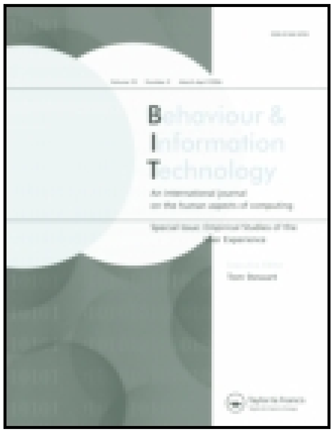




Behaviour & Information Technology
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/tbit20

# Police message diffusion on Twitter: analysing the reach of social media communications

Bob van de Velde[a], Albert Meijer[b] & Vincent Homburg[c]

[a] Communication Science, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands

[b] Utrecht University School of Governance, Utrecht University, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands

[c] Faculty of Social Sciences (Public Administration), Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands
Published online: 26 Aug 2014.

To cite this article: Bob van de Velde, Albert Meijer & Vincent Homburg (2015) Police message diffusion on Twitter: analysing the reach of social media communications, Behaviour & Information Technology, 34:1, 4-16, DOI: 10.1080/0144929X.2014.942754

To link to this article: http://dx.doi.org/10.1080/0144929X.2014.942754

*Behaviour & Information Technology*, 2015
Vol. 34, No. 1, 4–16, http://dx.doi.org/10.1080/0144929X.2014.942754



# Police message diffusion on Twitter: analysing the reach of social media communications

Bob van de Velde[a*], Albert Meijer[b] and Vincent Homburg[c]

[a]*Communication Science, VU University Amsterdam, De Boelelaan 1081, 1081 HV Amsterdam, The Netherlands;* [b]*Utrecht University School of Governance, Utrecht University, Bijlhouwerstraat 6, 3511 ZC Utrecht, The Netherlands;* [c]*Faculty of Social Sciences (Public Administration), Erasmus University Rotterdam, PO Box 1738, 3000 DR Rotterdam, The Netherlands*

(*Received 9 July 2013; accepted 29 June 2014*)

Social media are becoming increasingly important for communication between government organisations and citizens. Although research on this issue is expanding, the structure of these new communication patterns is still poorly understood. This study contributes to our understanding of these new communication patterns by developing an explanatory model of message diffusion on social media. Messages from 964 Dutch police force Twitter accounts are analysed using trace data drawn from the Twitter™ API to explain why certain police tweets are forwarded and others are not. Based on an iterative human calibration procedure, message topics were automatically coded based on customised lexicons. A principal component analysis of message characteristics generated four distinct patterns of use in (in)personal communication and new/versus reproduced content. Message characteristics were combined with user characteristics in a multilevel logistic general linear model. Our main results show that URLs or use of informal communication increases chances of message forwarding. In addition, contextual factors such as user characteristics impact diffusion probability. Recommendations are discussed for further research into authorship styles and their implications for social media message diffusion. For the police and other government practitioners, a list of recommendation about how to reach a larger number of citizens through social media communications is presented.

**Keywords:** micro-blog; Twitter; message diffusion; police



## 1. Introduction

Although social media were introduced to support networks of friends, they quickly made their entrance in the domain of politics and administration. Much has been written about the use of social media in democratic elections to mobilise voters and influence message frames (Lassen and Brown 2011; Kim and Park 2012; Stieglitz and Dang-Xuan 2013) and to increase civic engagement and political participation (Valenzuela, Park, and Kee 2009; see also Ellison, Steinfield, and Lampe 2011; Vitak and Ellison 2012 for analyses of social media use implications for social capital). At the same time, social media have also been embraced by government organisations to enhance their performance and legitimacy (Mergel 2012). Government organisations as diverse as the Federal Bureau of Investigations, the Environmental Protection Agency and a host of state and local governments are using Facebook, Twitter and YouTube to improve their communication with citizens. Social media are used not only to inform citizens about government activities and improve their services, but also to obtain information from citizens and engage them in processes of coproduction of government policies (Meijer and Thaens 2013; Meijer 2013).

In the USA and also in Canada, Australia, New Zealand, European, Latin American and Asian countries, the micro-blog platform Twitter has been adopted by a large number of government organisations to communicate with citizens. This social media platform is based on the premise ‘less-is-more’ (Finin and Tseng 2007): it limits the amount of characters per message to create bite-sized messages that can easily be consumed by users. This service has been broadly adopted not only by individuals, but also by companies, non-profit organisations, politicians and government (officials) (Burton and Soboleva 2011; Cho and Park 2011; Li et al. 2011; Picazo-Vela, Gutiérrez-Martinez, and Luna-Reyes 2011; Rojas, Ruiz, and Farfán 2011; Waters and Williams 2011). Some authors note that the medium can inform about societal trends (Asur and Huberman 2010; Khrabrov 2010; Bae and Lee 2012; Vergeer, Hermans, and Sams 2011), or discuss how it influences society (Chew and Eysenbach 2010; Christensen and Lægreid 2011).

One of the government domains that shows active social media communication is policing (Heverin and Zach 2010; Crump 2011; Procter et al. 2013). In part through word-of-mouth processes (Wang and Doong 2010), police departments all around the world have started adopting the

*Corresponding author. Email: r.n.vande.velde@vu.nl

medium. In the UK, the London riots provided an impetus for police officials to use Twitter and other social media to provide citizens with moment-to-moment updates to combat rumours, discuss incidents and reassure the public (Crump 2011). Elsewhere, police officials have started to adopt the medium to improve their information sharing capacity and thereby actively engage citizens in solving problems (Heverin and Zach 2010). The basic premise is that better-informed citizens can contribute to public safety by taking proper preventative measures, avoiding hazardous locations or providing relevant information to the police. Reaching citizens with police communications is essential to obtaining these objectives, but little is known about the audience being reached.

While some research has been done in the number of followers (Crump 2011), one cannot assume that followers actually read all messages. Active processing of these messages can be regarded as a better indicator of the diffusion of information through social media (Kulshrestha et al. 2012). Message forwarding in the context of Twitter takes place using ‘retweets’. This is especially important for ‘wanted’ and ‘warning’ messages, which may result in a crime solved or prevented for every person reached. However, message diffusion (retweets) via this type of medium is still poorly understood, especially in the context of policing. At this point, it is unclear which message and sender characteristics influence the diffusion of information through social media.

The current body of literature mainly consists of conference papers with explorative data rather than explanatory models. This means many causes have been identified, but have not been tested in a rigorous fashion. In addition, little theoretical attention has been paid to the relation between tweet and user characteristics. Such relations might entail bias towards well-connected users and their tweeting behaviour. This study seeks to build upon existing knowledge about the diffusion of government information through social media and test their applicability to messages authored by Dutch police officials. The research question is: how can differences in the retweet rate of messages authored by Dutch police officers be explained? On the basis of a basic analysis of these retweet patterns, this study contributes to our theoretical understanding of the structure of social media communications between government and citizens.

## 2. Literature review: predictors of message diffusion on micro-blogs

Although previous studies do not explicitly differentiate between message and user characteristics, this study will do so in order to better understand their relationship. As mentioned before, the separation has both methodological and theoretical reasons. Mainly, user characteristics are taken to hold across the messages created by that user, which means the effective sample size for these variables is significantly smaller. In a theoretical sense, user characteristics are contextual factors for individual tweets. The messages sent by popular users have a different starting point to those of unpopular users, which may affect their chances of diffusion regardless of their characteristics. For the sake of clarity, tweet and user[1] characteristics are discussed separately. The insights from literature are summarised in Table 1, based on the type of variables discussed by these authors.

### 2.1. *Message characteristics*

Most research has focused on message characteristics, which are often directly observable. Suh et al. (2010) find that the inclusion of web addresses and hashtags affects the chances of diffusion. The inclusion of hashtags signals the broader discussion a message is part of and increases searchability (Zappavigna 2011), whereas a web address is used to provide readers with additional information. This can be a news article, picture, video clip or (longer) blog posts. To this extent, the inclusion of resources such as URLs and hashtags provides additional information, which may heighten their information value to recipients and thereby increase chances of being forwarded. Such communicative capacity is deemed especially relevant for government actors, who aim to inform citizens (Ampofo, Anstead, and O’Loughlin 2011; Picazo-Vela, Gutiérrez-Martinez, and Luna-Reyes 2011).

Apart from these elements, the topic discussed has been linked to diffusion. Research into Twitter activity has shown that popular or current topics can increase both activity and information sharing (Hansen, Arvidsson, and Nielsen 2011), especially when they fit with audience expectations (Weng et al. 2010). A study into the use of Twitter by police departments has supported this finding in the law enforcement context (Heverin and Zach 2010). This implies that messages may differ in terms of appeal based on the topic they discuss. In the context of law enforcement, it may mean that missing person reports garner more sympathy and are, therefore, more often forwarded compared to traffic information, but small talk may actually reduce chances of being retweeted.

