

Social Media Monitoring for Crisis Communication: Process, Methods and Trends in the Scientific Literature

Aino Ruggiero, University of Jyväskylä, Finland

Marita Vos, University of Jyväskylä, Finland

Abstract

This literature review study aims at clarifying current knowledge on social media monitoring from the perspective of organizational communication and public relations. It also contributes to crisis communication by shedding light on how fast developing social media discourse can be followed and analysed in order to understand citizens' needs throughout all the phases of a crisis. The findings of this study reveal a number of insights in the scientific literature on the concept of monitoring, the monitoring process, methods, tools and solutions, methodological issues and trends covering the years 2009–2012. In the literature, social media monitoring is described as a process which comprises various steps: preparation, data collection, data analysis and reporting. The methodological issues discussed in the literature show the many challenges still to be addressed. Knowledge connecting the results of monitoring activities to communication strategy making is lacking, indicating a gap to be filled in future studies.

Keywords: social media, monitoring, crisis communication, public relations

Introduction

This paper aims at clarifying current knowledge on the monitoring of citizen interaction in social media. Specifically, the aim is to find out how the concept of monitoring, the process, monitoring methods, tools and solutions, and methodological issues are described and to identify trends in the recent scientific literature. This objective is pursued by undertaking a literature review. The underlying purpose is to contribute to crisis communication in the case of emergencies and disasters, including incidents of terrorism, by explaining how fast developing social media discourse can be followed and analysed. Monitoring of the online environment, listening to citizen interaction in the social media, is considered a first step in communication strategy-making and online participation by various organizations in the response network throughout a crisis (e.g. Reynolds & Seeger, 2005).

Social media form a group of diverse and fast developing communication media which enable the production and sharing of information through collaborative means (Sutton, 2009). The role of social media for crisis communication has been acknowledged in recent years with respect to e.g. the issuing of emergency warnings and alerts, receiving requests for assistance, supporting recovery efforts, and provision of situational awareness and real-time information through monitoring (Lindsay, 2010).

The importance of gaining fast insights into public perceptions and information needs that may be large in volume and also incorporate conflicting information becomes evident in crises which are unpredictable, unknown and fear-inducing (Sheppard, 2011). Due to the uncertainty and ambiguity involved, as in the anthrax crisis of 2001, there may be need to adjust information (Robinson & Newstetter, 2003), fine-tune communication and improve response (Rubin, Amlôt, & Page, 2011) as the crisis evolves. This calls for knowledge on how to monitor publics' perceptions and the debate and manage the communication process, all in real time.

This paper comprises four sections. Following the introduction, the method used, including the research questions, sample and data analysis, is described. The third section presents the findings of this study, and section four concludes.

Method

This study aims at clarifying monitoring of citizen interaction in social media. In particular, the focus is on describing the concept, process, methods, tools and solutions, and methodological issues, and on identifying trends in current scientific literature. The following research question was set:

How are the methods for monitoring social media interaction of citizens described in the academic literature?

The study contributes to a project identifying factors that need to be taken into consideration when monitoring the social media interaction of citizens in crisis situations in order to understand citizens' needs throughout all the phases of a crisis. As social media are a new research field with a limited volume of published research, this study takes a broader approach and reveals insights on social media monitoring as a basis for communication with publics.

Sample

The present study was conducted following the protocol of a systematic literature review (e.g. Jesson, Lacey, & Matheson, 2011). The search was conducted in multiple electronic databases via EBSCOhost, ProQuest and Web of Science, accessible via Jyväskylä University Library, in October 2012. The search terms used included ["social media"] and [monitoring or scanning or tracking or analytics] and [communication or "public relations"]. The selection of optimal search terms was a result of several try-outs. For example, the terms mapping, metrics and measurement produced results that did not support the aim of the present study, and were consequently discarded.

The search covered abstracts, titles and keywords, and was limited to scientific articles in peer-reviewed journals. Hence, articles in consumer or trade magazines, for example, were excluded. In addition, the availability of the full article in the Jyväskylä University Library databases or on the internet was a prerequisite. Throughout the process, record was kept of excluded articles. Table 1 clarifies the process.

Table 1 Search results per database (*Number includes some duplicates)

Keywords	Database	First results	After initial scanning	Final sample coded
[“social media”] and [monitoring or scanning or tracking or analytics] and [communication or “public relations”]	EBSCOhost	62 (15.10.2012)	38	17
	Web of Science	22 (16.10.2012)	9	6
	ProQuest	18 (17.10.2012)	2	1
	Total	102*	49	24

In the first step, the abstract, title and keywords of all the results from the different databases, 102 articles altogether (see table 1), were checked to assess their potential for the study. If needed, the introduction and/or conclusions were also read. Moreover, the scientific quality of the publication was checked and duplicates removed. After this initial scan, articles considered to be adequate in scope were exported to RefWorks, a system for managing references.

