publics (in the visualization below) showing less connection with the center of the network graph, such as YouTubers and gamer scenes, urban, memes and fan cultures. To clarify, the layout algorithm we used, ForceAtlas 2, clusters the network by virtue of repulsing nodes and attracting edges (Jacomy et al., 2014, p. 2).

Also, we found that some accounts that show interaction with a diverse range of Twitter accounts – situated in the center of the graph – are web care accounts, for example, the most prominent account in the reply network hosted by the Dutch National Railways. Other accounts that stand out are those of community police officers. This group is particularly active on Dutch Twitter and connects multiple publics through their accounts. They also connect local publics with the national networks.

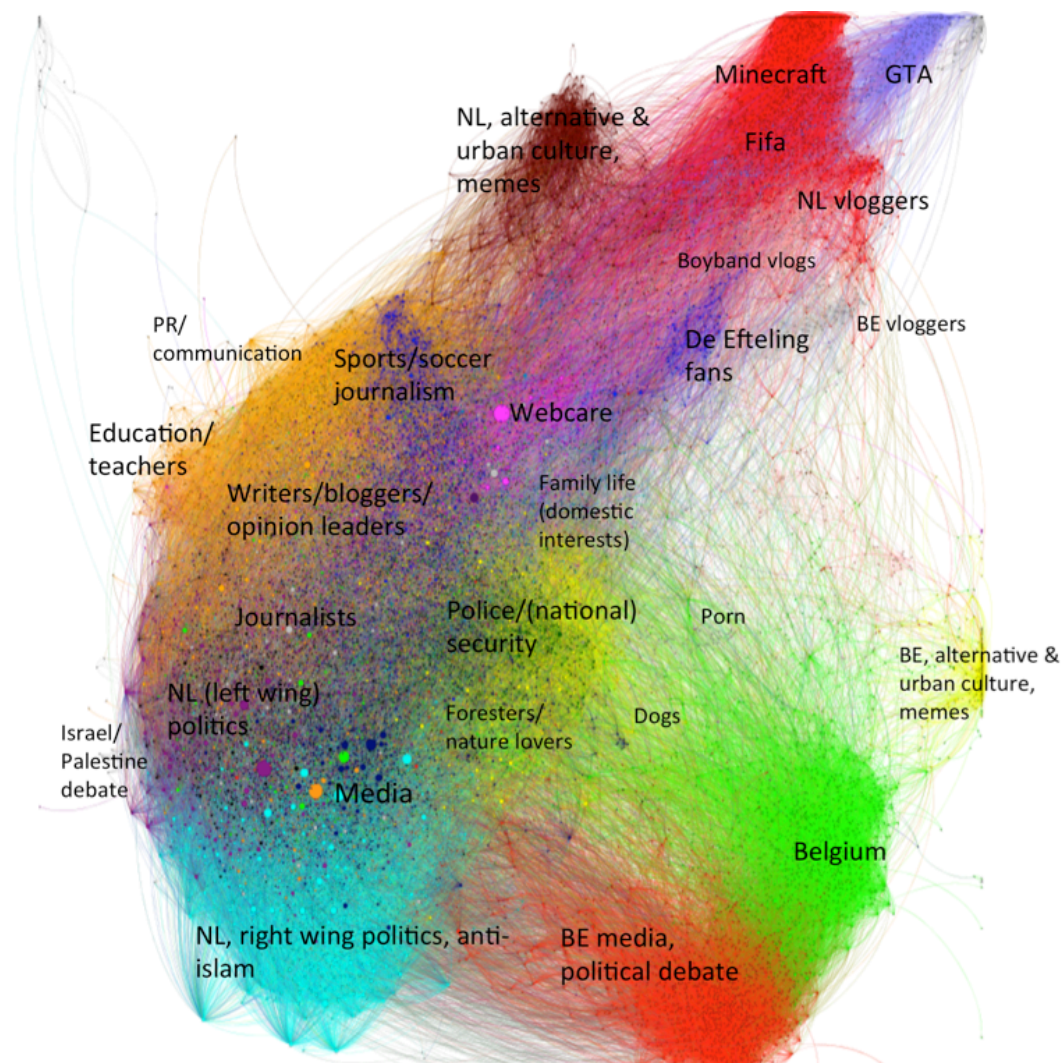


Figure 2. @reply practices in the active top of the Dutch speaking Twittersphere (degree 10) mapped by the application of Gephi's ForceAtlas 2 algorithm and partitioning (colors) based on statistical community detection (modularity)

Thus, our findings show the networked publics engaging in the “exchange of information and points of view” (Habermas, 1996, p. 360) on a national and local level. Many

networks appear to be topic-based, such as the Netherland's YouTube communities on Twitter, the fans of theme park De Efteling, or the community of police officers who cover local and national security topics. On a local level (cities or villages) we can observe the emergence of local Twitter elites. In both local samples these consist of active and highly connected citizens, journalists, and politicians.