In addition to the resources and topics of a tweet, the social characteristics are of interest. Suh et al. (2010) found a decreasing effect of mentions on retweet probability. But as with replies, these social factors are linked to user popularity (Burton and Soboleva 2011; Wigley and Lewis 2012). On the tweet level, both are expected to reduce diffusion because of their orientation towards specific rather than broad audiences.

### 2.2. *User characteristics*

In terms of user characteristics, there has been considerable attention to the style of communication employed on

Table 1. Overview of micro-blog literature discussing diffusion or related concepts.

| | Independent | Dependent | Source |
|---|---|---|---|
| Message characteristic | URL inclusion | Diffusion | (Hansen, Arvidsson, and Nielsen 2011; Suh et al. 2010) |
| | Hashtag inclusion | Diffusion | |
| | Mention | Diffusion | |
| | Reply | Audience size | (Burton and Soboleva 2011; Wigley and Lewis 2012) |
| | Topic discussed | User activity | (Heverin and Zach 2010[a]; Crump 2011[a]; Wang and Li et al. 2011) |
| User characteristic | Organisation type | Replies to audience | (Cho and Park 2011; Golbeck, Grimes, and Rogers 2010; Rojas, Ruiz, and Farfán et al. 2011; Waters and Jamal 2011; Waters and Williams 2011) |
| | Organisation type | Mentioning of others | |
| | Engagement | Diffusion | (Wigley and Lewis 2012; Zhang, Jansen, and Chowdhury 2011) |
| | Informativeness | Decreased audience | (Kwak and Chun 2011) |
| | Account age | Diffusion | (Suh et al. 2010) |
| | Total messages posted | Diffusion | |
| | In-links | Diffusion | (Lussier and Chawla 2011; Suh et al. 2010) |
| | Out-links | Diffusion | |

[a]Article specifically discusses police or police officer use of the medium.



the medium. Such stylistic differences can be divided into roughly two kinds: interactivity and authorship. Interactivity refers to the extent to which a user uses conversational tools such as mentions and replies. A number of studies show that organisational accounts often lack such interactive elements, opting instead for classic one-way communication (Golbeck, Grimes, and Rogers 2010; Cho and Park 2011; Rojas, Ruiz, and Farfán 2011; Waters and Jamal 2011; Waters and Williams 2011). This contradicts the advice of studies into business engagement, which finds significant positive effects of interaction on message diffusion (Zhang, Jansen, and Chowdhury 2011; Wigley and Lewis 2012). This study will employ the use of an interactive style as engagement, which in turn is expected to predict message diffusion. Authorship is another way for users to distinguish themselves. Research has shown that the role of users as information channels in part determines the retention of their audience (Kwak, Chun, and Moon 2011). This raises a point about the role of accounts as conduits for new information. Some users will frequently forward messages from others, thereby providing information from various sources, whereas others are expected to use their accounts mainly for their own information.

A user also brings some experience to the table. Although Suh et al. (2010) do not distinguish user and message-based causes of diffusion, they do find a significant influence of both account age and number of messages posted. Such influences may be the result of a learning effect. Prolonged use has been linked to socialisation or adaptation to audience preferences (Marwick and Boyd 2010). Because of this, experienced users are more capable of connecting with their audience. In addition, it allows for users to become accustomed to the medium and feel more secure in its use (Chen 2011). These factors should, therefore, increase message quality and thereby the chances of diffusion.

In addition to the type of relations a user has built, the quantity of relations has been connected to message diffusion (Suh et al. 2010). A user has a number of in-links, from which messages are received automatically. This provides a user with information about what is going on in their network and the ability to forward interesting news and expectations (Marwick and Boyd 2010; Weng and Lee 2010) – a benefit which has been linked to increased retweets (Lussier and Chawla 2011). A higher number of in-degree links is therefore hypothesised to increase the ability of a user to serve his or her followers. These followers form the audience of a Twitter user. They not only receive messages posted by a user, but also have the ability to forward these messages, thereby making them available to all of their followers (Bae and Lee 2012). In this way, followers are gatekeepers who decide whether or not to spread a message to their own audience. Having more followers is thereby expected to increase the chances of having a message forwarded.

### 2.3. *Building a model of message diffusion*

Together, the elements described in Sections 2.1 and 2.2 are combined in an explanatory model of message diffusion shown in Figure 1. In addition to relationships within each level, the model includes interaction between levels. The direct inter-level relation is represented by an arrow crossing the levels; the moderating effect uses the cross-level

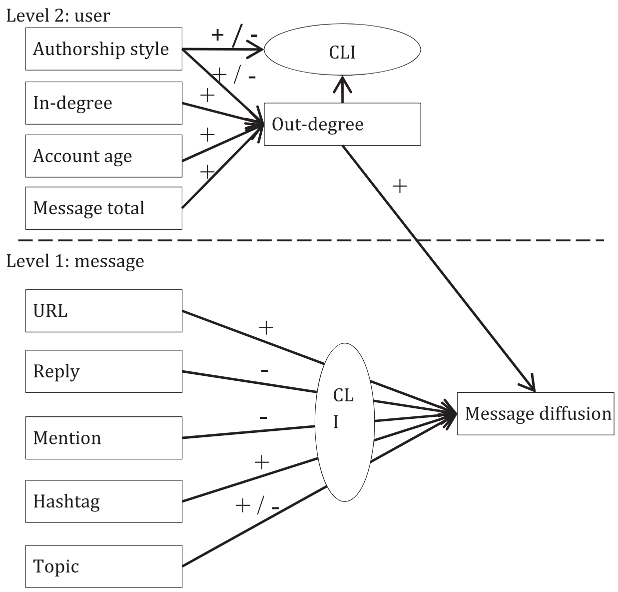


Figure 1. Model of message diffusion on the Twitter micro-blog service.

interaction (CLI) ellipse. Two of these indirect effects have been hypothesised. The first one is the moderating relation between authorship style and out-degree on each of the message-level characteristics. The second is the direct mediating effect of out-degree on message diffusion.

User characteristics such as authorship style, in-degree and experience are expected to contribute indirectly to message diffusion. Out-degree is expected to mediate their influence, and is itself expected to have a direct relation with diffusion, because in-degree and experience of an authorship help build an audience. A direct effect of other author characteristics is not expected, as these are reflected in the composition of message characteristics. Instead, in-degree, account age and message total (the last two variables being operationalisations of experience) are expected to enable a user to serve his/her audience, thereby increasing the out-degree. Similarly, authorship style is expected to impact the number of followers based on the personal or original nature of the user's tweets. Such stylistic properties may increase or decrease relational value for followers. The out-degree is expected to have a direct effect on message diffusion. This is because tweets that neither feature URLs, mentions or hashtags, nor reply to a user or discuss a measured topic, may benefit from greater exposure and thus greater likelihood of diffusion.

Finally, an interaction effect between levels is hypothesised to moderate the effect of message-level variables on message diffusion. The CLI ellipse denotes this cross-level moderating effect. The out-degree interaction is based on logic of increasing returns. When the audience size increases (big out-degree), the effect of a tweet-level characteristic is applicable to more 'gatekeepers'. This means that for each person reached, the effect of such a characteristic again applies. Authorship style is expected to moderate the effect of these characteristics based on audience expectations; users who have a more personal style are likely to draw more attention when compared to impersonal users. Because of this, their use of message elements such as URLs or hashtags may result in stronger effects.

## 3. Methods

In order to actually carry out the analyses, specific design choices with respect to the operationalisation of key variables, procedures for data gathering and application of statistical models and techniques had to be made. The line of reasoning that led to these specific choices for a research strategy is explained in the following subsections.

### *3.1. Measures*

In inclusion of directly observable elements such as hashtags, URLs and mentions are operationalised as their presence or absence in a given message. Replies were operationalised as given by Twitter, defined by the inclusion of a mention in the first character space of the message. Mentions were recoded to 0 if a message was a reply to reduce confounding mentions and replies. URL and mention variables are coded 1 if present and 0 if absent. Only hashtags were counted based on the use of symbol # in the message text in order to gauge the added effect of multiple hashtag inclusions. Table 2 summarises these operationalisations.