This resulted in a total of 49 articles that were read through more thoroughly and from which, on the basis of the inclusion/exclusion criteria, a further selection was made. The inclusion criteria were: focus on social media, perspective on communication (here limited to communication with members of external public groups), perspective on monitoring, English language, and scientific quality. In the context of the present study, monitoring is understood as following and analysing online discourse in the social media. Hence, articles focusing on monitoring for evaluation purposes, e.g. measuring the outcomes of campaigns, were excluded. As the field of study is multidisciplinary in nature, articles from advertising or marketing were included in the sample when pertinent to the research objective. Moreover, technical ICT-related articles were also included if they offered useful elements from a communication point of view. After applying the inclusion/exclusion criteria, the final sample for the literature review before coding numbered 24 articles.

Data Analysis

The 24 articles in the final sample were read through and initially coded by underlining relevant parts of the text. During this process, a preliminary list of codes was formulated. All the articles in the final sample were transported to ATLAS.ti (version 7), a qualitative data management programme, where the final analysis and coding were conducted. The unit of analysis was one article.

The analysis was conducted with the help of the preliminary code list, which was further refined during the process on the basis of data-driven thematic coding (Riessman&Quinney, 2005). Thematic analysis is a common approach used for analysing qualitative data (e.g. Schwandt, 2007). In this approach, the analyst codes sections of text based on whether they contribute to emerging themes (Schwandt, 2007). The aim is to reduce the data by segmenting, grouping and categorizing, and then to summarize and reconstruct them so as to capture important patterns and concepts (Ayres, 2008; Elo&Kyngäs, 2008). Following the coding process, the coded text segments were further grouped and categorized, yielding five main themes. In addition to identifying emerging themes in the literature, some simple quantitative analyses were conducted, covering, for example, the distribution of articles according to year of publication, the focus of the articles, and the types of social media studied. The findings, including the main themes answering the research question and the trends discovered in the literature, are reported next.

Results

This study yielded insights from the scientific literature on the concept of monitoring, the monitoring process, monitoring methods, tools and solutions, methodological issues and trends.

Monitoring

In this section the concept of monitoring is clarified, including how it is described, its purposes, and its objects and focuses.

Monitoring Described

In the literature, social media monitoring is described as a research approach covering both data collection and analysis. It is an observational, passive and quantitative approach which

shares commonalities with ethnography and quantitative surveys (Branthwaite & Patterson, 2011). Historically, social media monitoring originates from the tradition of media content analysis (Rappaport, 2010). The idea of analysing the social environment through news media monitoring can be traced back to the sociologist Alvan Tenney and his proposal to measure the “social weather” in 1912 (Bengston, Fan, Reed, & Goldhor-Wilcock, 2009). In the pre-computer era, these early content analytical approaches required considerable resources in terms of time, labour and money (Bengston et al., 2009; Branthwaite & Patterson, 2011).

The benefits that social media monitoring is much praised for, and that are expected from it, include access to authentic and honest data (Eysenbach, 2009) by providing an account of social life “as it happens” (Branthwaite & Patterson, 2011). Moreover, it enables the continuous, minute-by-minute (Branthwaite & Patterson, 2011), real-time data collection much needed by emergency managers, and tracking of fast changing perspectives over time (Kavanaugh et al., 2012; Sundaram, Xie, De Choudhury, Lin, & Natsev, 2012). It may also provide access to novel information (Deluca et al., 2012) and reveal hidden insights or even topics of controversy (Sobkowicz, Kaschesky, & Bouchard, 2012). This could also pave the way for two-way communication and partnering with citizens, e.g. in emergency situations (Lindsay, 2010). Monitoring is also said to be cost-effective, e.g. in comparison with traditional research methods (Bengston et al., 2009; Eysenbach, 2009; Kavanaugh et al., 2012). In contrast, other sources cite the uncertainty related to possible costs of monitoring and responding to potentially voluminous incoming messages in the case of a crisis (Lindsay, 2010).

Social media monitoring methods can be described as “listening solutions” that provide technology; services and know-how to help researchers and organizations listen, interpret and respond to what people are saying online (Rappaport, 2010). When compared to the traditional research approaches, the similarities that social media monitoring shares with quantitative research include large samples, numeric data and difficulty in assessing meanings, while among those it shares with qualitative approaches are the gathering of spontaneous views and opinions, and a need for rigorous semantic analyses (Branthwaite & Patterson, 2011). Monitoring can also be seen as “a 24x7 focus group”, which provides community insights and intelligence for e.g. communication planners (Hipperson, 2010). In

turn, the differences between monitoring and the traditional quantitative approaches include more imprecise sampling and lack of control and standardisation in conducting research, while the differences with the traditional qualitative approaches include larger samples, no direct contact with the targets of the research, and lack of non-verbal cues, feedback and contextual information (Branthwaite & Patterson, 2011).