In opposite to Wijk bij Duurstede, the Twitter networks of Almere show a higher number of political accounts in the top 100 most active accounts. Almere is one of two larger cities in the Netherlands, where the nationalist PVV (Freedom Party) is one of the biggest parties in the city council. This polarized political sentiment is also recognizable in the local Twitter conversations. The account of the local fraction leader (PVV Almere) is especially telling in its way to handle news: Retweets are often used to place a commentary or an endorsement. Without any commentary, tweets from known (international) rightwing sites are redistributed (e.g. Breitbart), and mainstream media reports are often accompanied by a cynical or critical comment. Moreover, local news is disseminated through tweets related to blogs and other online media dedicated to covering Almere, since a local newspaper is not available to refer to.

Discussion

In view of our findings, this paper addresses three points of discussion providing empirical, theoretical, and epistemological insights. Firstly, valorizing data from Twitter outside the high-awareness events and below the threshold of trending topics, we are able to make statements concerning everyday use of Twitter and to zoom in on the intertwined local and national networks and their various publics.

Secondly, on a theoretical level, we can address these phenomena as forms of "citizen microbroadcasting" (Erickson, 2010, p. 1201). This step provides the opportunity to draw from Habermas' (1989) notion of the "public sphere" and explore its "structural transformation" in local terms and related to topic-based publics, in the context of and through the rise of social media. While the many publics constitute both a multiplication and an increasing fragmentation of publics, they expose rather different media practices. The deliberation of societal questions, in order to support opinion formation and democratic organization is only one of many uses. However, these Twitter-based networks are still connected to other media referring to them as cultural references or as source of information. With an awareness of the techno-economical and socio-technical fragmentation of web platforms (Van Dijck, 2013, p. 163), and taking geographical distribution into consideration, we map a plurality of (interconnected) public spheres. Further investigation of interrelations between participants on local and national level, including professional media outlets, politicians, or members of civil public allows us to make an in-depth analysis of power relations and communicative effects.

Thirdly, on an epistemological note, the study raises the issue of completeness and representation of Twitter data sets and critically reviews the possibilities and limitations of social media metrics. Since we faced selection bias in monitoring all Dutch actors participating on a regular basis, we enhanced the data extraction method in order to improve the initial data material. Furthermore, this improvement of the collection method allows us to make inferences discussing the completeness of data mined through Twitter's APIs (see e.g. collection scheme as Figure 1). Therefore, a further elaboration of this aspect will respond to the swelling call for "data critique" (Rieder et al., 2015).

References

Boeschoten, T. (2015). *Stedelijke publieken op Twitter* (Master's thesis). <http://dSPACE.library.uu.nl/handle/1874/311619>.

boyd, d. (2010). Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications. In Papacharissi, Z., *Networked Self: Identity, Community, and Culture on Social Network Sites* (pp. 39-58). New York: Routledge.

Bruns, A., Burgess, J., & Highfield, T. (2014). A "Big Data" Approach to Mapping the Australian Twittersphere. In P. L. Arthur & K. Bode (eds.), *Advancing Digital Humanities: Research, Methods, Theories* (pp. 113-129). Basingstoke, Hampshire: Palgrave Macmillan.

Bruns, A. & Moe, H. (2014). Structural Layers of Communication on Twitter. In K. Weller et al. (eds.), *Twitter and Society* (pp. 15-28). New York: Peter Lang.

CBS (2016). StatLine. In *Centraal Bureau voor de Statistiek*. <http://statline.cbs.nl/Statweb/search/?Q=inwoneraantallen>.

Dijck, J. van (2013). *The Culture of Connectivity. A Critical History of Social Media*. New York: Oxford University Press.

Erickson, I. (2010). Geography and community: New forms of interaction among people and places. *American Behavioral Scientist*, 53(8), pp. 1194-1207. <http://journals.sagepub.com/doi/abs/10.1177/0002764209356250>.

Habermas, J. (1989). *The Structural Transformation of the Public Sphere*. Cambridge, MA.: MIT Press.

--- (1996). *Between Facts and Norms: Contributions to a Discourse Theory of Law and Democracy*. Cambridge: Polity.

Jacomy, M., Heymann, S., Venturini, T. and Bastian, M. (2014). ForceAtlas2. A Continuous Graph Layout Algorithm for Handy Network Visualization. *PLoS ONE* 9(6). <http://dx.doi.org/10.1371/journal.pone.0098679>.

Kerkhoven, M.C. van (2016). *Lost in transition: Media innovations in the Netherlands*. Ph.D thesis. University of Amsterdam. <http://hdl.handle.net/11245/1.535977>.

Moretti, F. (2013). *Distant Reading*. London: Verso.

Newcom (2016). *Nationale Social Media Onderzoek 2016*. www.newcom.nl/socialmedia2016.

Rieder, B., Abdulla, R., Poell, T., Woltering, R. and Zack, L. (2015). Data critique and analytical opportunities for very large Facebook Pages: Lessons learned from exploring "We are all Khaled Said", *Big Data & Society*, 2(2). <http://bds.sagepub.com/content/2/2/2053951715614980>.

Rogers, R. (2013). *Digital Methods*, Cambridge, MA.: MIT Press.

Weller, K., Bruns, A., Burgess, J., Mahrt, M. & Puschmann, C. (2014). *Twitter and Society*. New York: Peter Lang.

Wilken, R. (2014). Twitter and Geographical Location. In K. Weller et al. (eds.), *Twitter and Society* (pp. 155-167). New York: Peter Lang.