To measure the influence of topics, the qualitative framework of police message subjects listed by Heverin and Zach (2010) is adapted for the Dutch context, which yields eight topic categories dealing with crime/incident reporting, department (activity) information, event information, traffic information, prevention aimed information, (witnesses) wanted requests, missing person information and small talk (see Appendix 1 for a description). These topics are operationalised based on word use. In the exploratory phase, 200 tweets were coded by the principal researcher for the main topics, which were operationalised based on

Table 2. Overview of micro-blog literature discussing diffusion or related concepts.

| Variable | Type | Mean/ percentage | Standard deviation |
|---|---|---|---|
| Posts | continuous | 802 | 1098 |
| Account age | continuous | 461 | 286 |
| Friends | continuous | 252.7 | 470 |
| Followers | continuous | 1152 | 1125 |
| URL | binary | 23% | |
| Mentions | binary | 16% | |
| Reply | binary | 83% | |
| Hashtag | continuous (per message) | 0.7 | 1.58 |
| Topic | binary (per topic) | between 25% (crime/incident) and 2% (missing persons) | |

Heverin and Zach (2010). For each topic, a list of often-used words was formulated and a full lexicon for the analysis was designed (Appendix 1).

Crime and incident reports contain tweets which state a crime or incident which has occurred; the lexicon of this topic includes terms such as 'apprehended', 'burglary' and 'investigate'.[2] Departmental information covers tweets about meetings, office hours, upcoming projects and internal affairs. This list includes 'meeting', 'project' and 'office hours'.[3] The event topic considers tweets that discuss not only current or upcoming events such as marathons, parades and demonstrations, but also information events. This wordlist includes 'campaign', 'strikes' and 'educate'.[4] The traffic topic includes tweets about crashes, traffic jams and traffic controls and tweets that deal with driving. Keywords in this list are 'traffic', 'speed' and the abbreviation of 'near'.[5] Tweets discussing prevention include warnings and suggestions to prevent crime. This wordlist includes 'tips', 'prevent' and 'warning'.[6] Messages which call for witnesses or tips concerning suspects are coded by their use of police hotlines and words such as 'witness?' and 'seen something?'.[7] Missing person related tweets are coded using 'missing' and 'last seen'.[8] Small talk is perhaps the broadest category and includes tweets in which officers tell what they are doing, respond to others and communicate about their state of mind and opinions. This category includes smiley symbols, as well as words such as 'fun' and 'nice'.[9] Tweets were then automatically coded 1 on a topic if it contained an exact match of a topic word. In an iterative calibration process, a random sample of 100 tweets was checked for human–computer coding congruence (for a discussion of manual versus automatic coding, refer to Crowston et al. 2010). Based on the resulting insights, wordlists were again updated and the comparison was repeated.[10] After four rounds of calibration of the full list by the principal researcher, the error rate was reduced to 4.3%–10.7% false positives ($\alpha = 0.05$, $n = 200$). This means messages are unlikely to be wrongfully attributed to a topic.[11] Due to the chances of false negatives, the influence of topics is expected to be deflated. The false-positive rate may reduce statistical power to detect topic influence, but the low false-positive ratio preserves reliability.

In terms of in-degree and out-degree contacts, Twitter data offer elegant solutions. In-degree contacts come in the form of 'friends'.[12] This metric is the count of Twitter users whose posts are automatically forwarded to a user. Out-degree contacts are registered in a similar statistic called 'followers' and are the logical opposite of friends. This is the number of accounts which receive a posted message. Experience was operationalised using the creation date of each account and the number of statuses since creation. The former provides an operational measure of users' history with the medium and the latter reflects previous engagement. Both are used as interval-level variables with respect to their impact on message diffusion. All these variables are measured at the time of data collection and are, therefore, constant for each tweet of a user.

Authorship is the only indirectly measured concept. Different styles are operationalised as recurrent use of message types, topics and elements.[13] In terms of types, a message can be a reply (as on the tweet level) or a retweet. If a message is a retweet, we draw the distinction; a retweet can be internal (retweeted from another police account) or external (from general public). To enable this distinction, both types are separated. Internal retweets are coded 1 if the message is a retweet and the source is an account from the police account list used in data collection. Retweets which are external are coded as 0 to maintain the reduce confounding. The messages types, topics and elements used are aggregated to the user level as averages. Styles are then defined using an explorative principal component analysis, in order to re-express the use of message types and characteristics (Lattin, Carroll, and Green 2003). This principal component analysis is applied to a subset of the data, of which duplicate users are removed. This is done to prevent bias towards users with more tweets. Components are selected based on the scree-plot method (Lattin, Carroll, and Green 2003: 114) and variables loading more than 0.3 on a component are scaled together and used as authorship dimensions. These are subjected to a Guttman lambda six test to ensure coherence. For the statistical tests, these items are combined into a scale based on their average. These scales are normalised in order to avoid multicolinearity issues.

We define the dependent variable, message diffusion, as the reach of a message beyond its initial audience. As such message diffusion is understood in a binary fashion, either a message is forwarded or not. This decision is made to deal with the extremely skewed distribution of retweets, making the data unsuitable for reliable normal regression. A message which has been forwarded corresponds to a retweet on Twitter, as declared by the integrated retweet function. Such messages are considered diffused, whereas a message which has not been 'retweeted' is considered non-diffused. The retweet variable is coded 1 if forwarded and 0 if not.

### *3.2. Data collection*

For data collection the Twitter™ REST API was used, which provides a maximum of 200 of the most recent messages posted by an unshielded user. A list of accounts was generated through searches on the Internet and on Twitter. A preliminary list was published on a popular platform for social media in the Dutch police and the list was finalised with a few additional accounts. For each of the 1000 accounts in this list, a request was made for their messages, including retweets and user characteristics. The requests to this service were made on 27 April 2012, using a custom program which saved the responses in a comma-separated file. Message and user characteristics were used as supplied by Twitter™, including retweet counts, message text, followers, friends, creation date, total number of statuses and whether a message contains a URL, mention or reply.[14] In the literature, there are reports of ethical concerns about the use of data extracted from Internet sources (Buchanan and Zimmer 2013; see also Rogers 2013 for a description of how Twitter has settled into a dataset for academic research). As only publicly available tweets from unshielded users were used, the express ambition was to inform the general public, and personal characteristics were scrubbed from research outputs, and researchers refrained from asking each and every user to agree on the use of their Twitter data (see also Reinberg 2009).

### *3.3. Data analysis*

Since message diffusion is a binary variable, the model will be tested using a logistic regression model. This method is designed to estimate the odds of either value of a dichotomous-dependent variable (Pampel 2000), in this case the chance of diffusion. During this analysis, a subset of the data is used which excludes tweets, which are themselves retweets. This is because this study is interested only in tweets authored by police officials, not diffusion of messages only forwarded by police officials. The collection of data gives us multiple tweets per user, a form of cluster sampling. Because the sample is clustered, a multilevel model is used to control for dependent sample bias (Bickel 2007; Khan and Shaw 2011). In addition, this method allows CLI to be modelled using variables from different levels in conjunction (Bliese 2012). In this way, contextual variables can be tested which are aggregated from individual-level raw data. This prevents problems estimating the effect of authorship style, which is aggregated from individual-level data. In terms of the required sample size, our data fit the '10 observation-per-independent-variable' rule on both levels (Level 1 $n = 106462$, Level 2 $n = 964$) (Garson 2009). The sample size is also bigger than the minimum of 20 groups with 30 observations required for multilevel regression (Bickel 2007). In addition, multilevel logistic regression requires the fulfilment of regular logistic assumptions on each level (Gelman and Hill 2007). These are normality for continuous variables,[15] low multicolinearity, non-additivity and linearity (Hosmer and Lemeshow 2000). To deal with inflated deviance due to non-normally distributed values, a logarithmic transformation was applied to previous messages, friends and followers. Multicolinearity was checked using correlation tables and yielded no highly collinear relations. Finally, interaction effects and non-linear relations were taken into account by exploring all possible additions to the model. In addition, running the Cook test indicated no influential outliers (Cook's distance $>0.02$). Time of day and characters in a message were added as control variables. All relations improving model fit are adopted in the final model.[16]

Moderating effects are checked using interaction terms, which shows how differences in Variable A impact the relation between Variables B and C (Whisman and McClelland 2005). When these interactions prove significant, the relation is interpreted as moderating. For mediating effects, the Baron and Kenny test is applied, to test the direct effect of the mediator and independent variable, the independent variable on the mediator and whether there is a decreasing effect of the independent variable when the mediating variable is included (Baron and Kenny 1986).

## 4. Results

The application of the procedures as described previously resulted in a description of users' Twitter behaviours, as well as in an empirical test of the model derived in Section 2. The descriptions of users and messages produced by users, as well as the results of the model test (for details, refer to Appendix 2), are presented in the following subsections.