In addition to the function of monitoring as environment analysis and listening – the focus of the present paper – monitoring may also serve the purpose of evaluation, e.g. tracking the success of a message or a campaign (Zailskaite-Jakste & Kuvykaite, 2012). Moreover, methods for social media monitoring can be divided into active methods that aim at engaging, e.g. citizens, in social media interaction and consequently gathering data from them, as in the case of health practitioners tracking the progress of their patients through an e-health website (Laakso, Armstrong, & Usher, 2012), and passive methods that include passive and automatized data gathering and analysis (Eysenbach, 2009). This paper focuses on the latter, i.e. methods for monitoring naturally occurring interaction online.

Purpose of Monitoring

According to the literature, the purposes for conducting social media monitoring are manifold. First and foremost, monitoring is conducted for analysing the environment (Zailskaite-Jakste & Kuvykaite, 2012). The goal is to make sense of a vast amount of information and to see the big picture of the phenomena monitored (Kavanaugh et al., 2012). Monitoring provides a window on societal debate and sheds light on stakeholder perceptions, needs and attitudes of various kinds (Bengston et al., 2009; Chou, Hunt, Folkers, & Augustson, 2011). The objective may be identification of issues of concern related to e.g. public safety (Kavanaugh et al., 2012), trends in e.g. drug use (Deluca et al., 2012), or political opinions or reactions to public policies (Sobkowicz et al., 2012). Moreover, social media monitoring may be conducted to identify radical opinions (Yang, Kiang, Ku, Chiu, & Li, 2011), misinformation e.g. on vaccinations (Eysenbach, 2009), profile criticism (Keelan, Pavri, Balakrishnan, & Wilson, 2010) or negative information (Campbell, Pitt, Parent, & Berthon, 2011).

The results of monitoring can be further used for specific action purposes. For example, monitoring may serve as a support in policy-making (Sobkowicz et al., 2012) e.g. to make

more informed decisions (Bengston et al., 2009) and for different planning purposes (Hipperson, 2010), e.g. organization's social media communication (Zailskaite-Jakste & Kuvykaite, 2012). The aim of monitoring may include e.g. responding to rumours, mediating reactions, correcting inaccuracies and responding to protests (Gallaugher & Ransbotham, 2010). In the context of disasters and emergencies, monitoring can serve to establish situational awareness and to enhance communication with citizens (Lindsay, 2010), and to provide an early warning system to identify potential problems (Gallaugher & Ransbotham, 2010) and paracrises (Coombs & Holladay, 2012).

Groups of users that may benefit from social media monitoring in the business sector include marketing specialists, advertisers (Campbell et al., 2011; Hipperson, 2010; Rappaport, 2010) and brand managers (Branthwaite & Patterson, 2011), and in the governmental sector decision and policy makers and officials (Kavanaugh et al., 2012; Sobkowicz et al., 2012; Yang et al., 2011) from various areas, such as health and the environment (Bengston et al., 2009; Eysenbach, 2009; Keelan et al., 2010). Monitoring is often specifically the task of public information officers (Sutton, 2009) and closely related to the work of crisis and emergency officials and managers (Coombs & Holladay, 2012; Lindsay, 2010). Moreover, other users of monitoring found in the literature include health professionals and organizations (Chou et al., 2011; Deluca et al., 2012; Laakso et al., 2012), higher education institutions (Zailskaite-Jakste & Kuvykaite, 2012) and researchers from different fields (Boyd & Crawford, 2012). Finally, monitoring tools may also be used by citizen groups (Kavanaugh et al., 2012). In this study, however, the focus is confined to monitoring by organizations and professionals.

Object and Focus of Monitoring

Monitoring can be targeted according to e.g. stakeholder group (who interacts), type of information (what is talked about), type of message (what kind of interaction is going on), or type of social media (where the interaction takes place). Objects of monitoring in Twitter include electronic word-of-mouth communication between consumers and businesses (Zhang, Jansen, & Chowdhury, 2011) and community-related information (Kavanaugh et al., 2012). The object of monitoring may be to learn about connections in social networks. With regard to health-related information, for example, it may be useful to know what relationships subsist between people in order to contain the spread of an infectious disease (Eysenbach, 2009). Other kinds of targets monitored have included Facebook posts and comments, community-

related videos in YouTube (Kavanaugh et al., 2012), crisis-related news content in blogs (Sutton, 2009) and radical opinions in extremist forums (Yang et al., 2011).

The focus of monitoring, on the other hand, defines the more specific emphasis of monitoring activities. For example, it could be information prevalence (the number of occurrences of certain keywords or concepts), information incidence (the volume of new information units created per time unit of interest), or concept co-occurrence (occurrence of certain concepts together) (Eysenbach, 2009). Other focuses may be topics, subtopics, associated sentiments, communication patterns (Sobkowicz et al., 2012), changes in discursive patterns over time (Bruns, Burgess, Highfield, Kirchoff, & Nicolai, 2011), trajectory of retweets (Zhang et al., 2011) or biographical information about those interacting, including influential users (Kavanaugh et al., 2012). Other examples of this are given throughout this report.

Monitoring Process

In the literature, the process of monitoring is described as consisting of certain steps or phases; for example, Bengston et al. (2009) mention five steps for the fast monitoring of issues, while Bruns and Liang (2012) outline three phases for analysing tweets. Below, these and the other steps described in the literature are summarized and explained under four categories: preparation, data collection, data analysis and reporting. In this section they are explained as steps in the monitoring process although in practice the steps to be taken will also depend on the solutions, tools and methods chosen. Methods, tools and solutions for monitoring are explained in detail in the following subsections; here, we focus on the process.

Preparation

The engagement of an organization in social media interaction, including monitoring activities, should, as suggested in the literature, be guided by commonly agreed upon guidelines (Gallaughar&Ransbotham, 2010). Moreover, the choices made throughout the monitoring process depend on the organization's objectives (Rappaport, 2010) and can be explained by reference to specific social media monitoring strategies (Sutton, 2009).

The preparation phase includes the definition of the problem and identification of the issue to be monitored (Bengston et al., 2009). Other considerations are identification of the resources needed for monitoring, including the budget and the personnel to be responsible for

implementing the monitoring activities (Bengston et al., 2009; Rappaport, 2010; Sutton, 2009). This phase also includes identification of timing, including dates and timetable; geographic focus, i.e. where (e.g. country) and on what level (e.g. national or local); and key stakeholder groups who have a stake in the issue and whose perspectives on the issue of interest it may be relevant to know (Bengston et al., 2009).

The preparation phase also involves a choice among the available listening tools and solutions (Rappaport, 2010). Moreover, the message type, e.g. tweets, Facebook or blog posts, to be monitored and the focus of interest in the monitored messages, e.g. favourable / unfavourable attitudes, need to be defined (Bengston et al., 2009). The choice of data sources, i.e. the key media to be monitored, will depend on the problem or issue monitored, time and budget resources (Bengston et al., 2009; Sutton, 2009). The identification of relevant social media platforms may also help in limiting an otherwise overwhelming amount of information (Deluca et al., 2012).

Data Collection

Once the relevant choices in the preparation phase have been made, data in the relevant social media are collected. This can be done via programme interfaces or running search queries which vary from simple keyword searches to more sophisticated searches using e.g. Boolean operators (Rappaport, 2010). The collection of data is preceded by setting the search terms, taking into account possible synonyms (Bengston et al., 2009). Bruns and Liang (2012), in describing the criteria for an advanced, custom-made system for analysing tweets, propose parameters which the end-user needs to fill in the system to initiate the data collection: keywords, or search terms; language; results type, recent or mixed; and frequency of data collection.

In the case of a manual, researcher-driven monitoring process, the search is followed by downloading and/or saving the search results (Bengston et al., 2009). A central step in the data collection process is the archiving of the data collected. This allows not only historical archiving and more analytical flexibility (Deluca et al., 2012), but also enhances methodological quality, which is especially important when monitoring is conducted for research purposes.

Data Analysis

Following the data collection is the analysis phase. While the use of open-source tools may provide a means to capture the data, the analysis often has to rely on other, either computer-assisted or manual, tools and methods. Specific areas of analysis include general statistical analysis and activity metrics, network analysis, and textual analysis (Bruns & Liang, 2012). The process will, of course, depend on the data and method chosen. An analysis of textual data may focus, for example, on the flow, volume, overall tone and temporal evolution of the discussion (Bengston et al., 2009). The analysis of tweet datasets could focus on frequency over time of tweets, users, keywords, replies and retweets, or changes over time of keywords and phrases (Bruns & Liang, 2012).

Reporting

Last, the findings are reported. The form of the report again depends on the methods and tools chosen as well as the purpose of the monitoring. In a crisis context, for example, publication of the findings should be rapid (Bruns & Liang, 2012). When monitoring is conducted to support decision-making, short and compact outputs may be preferable (Bengston et al., 2009). The presentation of findings should be supported by illustrative visualization, such as bar charts of recurrent topics of Facebook posts, tag clouds on YouTube videos, pie charts of followers and followers of followers (Kavanaugh et al., 2012) or a map illustrating hot zones of blogging activity, including the distribution of activity and overall tone of the blogs (Keelan et al., 2010).