### *4.1. Descriptive statistics: what are the police doing on Twitter?*

The 964 accounts gathered range from those of street police officers (76%) to those of district-level managers (3%). Their specific functions range from general community officer (65%), other (14%),[17] youth officer (4%), manager (2%) to PR representative (1%). As such, this study had a high percentage of local accounts as compared to the sample examined by Crump (2011). This could be the result of the strong change in the force-to-local accounts ratio discussed by this author, or cultural effects (Poblete et al. 2011). The total amount of posts per account ranges from 1 to 10,357, with a mean of 802 (standard deviation (SD) = 1098). The oldest account is just under 4.5 years (1645 days) old, whereas the newest account started 35 days before collection. The amount of followers range from 6 to 21,084, which is positively skewed with an average of 1152 and a median of 560 (SD = 1125), just as for friends, which vary between 0 and 5715, with a mean of 252.7 (SD = 470).[18]

The most discussed topic matched that of earlier studies (Heverin and Zach 2010), which is crime and incident reporting (25%). In our coding scheme, this topic is followed by small talk (15%) and (witnesses) wanted/'look out for' messages (11%). About 23% of these messages contain a URL, which most often links to the official police website (27%), followed by the yfrog™ and twitpic™ picture sharing service (12% and 3%), and the Youtube video sharing at 3%. Hashtags use averaged at 0.67 per tweet, with a maximum of 7 in one tweet. In all messages, 20% formed a reply but only 5.5% used a mention. Most tweets (41%) were sent in the afternoon, 28% in the evening, 26% in the morning and 5% at night.[19]

The first two components estimated by the principal component analysis conform to theoretical expectations. On the first component, there are four elements, which load above 0.3. The user average of small talk and replies load positively. Negatively contributing are URL use and crime/incident reporting. Both reply and small talk are means of communicating on a personal level, by addressing one specific person or discussing topics more personal in nature. In this sense, the use of URLs (which mostly refer to the official police website) and crime reporting are more impersonal in nature. Combined, the first component is taken to cover the personal (positive) to impersonal (negative) dimension (Guttman's lambda six = 0.716). There are three variables that load on the second component. These are internal retweets, external retweets and mentions, which all load negative. As a combined measure, these items are taken to denote the authorship–messenger dimension (Guttman's lambda six = 0.919). The higher the scale, the less a user forwards messages or mentions others. As such, the higher a user scores on the authorship scale, the more the more tweets are self-authored and the less reference is made to others. Together, these stylistic dimensions explain the extent to which a user chooses an (im)personal tone and acts more as a messenger. In our sample, there is a tendency towards a personal style (negatively skewed distribution). In terms of authorship, the distribution strongly negatively biased in favour of authorship over messenger properties (negatively skewed). Both these dimensions will be used in the model.

### 4.2. *Model results: predictors of retweets*

The observed best-fit model significantly reduced variance compared to a null model ($\rho < 0.01$, McFadden's pseudo-$R^2 = .14$). All effects mentioned are significant at the $\alpha = 0.01$ level, unless specified otherwise; insignificant predictors are predictors with a $\rho$-value greater than 0.05. The effect sizes mentioned are deviations from a 'joe average' user's tweet, with average characteristics and lacking any specific topic or element (which has a base chance of 1.26% of being retweeted). Such comparisons to a baseline tweet take non-linear and interaction effects into account. For comparisons concerning continuous variables, the difference between the first and third quartiles is used. Note that all estimates are based on a logarithmic scale and can therefore not be directly added.

Findings regarding user-level variables support the expected direct effect of followers on retweet probability, with an increase of 0.56% over generic tweets. In addition, the number of friends has no statistically significant direct effect on retweet probability. Contrary to the expectations of the theoretical model, experience in terms of days active has a strong direct negative effect on retweet probabilities (−1.17%). This is also true for the total number of messages posted by a user, although this effect is smaller (−0.25%). There was also a direct effect of authorship, contrary to the theoretical model; original authorship reduces retweet chances (−0.40%). The effect of sending more personal messages is smaller (−0.03%). The effect of authorship styles diminishes as a user gets more followers.

Message-level variables both support and contradict expectations. Supporting the theoretical model, URLs and hashtags have a positive effect on retweet probabilities with 1.77% and 0.24–0.38% respectively. Contrary to the model, mentions also contribute to message diffusion, with an increase of 1.04% over generic tweets. Replies do conform to theory and reduce the chances of a retweet with 0.39%. In terms of topics, significant effects are observed for all but event-related messages. The strongest effects are related to missing person (+3.04%) and (witnesses) wanted (+1.56%) tweets.[20] In terms of interaction effects, combinations of replies and URLs increase chances of diffusion. URLs in traffic messages and replies about (witnesses) wanted combinations have diminished effects of diffusion.

Conforming to the model, the data support the strengthening effect of followers on the relation between replies, URLs and message diffusion. But contrary to the model, mentions become less, rather than more influential, when an account has a bigger audience (a −0.82% decrease in likelihood). The moderating role of authorship is supported. Personal authorship increases the effect of URLs beyond the decrease resulting from the direct effect of this style. Crime- and incident-related messages are by contrast less likely to diffuse if sent by personal-type authors. The moderating role of original authorship is statistically significant but small and only applicable to the use of mentions. Both experience measures have an unexpected moderating effect, with the number of messages reducing the effect of mentions by 1.55%, yet increasing the chances for replies by 0.25%. The age of an account also increases chances for replies to diffuse, perhaps suggesting reputation effects. By contrast, both experience variables are negatively linked to traffic message diffusion.

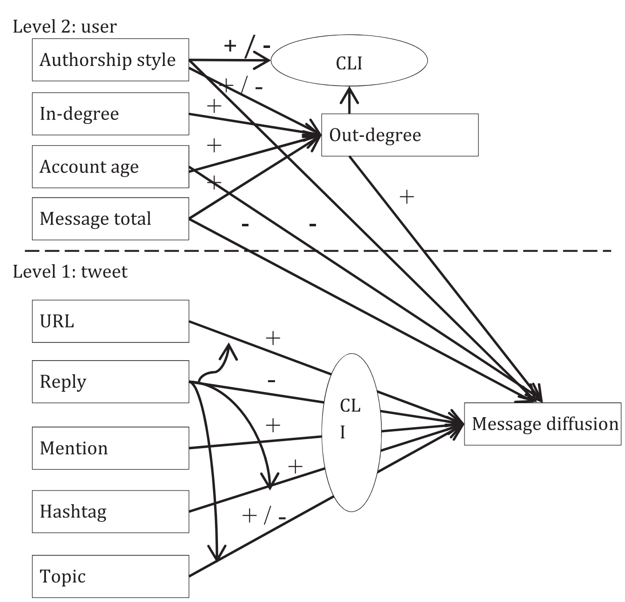


Figure 2. Model as observed.

In addition to the moderating effects, findings support a moderate mediating effect between followers, account age and in-degree (friends). When not controlling for the effect of followers, both these variables get higher coefficient estimates. In addition, a separate general linear analysis showed the experience and in-degree variables to be significant predictors of followers. Together, this conforms to Baron and Kenny's statistical operationalisation of mediating variables. Figure 2 summarises the observed results.

## 5. Conclusions and discussion

This study aimed to explain why some police-authored tweets are forwarded whereas others are not. Our research question was: how can differences in the retweet rate of messages authored by Dutch police officers be explained? The results show that both the user and the message matter, as tweet and user characteristics, influence the chances of being retweeted. Differences in retweet probability for diverse topics were found, suggesting that the type of information provided is important to an audience (as suggested by Marwick and Boyd 2010). In addition, the inclusion of web addresses, hashtags and mentions was significantly related to retweet probabilities. This suggests that other users prefer messages which include more content, perhaps because these are less dependent on – or more clearly tied to – outside context for interpretation (in line with Zappavigna 2011). Interaction with the audience by way of replies was also related to message diffusion, which emphasises the effect of engagement. Various user-level characteristics such as audience size, previous engagement and authorship style were of direct influence on retweet probabilities. In addition, user-level characteristics significantly influence the effect of tweet-level characteristics such as topics discussed, elements included and interactivity.

A first contribution to the literature on social media in government concerns the *interaction effects between message and sender*. In terms of tweet-level characteristics, our findings deviate from the model of Suh et al. (2010). Mentions were found to have a significant positive effect on retweet probabilities in both statistical and practical terms. The same is true for replies, which may decrease retweet probabilities, but diminish as followers increase. In addition to these direct effects, this study presents supporting evidence of interaction effects on replies, URL inclusion, various topics and mentions. Such interaction effects have remained largely untested in the message diffusion literature dealing with micro-blog use. Parts of the multi-level model used in this paper may be of interest as components of an alternative conceptualisation of social capital, a burgeoning field of interest in social media research (see for example Ellison, Steinfield, and Lampe 2011; Vitak and Ellison 2012).