The need for decision-making based on the results of the monitoring is mentioned in several articles (Gallaughier & Ransbotham, 2010; Kavanaugh et al., 2012) but addressed more thoroughly only in one article, where interpretation of the results and selection of appropriate communication strategies are discussed (Coombs & Holladay, 2012). According to Sutton (2009), current processes are too slow to meet the demands of rapidly evolving online arenas, and hence responding to misinformation online is likely to emerge as a critical issue in future disaster management.

Monitoring Methods

The monitoring methods described in the literature cover textual analysis and network analysis.

Textual Analysis

The majority of the articles in the final sample described textual analysis methods in which the researcher has a central role, while other articles gave examples of computerized content analysis techniques. The researcher-driven manual methods and applications are described first, followed by computerized content analysis techniques.

Researcher-Driven Analysis

A common approach to research-driven content analysis of social media interaction, according to the literature, is first to conduct a search in the selected social media and then to analyse the data following different forms of textual analysis. For example, in a study analysing vaccine-related blog discourse in MySpace, data were collected using key words in the blog search engine and subsequently analysed with a focus on the frequency of vaccine-related arguments and overall tone – positive, negative or ambivalent – of the blogs (Keelan et al., 2010). In another study, comments posted on a county government’s Facebook page were content-analysed by topic, tweets from local civic organizations were analysed by popular topics and a tag analysis was conducted to YouTube videos pertaining to the county in question (Kavanaugh et al., 2012).

A textual analysis was also conducted in a study focusing on business engagement in Twitter, where tweets were classified into six groups: humorous, anecdotal, philanthropic, news, philosophical, and promotional (Zhang et al., 2011). Moreover, Bengston et al. (2009), in explicating their rapid issue tracking approach, give examples of cases where data from diverse online news media sources were collected via search and then analysed for e.g. main beliefs and favourable/unfavourable attitudes towards the issue of interest. Bruns and Liang (2012) describe content analysis of tweets, where the focus is on the keywords, terms and phrases used in order to map e.g. the overall distribution of keywords, and the occurrence over time and co-occurrence of keywords.

A different approach was taken in a study on cancer survivor stories on YouTube, where a narrative analysis covering both thematic (content-level) and discursive (linguistic-level) elements was conducted on transcribed video clips (Chou et al., 2011). The authors propose their method as a complementary technique to traditional content analysis, which could also

be conducted through a semi-automated coding scheme using natural language processing and computational techniques.

Computer-Assisted Analysis

Next to researcher-driven, manual analytical techniques are techniques involving software tools and programmes. For example, in a study on comments about consumer-generated ads on YouTube, first, a correspondence analysis was conducted with WordStat, a content analysis software program, to map brand personality dimensions, and second, a Bayesian machine learning-based content analysis was conducted with Leximancer, a software tool enabling both a thematic and relational analysis of the data, including main concepts and their interrelations (Campbell et al., 2011). In another study, machine learning and semantic-oriented approaches were combined to identify radical opinions in extremist web forums using four types of text features – syntactic, stylistic, content-specific, and lexicon – as text classification predictors, and three classification techniques – SVM, Naïve Bayes, and Adaboost (Yang et al., 2011).

Moreover, a paper presenting an opinion formation framework in which content analysis of social media is one building block proposes a combination of natural language processing (NLP) and semantic web approaches (SW) to detect political opinions, including topics and sentiments, automated in real time (Sobkowicz et al., 2012). However, the use of computer software and tools and the development of these techniques also involve researcher-driven efforts, as for example machine learning based methods may require a large amount of manually labelled training data to ensure accurate learning (Yang et al., 2011).

Statistics and metrics can add to the analysis, for example, in the case of Twitter, time-based metrics can be used to calculate e.g. the volume of tweets and keywords over time (Bruns & Liang, 2012).

Network Analysis

Besides methods that are used for analysing textual data online, the literature describes network analysis as another approach for analysing how information spreads in the social media. For example, a paper reporting research results on the networked public sphere in Australia, describes methodologies for mapping the blogosphere and tracking information

dissemination (Bruns et al., 2011). The described approach includes: (1) web crawling to identify networks of links combined with manual coding to refine the link lists, (2) automated content harvesting through Really Simple Syndication (RSS) feed alerts to capture new blog posts, (3) distinguishing blog content from other content, e.g. comments and the blogroll, using custom-made tools, (4) conducting textual analysis via e.g. automated concept mapping tools, and (5) combining textual analysis with link network mapping to produce both long-term maps and snapshots of current activity, including interlinkages and themes.

A related article describing French political blogging focuses on two types of networks: blogroll link networks describing the more permanent set of connections between blogs, and in-post link networks representing topical discussion networks in the blogosphere (Highfield, Kirchhoff, & Nicolai, 2011). Bruns and Liang (2012) describe different approaches to the network analysis of tweets: user-to-user messaging networks and keyword co-occurrence networks for homogenous networks; and user-and-URL networks, user-and-keyword networks, and user-and-hashtag networks for hybrid networks.

Tools and Solutions

Tools and solutions for social media monitoring include publicly available, open-source tools, and custom-made tools. Most of the ready-made tools focus on the content and frequency of the information shared while some also measure emotions in sentiment analysis. An issue which currently limits research in this area is the fact that many studies employ custom-made or commercial tools; these are discussed only to a limited extent and thus cannot be scrutinized, replicated nor learnt from by other researchers (Bruns & Liang, 2012). Therefore, a more systematic and open discussion on the research tools available is called for. The division of tools into free, open-source vs. often expensive custom-made tools also sets the border between more limited solutions vs. more advanced systems (Bruns & Liang, 2012).

Currently available tools and solutions mentioned in the literature include dashboards services that provide an overview of online activities, such as HootSuite, Netvibes and Trackur. A study on social media use by government officials showed that at that time geo-mapping features, in particular, were lacking and cited the open-access place-based application Locast as an example of emerging new solutions (Kavanaugh et al., 2012). Other

open access options for social media monitoring mentioned in the literature include various search solutions offered by Google, the Social Mention platform for tracking news, blog posts, videos and images, Addict-o-matic (Zailskaite-Jakste & Kuvykaite, 2012), Technorati.com for searching blogs (Bengston et al. 2009), and BotBox, a commercial tool for online news content, including blogs (Sutton, 2009).

Analytical tools for Twitter mentioned in the literature include Tweet Archivist, 140kit (Kavanaugh et al., 2012) and TweetDeck, an aggregation tool (Gallaughier & Ransbotham, 2010). Bruns and Liang (2012) discuss your Twapperkeeper (now in HootSuite), for tracking Twitter activities through its Application Programming Interface (API), the search API and the streaming API, and point out that the data analysis relies on additional tools and the analytical skills of the researcher. Other tools mentioned include Gawk, an open-source command-line tool which can be used to further process data into e.g. tweet statistics and metrics (Bruns & Liang, 2012), and the open source visualization tools Wordle, used also for YouTube video tags (Kavanaugh et al., 2012), and Gephi for visualizing networks, used for e.g. Twitter datasets (Bruns & Liang, 2012) and blogs (Highfield et al., 2011). The development of tools concerning social media monitoring is rapid and consequently new software mergers and names may arise.

Methodological Issues

Methodological issues may apply to more aspects of the social media monitoring process, but here they are explained in relation to research subjects, data, tools, type of social media, and sample.

Issues Related to Research Subjects

According to Bruns et al. (2011), the debate on ethics when crawling, tracking, and capturing social media content is only just starting. In fact, ethical questions are important throughout the process, from data collection to publication (Boyd & Crawford, 2012).

Ethical issues which have to do with research subjects, i.e. people whose interaction online is being monitored, centre on issues of privacy and informed consent (Eysenbach, 2009). According to some authors, the fact that social media content is publicly available does not mean that the producers of such content intended it to be utilized by anyone, and probably

rarely think of researchers as a target audience (Boyd & Crawford, 2012). This may also depend on the type of social media used, as Twitter is more clearly recognizable as a public place than Facebook, for example. Privacy issues may be especially sensitive in the case of disasters and emergencies (Lindsay, 2010), as some people stop tweeting when a crisis occurs, for example, to prevent theft in the case of an evacuation. Although it would not be reasonable to require consent from each person engaging in online discourse, these are questions that need to be tackled in research ethics and methodology (Bruns et al., 2011).

Issues Related to Data

Other methodological issues, related to accuracy and objectivity, have to do with the data being collected. The accuracy of large data sets drawn from the internet may be threatened by the fact that they are prone to outages and losses, and thus unreliable (Boyd & Crawford, 2012). The limitations of data collected in the social media have also to do with the limited archiving capacity of certain social media (Bruns & Liang, 2012). Due to the difficulty of studying historical events via e.g. Facebook and Twitter, researchers may more likely end up studying issues located more in the present or immediate past (Boyd & Crawford, 2012). Moreover, subtle meanings, attitudes and motivations are hard to interpret and understand (Branthwaite & Patterson, 2011). Among other things, this has to do with a lack of contextual information. When taken out of context, data, such as large data sets of networks, may lose or change their meaning (Boyd & Crawford, 2012). An accurate interpretation of plots and maps produced by complicated software may also be challenging (Campbell et al., 2011), leading to apophenia – the identification of patterns where they do not exist (Boyd & Crawford, 2012). The potential malicious use of social media, e.g. intentional provision of inaccurate information by terrorist groups or mischievous individuals may also threaten data accuracy (Lindsay, 2010).