A second contribution to the literature concerns the identification of *message content* as a key factor. Drawing from descriptive analyses by Crump (2011) and Heverin and Zach (2010), tweets were coded for topics often discussed by police officials. This study used these topics in a quantitative analysis in order to predict retweets and found supporting evidence to their effect. The strongest effects are missing person reports, tweets with URLs and wanted messages stimulating message forwarding. The positive interaction between personal messages and URLs and high message counts and mentions also increase the odds of retweets. Small talk has the strongest diminishing effect.

A third addition to the literature concerns the *authorship style*. The analysis of this factor builds upon qualitative and descriptive studies concerning audience engagement on the platform (Kwak, Chun, and Moon 2011; Zhao, Zeng, and Zhong 2011). By aggregating the use of specific tweet characteristics employed by a user, this study sought to quantitatively operationalise authorship style. Findings support the hypothesis that differences in authorship style significantly impact retweet probabilities. Apart from direct effects, authorship has been observed to interact with the diffusion probability of specific topics, as with URL inclusion. This adds another dimension to predictive models of message diffusion, based on the distinction between tweet- and user-level characteristics.

For the police, and possibly government organisations at large, this study has generated some applicable insights to increase message diffusion as a means to foster public safety through public participation and civic engagement (Valenzuela, Park, and Kee 2009; Meijer 2013). Findings show which message characteristics can be manipulated to increase the probability of being retweeted. Tweet characteristics to maximise are, in order of effect size, send replies with URLs, include URLs, use mentions to show you are socially engaged, include hashtags to increase searchability, write longer tweets and send tweet in the afternoon or evening when more people listen. In terms of accounts, the recommendations are less straightforward. The police, and other government organisations, should note that having more followers is better but reduces the effect of replies and mentions, older accounts have less chance of getting retweets unless there are enough followers, having posted a lot reduces the chances of being retweeted and adopting a personal style increases chances for messages containing a URL but decreases retweet probability for those without one. Finally, sending original tweets rather than avoiding to retweet others can be expected to result in a larger number of retweets. In order to analyse whether and if so by means of which exact mechanisms retweets may eventually foster public safety, future research should be targeted at developing a better understanding of the intentions and behaviours of Twitter users, both senders (police officials in this case) and recipients (citizens). Here, qualitative research in the form of in-depth qualitative interviews and observations may complement the quantitative approaches demonstrated in this paper.

However, there are also interesting avenues for new quantitative approaches to the study of message diffusion. Further analysis of user-aggregated data may yield interesting results. The outcomes of this study seem to indicate that message diffusion depends on the use of information queues coupled with audience expectations, indicating a model of tweet match with audience expectations in the context of specific accounts or account typologies. This model would explain the findings about topics and authorship-style interactions, as well as changes between mention-effects for low and high follower accounts. A user-dependent theory of audience expectations may explain why accounts with more original content rather than a preselection of important content are less successful: these accounts may be more individual-to-individual communication rather than providing a news gatekeeper function expected by followers of officials. Such analyses may also uncover learning effects, such as diminished effects of URLs for users which often employ this feature. Understanding of authorship style can perhaps be enhanced with the application of sentiment analysis (Barbosa and Feng 2010), social network theory (Butts 2008) and improved topic detection (Cataldi, Di Caro, and Schifanella 2010). In addition, the role of authorship may differ by cultural setting, as research points out international differences in Twitter use (Poblete et al. 2011) and e-governance in general (Zhao 2011). The sample used in this research was limited to the Dutch policing context. Other research (for example Ellison, Steinfield, and Lampe 2011) suggests that contextual factors shape audience expectations and diffusion patterns. Finally, this study has observed the relation of authorship and retweet probability in the context of police officials. Further research is required to understand the extent to which authorship is applicable to message diffusion in general. Besides additional research into user-level characteristics, increased specification of topics may be beneficial. The rudimentary coding employed by this study suggests that topics can be successfully applied in predictive research (see for instance Shamma, Kennedy, and Churchill 2009), but are applicable to limited and highly specific contexts. Besides testing the influence of topics for police officials of different nationalities, an improved method of topical coding might yield additional insights into this phenomenon. Although vertical coding, such as that employed in this study, are supported as viable options for well-defined populations, more general approaches may be developed to understand general topics of interest. Some studies aimed at methodological development show promising inroads into broader application of topic-based research (Michelson and Macskassy 2010; Yang et al. 2011).

This study has presented some in-depth insights into the new social media communication patterns between government and citizens. The findings show that citizens value *interactions* (through mentions), they value messages with much *content* (in terms of URLs and relevant subjects), they value *long-term commitment* (in terms of the age of accounts and the total number of messages posted) and they value a *personal style* (and original tweets). This shows that there is no quick fix to reaching a large number of citizens through social media: it takes hard work, perseverance and sound knowledge of the subjects that are of interest to citizens.

## Notes

1. 'User' is used synonymously with 'account', as accounts managed by more than one person are still perceived under one name, as one 'source'. This conforms to the literature on the organisational use of Twitter accounts, which takes the organisation as the user. Note that most accounts in this study are tended to by no more than one person.
2. 'aangehouden', 'inbraak', 'onderzoek'.
3. 'overleg', 'project', 'spreekuur'.
4. 'campagne', 'acties', 'voorlichting'.
5. 'verkeer', 'snelheid', 'thv' (ter hoogte van).
6. 'tips', 'voorkomen', 'waarschuwing'
7. '0900-8844', 'getuige?', 'iets gezien?'.
8. 'vermist', 'vermiste', 'laatst gezien'.
9. ';-)', 'leuk', 'mooi'. For the complete wordlists used for each topic, consult Appendix 1.
10. This approach was deemed most feasible considering the amount of data (130,000+ tweets) and vertical nature of semantic entities (all drawn from police-authored communication). The latter provides less ambiguity and thus increased reliability for simple coding.
11. Due to the rudimentary approach of this method, between 27.5% and 40.5% of the messages are not attributed to the right topic and thus default to 'generic' tweets.
12. These are the accounts followed by a user, contrary to following the user.
13. As an example: mentions often denote conversation (Honeycutt and Herring 2009), users with a high mention average can therefore be characterised as more interaction oriented relative to those who do not use mentions. By examining such signals across multiple variables, the reliability of stylistic distinctions is increased.
14. Out of 1000 accounts listed, 36 accounts were deleted because they appeared to be deleted or shielded in the mean time, making the effective sample size 964.
15. Although not a hard requirement (Harrell 2001), non-normal distributions may inflate coefficients.
16. Although our software noted false convergence, the multicolinearity check produced no problems. In addition, neither coefficients nor standard errors are visibly inflated (Hosmer and Lemeshow 2000), nor do iteration reports show deviance between iterations. Finally, the R packages applied is known to be stringent with estimation procedures, which often yields false positives on convergence checks (Bates 2009). In such cases, deviance is overestimated, making the results more conservative than optimal, but equally reliable.
17. The other category of functions includes 'animal cops' and 'loverboy team cops'.
18. Both follower and friend counts were normalised to compensate for their skewed distribution (see Section 3).
19. Morning is 06:00–11:59:59, afternoon is 12:00–17:59:59, evening is 18:00–23:59:59 and night is between 0:00 and 05:59:59.
20. Other effects are department (−0.73%), small talk (−0.51%), prevention (+0.39%), crime or incident (+0.16%) and traffic (+0.05%, $\rho < 0.05$).

## References


Ampofo, L., N. Anstead, and B. O'Loughlin. 2011. "Trust, Confidence, and Credibility: Citizen Responses on Twitter to Opinion Polls During the 2010 UK General Election." *Information, Communication & Society* 14 (6): 850–871.

Asur, S., and B. A. Huberman. 2010. "Predicting the Future with Social Media." arXiv:1003.5699. http://arxiv.org/abs/1003.5699.

Bae, Y., and H. Lee. 2012. "Sentiment Analysis of Twitter Audiences: Measuring the Positive or Negative Influence of Popular Twitterers." *Journal of the American Society for Information Science and Technology* 63 (12): 2521–2535.

Barbosa, L., and J. Feng. 2010. "Robust Sentiment Detection on Twitter from Biased and Noisy Data." In *Proceedings of the 23rd International Conference on Computational Linguistics: Posters*, 36–44. Stroudsburg, PA: Association for Computational Linguistics.

Baron, R. M., and D. A. Kenny. 1986. "The Moderator–Mediator Variable Distinction in Social Psychological Research: Conceptual, Strategic, and Statistical Considerations." *Journal of Personality and Social Psychology* 51 (6): 1173–1182.

Bates, D. 2009. Online Response to Convergence Issues in CRAN R LME4 Package. http://permalink.gmane.org/gmane.comp.lang.r.lme4.devel/2956

Bickel, R. 2007. *Multilevel Analysis for Applied Research*. New York: Guilford Press.

Bliese, P. 2012. *Multilevel Modeling in R: A Brief Introduction to R, the Multilevel Package and the nlme Package (2.4)*. http://cran.r-project.org/doc/contrib/Bliese_Multilevel.pdf

Buchanan, Elizabeth, A., and Zimmer, Michael. 2013 "Internet Research Ethics." In *The Stanford Encyclopedia of Philosophy (Fall 2013 Edition)*, edited by Edward N. Zalta. http://plato.stanford.edu/archives/fall2013/entries/ethics-internet-research/.

Burton, S., and A. Soboleva. 2011. "Interactive or Reactive? Marketing with Twitter." *Journal of Consumer Marketing* 28 (7): 491–499.

Butts, C. T. 2008. "Social Network Analysis: A Methodological Introduction." *Asian Journal of Social Psychology* 11 (1): 13–41.

Cataldi, M., L. Di Caro, and C. Schifanella. 2010. "Emerging Topic Detection on Twitter Based on Temporal and Social Terms Evaluation." In *Proceedings of the Tenth International Workshop on Multimedia Data Mining*, 4. New York: ACM.

Chen, G. M. 2011. "Tweet This: A Uses and Gratifications Perspective on How Active Twitter use Gratifies a Need to Connect with Others." *Computers in Human Behavior* 27 (2): 755–762.

Chew, C., and G. Eysenbach. 2010. "Pandemics in the Age of Twitter: Content Analysis of Tweets During the 2009 H1N1 Outbreak." *PloS One* 5 (11): e14118.

Cho, S. E., and H. W. Park. 2011. "Government Organizations' Innovative Use of the Internet: The Case of the Twitter Activity of South Korea's Ministry for Food, Agriculture, Forestry and Fisheries." *Scientometrics* 90 (1): 9–23.

Christensen, T., and P. Lægreid. 2011. "Democracy and Administrative Policy: Contrasting Elements of New Public Management (NPM) and Post-NPM." *European Political Science Review* 3 (1): 125–146.

Crowston, K., X. Liu, E. Allen, and R. Heckman. 2010. "Machine Learning and Rule-Based Automated Coding of Qualitative Data." ASIST 2010, Pittsburgh, PA, October 22–27.

Crump, J. 2011. "What Are the Police Doing on Twitter? Social Media, the Police and the Public." *Policy & Internet* 3 (4): 1–27.

Ellison, N. B., C. Steinfield, and C. Lampe. 2011. "Connection Strategies: Social Capital Implications of Facebook-Enabled Communication Practices." *New Media Society* 13 (6): 873–892.

Finin, T., and B. Tseng. 2007. "Why We Twitter: Understanding Microblogging." Joint 9th WEBKDD workshop and 1st SNA-KDD Workshop '07. 12 August 2007, California: ACM.

Garson, D. 2009. "Quantitative Research in Public Administration." http://faculty.chass.ncsu.edu/garson/PA765/pa765syl.htm.

Gelman, A., and J. Hill. 2007. *Data Analysis using Regression and Multilevel/Hierarchical Models*. Cambridge: Cambridge University Press.

Golbeck, J., J. M. Grimes, and A. Rogers. 2010. "Twitter use by the US Congress." *Journal of the American Society for Information Science and Technology* 61 (8): 1612–1621.

Hansen, L. K., A. Arvidsson, F. A. Nielsen, E. Colleoni, M. Etter. 2011. "Good Friends, Bad News – Affect and Virality in Twitter." In *6th International Conference, FutureTech 2011, Loutraki, Greece, June 28–30*, Vol. 2, edited by J. J. Park, L. T. Yang, and C. Lee, 34–43. Berlin: Springer-Verlag. doi:10.1007/978-3-642-22309-9_5

Harrell, F. E. 2001. *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer.

Heverin, T., and L. Zach. 2010. "Twitter for City Police Department Information Sharing." *Journal of the American Society for Information* 47 (1): 1–7.

Honey, C., and S. C. Herring. 2009. "Beyond Microblogging: Conversation and Collaboration via Twitter." In *System Sciences, 2009. HICSS'09. 42nd Hawaii International Conference*, 1–10. Big Island: IEEE.

Hosmer, D. W., and S. Lemeshow. 2000. *Applied Logistic Regression.* Hoboken, NJ: John Wiley & Sons.

Khan, H. R., and J. E. H. Shaw. 2011. "Multilevel Logistic Regression Analysis Applied to Binary Contraceptive Prevalence Data." *Journal of Data Sciences* 9 (1): 93–110.

Khrabrov, A. 2010. "Discovering Influence in Communication Networks Using Dynamic Graph Analysis." Proceedings of the 2010 IEEE Second International Conference on Social Computing (SOCIALCOM '10), Minneapolis, MN, August 20–22.

Kim, M., and H. W. Park. 2012. "Measuring Twitter-Based Political Participation and Deliberation in the South Korean Context by Using Social Network and Triple Helix Indicators." *Scientometrics* 90 (1) (January 2012): 121–140. http://dx.doi.org/10.1007/s11192-011-0508-5

Kulshrestha, J., F. Kooti, A. Nikravesh, and P. K. Gummadi. 2012. "Geographic Dissection of the Twitter Network." In *Proceedings of the Sixth International AAAI Conference on Weblogs and Social Media*. Dublin: AAAI. http://www.aaai.org/ocs/index.php/ICWSM/ICWSM12/paper/view/4685

Kwak, H., H. Chun, and S. Moon. 2011. "Fragile Online Relationship: A First Look at Unfollow Dynamics in Twitter." In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 1091–1100. New York: ACM.

Lassen, D. S., and A. R. Brown. 2011. "Twitter: The Electoral Connection?" *Social Science Computer Review* 29 (4): 419–436.

Lattin, J., J. D. Carroll, and P. E. Green. 2003. *Analyzing Multivariate Data*. Belmont, CA: Brooks/Cole.

Li, L. T., S. Yang, A. Kavanaugh, E. A. Fox, S. D. Sheetz, D. Shoemaker, and T. Whalen. 2011. "Twitter Use During an Emergency Event: The Case of the UT Austin Shooting Categories and Subject Descriptors." In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, 335–336. New York: ACM.

Lussier, J. T., and N. V. Chawla. 2011. "Network Effects on Tweeting." In *Discovery Science: 14th International Conference, DS 2011, Espoo, Finland, October 5–7*, edited by T. Elomaa, J. Hollmén, and H. Mannila, 209–220. Berlin: Springer-Verlag. doi:10.1007/978-3-642-24477-3_18

Marwick, a. E., and D. Boyd. 2010. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13 (1): 114–133.

Meijer, A. 2013. "New Media and the Coproduction of Safety: An Empirical Analysis of Dutch Practices." *American Review of Public Administration*. doi:10.1177/0275074012455843

Meijer, A. J., and M. Thaens. 2013. "Social Media Strategies: An Empirical Analysis of Emerging Social Media Strategies of North American Police Departments." *Government Information Quarterly* 30 (4): 343–350.

Meijer, A., S. Grimmelijkhuijsen, D. Fictorie, and A. Bos. 2011. "Politie en Twitter: Coproductie en Community Policing in Het Informatietijdperk." *Bestuurskunde* 3 (3): 1–23.

Mergel, I. 2012. "The Social Media Innovation Challenge in the Public Sector." *Information Polity* 17 (3): 281–292. doi:10.3233/IP-2012-000281

Michelson, M., and S. A. Macskassy. 2010. "Discovering Users' Topics of Interest on Twitter: A First Look." In *Proceedings of the Fourth Workshop on Analytics for Noisy Unstructured Text Data*, 73–80. New York: ACM.

Pampel, F. C. 2000. *Logistic Regression: A Primer*. Thousand Oaks, CA: Sage.

Picazo-Vela, S., I. Gutiérrez-Martinez, and L. F. Luna-Reyes. 2011. "Social Media in the Public Sector: Perceived Benefits, Costs and Strategic Alternatives." In *Proceedings of the 12th Annual International Digital Government Research Conference: Digital Government Innovation in Challenging Times*, 198–203. New York: ACM.

Poblete, B., R. Garcia, M. Mendoza, and A. Jaimes. 2011. "Do All Birds Tweet the Same?: Characterizing Twitter Around the World." In *Proceedings of the 20th ACM International Conference on Information and Knowledge Management*, 1025–1030. New York: ACM.