A concern related to the data being collected is limited access. The right to social media data is often owned by (internet) companies which restrict access to the data in keeping with their own interests (Boyd & Crawford, 2012). Limited and paid access to research data creates a divide between the industry and the academy and hinders evaluation of the methodological claims of such studies (Boyd & Crawford, 2012.). The same goes for many of the tools and solutions used for gathering and analysing data in social media studies: limited access to

custom-made tools undermines the replicability and translatability of such studies to other contexts (Bruns & Liang, 2012).

Issues Related to Tools and the Type of Social Media

Data scalability issues in the case of tools arise from the enormous diversity, speed and volume of social media data, which challenge their storage and analysis (Sundaram et al., 2012). In particular, two areas are affected, storage space and computing power (Bruns & Liang, 2012): tools should be designed in such a way that they are able to continuously up-scale both of these. A critical factor for emergency situations, on the other hand, is a potential power outage, which could hinder the effective use of social media and other technologies (Lindsay, 2010).

The limitations of the existing tools and social media functions mentioned in the literature include the failure of available solutions to recognize subtle emotions, such as humour, sarcasm, irony or provocation (Sobkowitz et al., 2012), and the limited search functions of e.g. Facebook (Boyd & Crawford, 2012). In the case of Twitter, the limited archiving capacity necessitates immediate and on-going monitoring when an issue of interest, or a sudden event, like a crisis, takes place (Bruns & Liang, 2012).

Issues Related to the Sample

Sample-related methodological issues include the representativeness of the sample and generalizability of the findings. For example, the amount of information that can be retrieved per minute is much smaller than the amount that is being produced (Sundaram et al., 2012). As a result of these technical limitations, the quality of the sample may be hard to evaluate. In Twitter, for example, no dataset of tweets can be guaranteed to be comprehensive (Bruns & Liang, 2012).

Moreover, social media users do not equal whole populations of people. For example, in a study on cancer survivors' narratives on YouTube it was found that the video posters were mainly upper-middle-class Americans of European descent (Chou et al., 2011). Furthermore, in the case of e.g. blogs, the culture prevalent in the social media influences people's self-presentation enabling them to create multiple persona, exaggerating and transforming their everyday personalities (Branthwaite & Patterson, 2011). These challenges related to sampling

increase when multiple datasets are combined: when the source is unknown, understanding the sample is hard, and a larger sample is not necessarily a better sample (Boyd & Crawford, 2012).

Trends in the Literature

Although the literature search was broad, the number of relevant articles found was somewhat limited, confirming that this is a young area of research. The final sample comprised 24 articles published during the years 2009–2012. The articles were distributed by year of publication as follows: three in 2009, five in 2010, seven in 2011 and nine in 2012, indicating a growing interest in the topic.

The majority (14) of the articles included empirical data, the rest being coded as non-empirical. Of the empirical articles, four were based on a case study methodology. The non-empirical articles consisted of a diversity of article types.

The reviewed topic, social media monitoring, was studied and discussed in various contexts and from several disciplinary perspectives. The articles in the final sample can be divided into four main groups according to the context of application: business sector, crisis communication and management, political communication and governmental sector, and health sector. The volume of coverage across the different contexts was rather equal: business-related issues were addressed in seven articles, while the other groups each contained five articles. Two articles did not fall into any of the four groups.

Next, the focus of the articles and the types of social media studied are treated in greater detail.

Focus of the Articles

The majority (13) of the articles focused on the development of methods and tools. The topics covered included infrastructure for tracking and analysing tweets (Bruns & Liang, 2012), frameworks for opinion mining in social media with respect to modelling, simulating, and forecasting political opinions (Sobkowicz et al., 2012), development of tools for collecting, monitoring, and analysing radical opinions in messages in extremist web forums (Yang et al., 2011), a review of solutions and methods available for social media monitoring (Rappaport,

2010), an approach designed for decision makers and communication professionals for rapid analysis of online issues (Bengston et al., 2009), and progress in the development of methodologies to map blog networks online (Bruns et al., 2011; Highfield et al., 2011).

The second biggest group of articles(6)focused on the use of social media monitoring in different contexts. One was an exploratory study on social media use and analysis by government and community organizations from the perspective of both routine and critical incidents, includingemergencies (Kavanaugh et al., 2012). Other articles addressed, for example, social media use, including monitoring by emergency managers (Lindsay, 2010), and business engagement in online word-of-mouth communication (Zhang et al., 2011).

The development of theories and models was addressed in three articles. In the most recent of these, Coombs and Holladay (2012) introduce a new concept, paracrisis, and explain how to evaluate and respond to paracrises. The other two articles in this group concerned models which explain communication in the social media context, including monitoring (Gallaughier&Ransbotham, 2010; Zailskaite-Jakste&Kuvykaite, 2012).