Procter, R., J. Crump, S. Karstedt, A. Voss, and M. Cantijoch. 2013. "Reading the Riots: What Were the Police Doing on Twitter?" *Policing and Society: An International Journal of Research and Policy* 23 (4): 413–436. doi:10.1080/10439463.2013.780223

Reinberg, C. 2009. "Are Tweets Copyright-Protected?" http://www.wipo.int/wipo_magazine/en/2009/04/article_0005.html

Rogers, R. 2013. "Debanalizing Twitter: the Transformation of an Object of Study." In *Proceedings of the 5th Annual ACM Web Science Conference* (*WebSci '13*), 356–365. New York, NY: ACM. http://doi.acm.org/10.1145/2464464.2464511

Rojas, J. M., C. J. Ruiz, and C. Farfán. 2011. "E-Participation and E-Governance at Web 2.0 in Local Governments of Colombia." In *Proceedings of the 5th International Conference on Theory and Practice of Electronic Governance*, 301–304. New York: ACM.

Shamma, D. A., L. Kennedy, and E. F. Churchill. 2009. "Tweet the Debates: Understanding Community Annotation of Uncollected Sources." In *Proceedings of the First SIGMM Workshop on Social Media (WSM '09)*, 3–10. New York, NY: ACM. http://doi.acm.org/10.1145/1631144.1631148

Stieglitz, S., and L. Dang-Xuan. 2013. "Social Media and Political Communication: a Social Media Analytics Framework." *Social Network Analysis and Mining* 3 (4): 1277–1291.

Suh, B., L. Hong, P. Pirolli, and E. H. Chi. 2010. "Want to be Retweeted? Large Scale Analytics on Factors Impacting Retweet in Twitter Network." In *2010 IEEE Second International Conference on Social Computing*, 177–184. IEEE.

Valenzuela, S., N. Park, and K. F. Kee. 2009. "Is There Social Capita in a Social Network Site? Facebook Use and College Students' Life Satisfaction, Trust, and Participation." *Journal of Computer-Mediatied Communication* 14 (4): 875–901. doi:10.1111/j.1083-6101.2009.01474.x

Vitak, J., N. B. Ellison. 2012. "'There's a Network Out there you Might as Well Tap': Exploring the Benefits of and Barriers to Exchanging Informational and Support-Based Resources on Facebook." *New Media Society*. doi:10.1177/1461444812451566

Vergeer, M., L. Hermans, and S. Sams. 2011. "Is the Voter only a Tweet Away? Micro Blogging During the 2009 European Parliament Election Campaign in The Netherlands." *First Monday* 16 (8). http://dx.doi.org/10.5210/fm.v16i8.3540

Wang, H.-C., and H.-S. Doong. 2010. "Does Government Effort or Citizen Word-of-Mouth Determine e-Government Service Diffusion?" *Behaviour & Information Technology* 29 (4): 415–422.

Waters, R. D., and J. Y. Jamal. 2011. "Tweet, Tweet, Tweet: A Content Analysis of Nonprofit Organizations' Twitter Updates." *Public Relations Review* 37 (3): 321–324.

Waters, R. D., and J. M. Williams. 2011. "Academic Paper Squawking, Tweeting, Cooing, and Hooting: Analyzing the Communication Patterns of Government Agencies on Twitter." *Journal of Public Affairs* 11 (4): 353–363.

Weng, J., and B. Lee. 2010. "Event Detection in Twitter." Proceedings of the Fifth International AAAI Conference on Weblogs and Social Media, Barcelona, Spain, July 17–21.

Weng, J., E. P. Lim, J. Jiang, and Q. He. 2010. "Twitterrank: Finding Topic-Sensitive Influential Twitterers." In *Proceedings of the Third ACM International Conference on Web Search and Data Mining*, 261–270. New York: ACM.

Whisman, M. A., and G. H. McClelland. 2005. "Designing, Testing, and Interpreting Interactions and Moderator Effects in Family Research." *Journal of Family Psychology* 19 (1): 111.

Wigley, S., and B. K. Lewis. 2012. "Rules of Engagement: Practice what you Tweet." *Public Relations Review* 38 (1): 165–167.

Yang, H., S. Chen, M. R. Lyu, and I. King. 2011. "Location-Based Topic Evolution." In *Proceedings of the 1st International Workshop on Mobile Location-Based Service*, 89–98. New York: ACM.

Zappavigna, M. 2011. "Ambient Affiliation: A Linguistic Perspective on Twitter." *New Media & Society* 13 (5): 788–806.

Zhang, M., B. J. Jansen, and A. Chowdhury. 2011. "Business Engagement on Twitter: a Path Analysis." *Electronic Markets* 21 (3): 161–175.

Zhao, F. 2011. "Impact of National Culture on e-Government Development: A Global Study." *Internet Research* 21 (3): 362–380. doi:10.1108/10662241111139354

Zhao, L., Y. Zeng, and N. Zhong. 2011. "A Weighted Multi-Factor Algorithm for Microblog Search." In *Active Media Technology*, 153–161. Heidelberg: Springer.

**Appendix 1. List of subcodes used for various message types**

| wanted | | department | | missing | | smalltalk | | crime/Incident | |
|---|---|---|---|---|---|---|---|---|---|
| 2298 | zoekt | 2934 | overleg | 2200 | vermist | 0 | zo ga hierna | 99 | aanhouden |
| 1848 | bel 0900-8844 | 139 | discussie | 877 | vermiste | 4 | nice! | 5953 | aangehouden |
| 264 | bel 09008844 | 324 | buro | 59 | laatst gezien | 21 | prima! | 4 | houdt aan |
| 1081 | situatie | 393 | opleiding | | | 235 | heerlijk | 376 | vlucht |
| 3294 | getuige? | 19 | excuses voor | | | 1726 | leuk | 2390 | gestolen |
| 1667 | iets gezien | 1 | manual | | *Event* | 833 | lekker | 432 | opgepakt |
| 467 | opsporing verzocht | 1 | gratis hulp | 32 | deelname | 61 | genoten | 162 | opengebroken |
| 0 | ##### | 3 | beslis mee | 1100 | campagne | 604 | vrij | 324 | opgelost |
| 55 | SMS-Alert | 60 | procedure | 206 | acties | 203 | einde dienst | 66 | gearresteerd |
| 36 | Allert | 406 | project | 333 | evenement | 2 | begin dienst | 324 | opgelost |
| 309 | help | 150 | cijfers | 326 | opstelten | 311 | bang | 19 | veroordeling |
| 258 | signalement | 48 | assisteren | 14 | actie voeren | 143 | goeie | 1008 | rijbewijs |
| 28 | weet meer | 3 | diensten zijn druk | 429 | voorlichting | 14 | oh ja | 312 | arresteer |
| 116 | bellen met | 3389 | verleg | 104 | opkomst | 1 | meningen | 83 | opstoot |
| 1371 | gezocht | 77 | aftrap | 21 | live tijdens | 1040 | (winks) | 2390 | gestolen |
| 779 | alert | 1035 | spreekuur | 22 | EK | 516 | :-) | 973 | drug |
| 23 | getuige gezocht | 30 | extra inzet | 22 | WK | 116 | vriendin | 24 | explosieven |
| 110 | herkent u | 9 | collegas in dienst | 725 | actie | 8 | success | 132 | opgerold |
| 66 | herken je | 84 | teamchef | 55 | demonstratie | 19 | tot ziens | 1258 | mishandeling |
| 0 | #### | 72 | praten over | | | 0 | wrijf het maar in | 1241 | ongeval |
| 8 | informatie is welkom | 662 | briefing | | | 1017 | fijn | 1010 | brand |
| 18 | vluchtauto | 20 | om de tafel | | | 27 | gecertificeerd | 811 | anhoudingen |
| 180 | verdachts gezien | 71 | focus | | | 747 | dagdienst | 7 | ontruimt |
| 26 | bel of mail | | | | | 2372 | mooi | 44 | explosief |
| 10 | uit kijken naar | *Prevention* | | | | 1124 | aanwezig | 3074 | onderzoek |
| 229 | wie herkent | 55 | bedenk | | | 205 | zometeen | 209 | verklaring |
| 161 | verdachte personen | 79 | folder | | | 1800 | rustig | 28 | aan te houden |
| 27 | verdacht persoon | 2 | empty | | | 148 | enthousiast | 2112 | anhouding |
| | | 1105 | waarschuw | | | 64 | ha | 2151 | melding |
| | *traffic* | 17 | riskeer | | | 365 | late dienst | 652 | overvallen |
| 77 | fietscontrole | 32 | houdt rekening | | | 183 | prachtig | 3167 | overval |
| 2 | aanrijding plaatsgevonden | 36 | oppassen | | | 945 | de wijk in | 80 | ontruimd |
| 2826 | verkeer | 168 | denk aan | | | 154 | toetsen | 2154 | aangetroffen |
| 70 | radarcontrole | 119 | opvallende | | | 72 | weet ik | 848 | ingebroken |
| 293 | a6 | 1127 | toezicht | | | 349 | geniet | 21 | escalatie |
| 337 | thv | 168 | NIET! | | | 4 | briefings | 10 | pakt op |
| 103 | a1 | 38 | géén | | | 20 | goed om te zien | 10 | opvarende |
| 95 | a2 | 8 | ga niet in op | | | 593 | ik heb | 20 | betrapte |
| 33 | a3 | 240 | goed op | | | 386 | jammer | 324 | opgelost |
| 120 | a4 | 489 | preventie | | | 538 | bezig met | 15 | geruime tijd |
| 19 | a5 | 498 | vooral | | | 1499 | bezoek | 1656 | nieuws: |
| 75 | a6 | 5 | voordoen als | | | 14 | je bent welkom | 34 | omgekomen |
| 63 | a7 | 0 | voordoet als | | | 1040 | gezellig | 135 | geruimd |
| 22 | a8 | 18 | zegt het voort | | | 359 | gefeliciteerd | 2767 | overlast |
| 43 | a9 | 3 | pas ook op | | | 271 | trots | 112 | overgedragen |
| 3 | u mag hier | 21 | reken op | | | 310 | koffie | 3 | niet in gevaar |
| 1 | je mag hier | 88 | levensgevaarlijk | | | 67 | voelt | 1356 | hennep |
| 16 | aanrijding geweest | 156 | gevaarlijke | | | 25 | #loesje | 3 | achtervolgen |
| 8 | total loss | 3 | verassen | | | 56 | zonde | 847 | slachtoffer |
| 23 | busbaan | 38 | niet toegestaan | | | 199 | afscheid | 5012 | inbraak |
| 110 | aanrijding: | 0 | wordt hier vaak | | | 56 | drukke dienst | 591 | gepleegd |
| 1329 | snelheid | 972 | voorkom | | | 91 | kennis gemaakt | 2332 | inbraken |
| 10 | doorstroming | 190 | meld misdaad anoniem | | | 499 | surveillance | 284 | vechtpartij |
| 33 | dichte mist | 33 | fraudealert | | | 529 | :) | 24 | wiet |

**Appendix 2. Best linear unbiased estimator model**

| AIC | BIC | logLik | deviance | | |
|---|---|---|---|---|---|
| 97,624 | 98,093 | −48,763 | 97,526 | | |
| | *Random effects*: | | | | |
| Groups | Name | Variance | SD | | |
| name | (Intercept) | 0.024146 | 0.15539 | | |
| Number of obs: 106,462, groups: (name), 964 | | | | | |
| | *Fixed effects:* | | | | |
| | Estimate | Standard error | $z$ value | $Pr(>\|z\|)$ | |
| (Intercept) | −7.93E+00 | 2.45E−01 | −32.37 | < 2e−16 | *** |
| | *User-level effects* | | | | |
| log(stats + 1) | −1.95E−01 | 6.54E−02 | −2.98 | 0.0029 | ** |
| log(friends + 1) | 6.74E−02 | 5.36E−02 | 1.26 | 0.209068 | |
| log(followers + 1) | 1.43E+00 | 5.01E−02 | 28.47 | < 2e−16 | *** |
| Personal style | −2.06E−01 | 2.29E−02 | −8.97 | < 2e−16 | *** |
| Original authorship | −3.64E−01 | 7.21E−02 | −5.06 | 4.27E−07 | *** |
| Days active | −3.29E−03 | 2.73E−04 | −12.04 | < 2e−16 | *** |
| Message-level effects | | | | | |
| Time of day | 4.81E−02 | 9.65E−03 | 4.98 | 6.24E−07 | *** |
| Characters in tweet | 9.58E−03 | 3.09E−04 | 31.04 | < 2e−16 | *** |
| Hashtag # | 1.98E−01 | 1.75E−02 | 11.31 | < 2e−16 | *** |
| Reply (y/n) | 1.29E+00 | 3.08E−01 | 4.2 | 2.70E−05 | *** |
| Url (y/n) | 1.29E+00 | 1.57E−01 | 8.21 | < 2e−16 | *** |
| Crime/incident (CI) | 1.23E−01 | 1.89E−02 | 6.53 | 6.73E−11 | *** |
| Departmental (DP) | −8.72E−01 | 4.65E−02 | −18.76 | < 2e−16 | *** |
| Traffic (TR) | 6.49E−01 | 2.70E−01 | 2.41 | 0.01616 | * |
| Prevention (PR) | 2.74E−01 | 3.91E−02 | 6.99 | 2.69E−12 | *** |
| Look out for (LOF) | 8.22E−01 | 2.86E−02 | 28.78 | < 2e−16 | *** |
| Missing person (MIS) | 1.26E+00 | 9.18E−02 | 13.73 | < 2e−16 | *** |
| Small talk (ST) | −5.23E−01 | 2.69E−02 | −19.46 | < 2e−16 | *** |
| Mentions | 1.26E+00 | 2.64E−01 | 4.77 | 1.82E−06 | *** |
| | *Interaction effects* | | | | |
| I(log(stats + 1)* log(stats + 1)) | −3.33E−02 | 6.42E−03 | −5.19 | 2.14E−07 | *** |
| I(Hash*Hash) | −2.28E−02 | 2.48E−03 | −9.2 | < 2e−16 | *** |
| log(stats + 1):mention | 2.40E−01 | 3.63E−02 | 6.62 | 3.48E−11 | *** |
| log(followers + 1):mention | −4.40E−01 | 5.20E−02 | −8.46 | < 2e−16 | *** |
| DP:mention | 5.48E−01 | 1.46E−01 | 3.75 | 0.000176 | *** |
| Original authorship:mention | 1.05E−01 | 2.95E−02 | 3.54 | 0.0004 | *** |
| log(stats + 1):log(friends + 1) | 4.02E−02 | 9.11E−03 | 4.42 | 9.96E−06 | *** |
| Reply:log(stats + 1) | 2.70E−01 | 4.91E−02 | 5.5 | 3.75E−08 | *** |
| Url:log(stats + 1) | 1.53E−01 | 2.44E−02 | 6.26 | 3.76E−10 | *** |
| log(*stats*+1):TR | 1.61E−01 | 3.87E−02 | 4.17 | 3.01E−05 | *** |
| log(friends + 1):log(followers + 1) | −5.42E−02 | 1.04E−02 | −5.19 | 2.07E−07 | *** |
| Reply:log(followers + 1) | −9.14E−01 | 5.72E−02 | −15.99 | < 2e−16 | *** |
| Url:log(followers + 1) | −2.73E−01 | 3.37E−02 | −8.12 | 4.58E−16 | *** |
| log(followers + 1): original authorship | 3.39E−02 | 1.15E−02 | 2.95 | 0.003199 | ** |
| Dayscale: personal style | −4.35E−02 | 1.13E−02 | −3.85 | 0.000119 | *** |
| Characters:TR | −8.04E−03 | 1.44E−03 | −5.57 | 2.50E−08 | *** |
| Hash:Reply | 1.46E−01 | 2.76E−02 | 5.29 | 1.22E−07 | *** |
| Reply:Url | 6.10E−01 | 1.22E−01 | 5.01 | 5.36E−07 | *** |
| Reply:LOF | −7.74E−01 | 2.06E−01 | −3.77 | 0.000165 | *** |
| Reply:ST | 5.90E−01 | 9.96E−02 | 5.93 | 3.09E−09 | *** |
| Url:TR | −5.35E−01 | 9.32E−02 | −5.74 | 9.19E−09 | *** |
| Url: personal style | 5.35E−01 | 2.33E−02 | 22.99 | < 2e−16 | *** |
| CI: personal style | 1.53E−01 | 1.71E−02 | 8.94 | < 2e−16 | *** |
| DP:ST | 4.02E−01 | 1.11E−01 | 3.62 | 0.000296 | *** |
| PR:LOF | −6.11E−01 | 1.44E−01 | −4.25 | 2.18E−05 | *** |
| Reply:days active | 7.18E−04 | 1.81E−04 | 3.96 | 7.63E−05 | *** |
| log(followers + 1):days | 3.60E−04 | 4.14E−05 | 8.69 | < 2e−16 | *** |
| TR: days active | −5.13E−04 | 1.50E−04 | −3.42 | 0.000628 | *** |

Significant codes: $^{*}p \leq 0.05$, $^{**}p \leq 0.01$, $^{***}p \leq 0.001$.