Methodological discussions were the focus of two articles. One of them discussed methodological issues related to big data in the social media context (Boyd & Crawford, 2012) and the other compared traditional qualitative and quantitative research approaches with respect to social media monitoring (Branthwaite & Patterson, 2011).

Types of Social Media Studied

The social media platform YouTube was addressed in seven articles, e.g. a paper on multimedia semantics (Sundaram et al., 2012), comments on consumer-generated advertisements (Campbell et al., 2011) and a narrative analysis of cancer survivors (Chou et al., 2011). In some articles, YouTube was studied alongside other platforms (e.g. Kavanaugh et al., 2012).

Twitter and blogs were addressed in five articles. These included a paper presenting two approaches for studying Twitter, one an open-source and the other a more advanced, custom-made system (Bruns & Liang, 2012), and a paper studying business engagement in Twitter (Zhang et al., 2011). Articles focusing on blogs addressed e.g. networks of political blogs

(Bruns et al., 2011; Highfield et al., 2011) and vaccine debates in MySpace (Keelan et al., 2010).

Forums and platforms were addressed in four articles (e.g. Yang et al., 2011). Flickr was studied in one article (Sundaram et al., 2012), and monitoring of Foursquare, along with other platforms, in another (Gallaughier & Ransbotham, 2010). Facebook was not the focus of any single article, but appeared in three articles in combination with Twitter and YouTube (Deluca et al., 2012; Gallaughier & Ransbotham, 2010; Kavanaugh et al., 2012).

Conclusions and Preliminary Factors for Monitoring

Monitoring activities are crucial in crisis communication in order to know citizens' needs and to obtain feedback on public perceptions and reactions. Monitoring includes both traditional and social media; however, it is the latter in particular that response organisations need more knowledge about. Information exchange via these new channels has greatly increased over the last few years, and as the numbers of citizens who are active online increase, so also do expectations that authorities will respond to what is going on in the online environment.

This literature review describes rather than attempts to resolve the current challenges involved in social media monitoring. Figure 1 summarizes the main themes found in the literature.

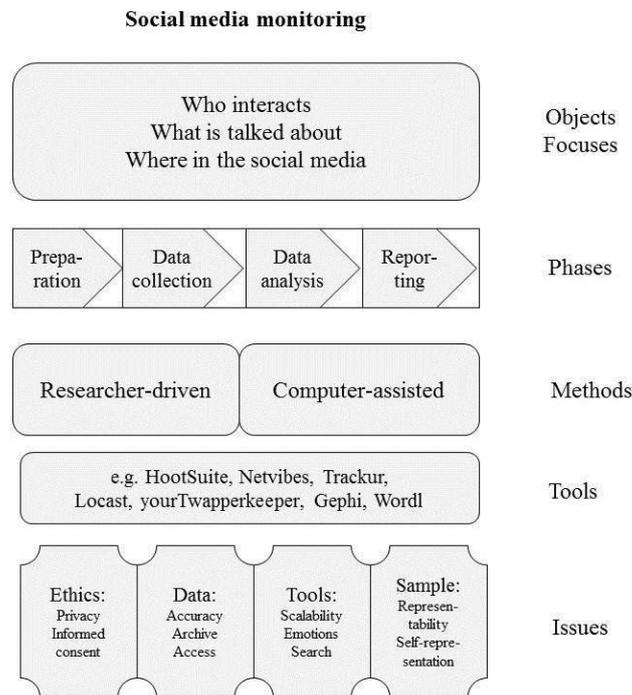


Figure 1 Overview of the main themes found in the literature

Social media monitoring methods can be seen as “listening solutions” that help organizations listen, interpret and respond to what people are saying online (Rappaport, 2010). Monitoring activities reveal who interact, what is talked about and in which media. The monitoring process consists of several steps: preparation, data collection, data analysis and reporting. The work needs adequate and qualified manpower. Textual and network analysis are the main data analytical procedures, and these may be either computer-assisted or researcher-driven. The process is facilitated by tools that currently include Hootsuite, Netvibe and Trackur, and also the open access tools Locast, Gephi and Wordl.

Methodological issues which have to do with research ethics concern lack of privacy and informed consent. Although social media content is publicly available, people may not be aware of this, and requiring the consent of each person involved is not feasible. Another problem is that the large datasets that are often used may not be comprehensive, or access to them may be restricted such as in cases where companies sell the data. Many of the current tools have limited scalability, including lack of storage and computing power, and may allow little inclusion of emotions, or offer limited search and archiving functions.

The publications found were from the period after 2009, and the sample demonstrates that increasing interest has been shown recently in this topic. The majority of the articles included empirical data. The topic has been studied from the perspective of different disciplines. The main foci concern the development of methods and tools, monitoring in different contexts, models and methodological issues. The social media that have been investigated include YouTube, Twitter, blogs, forums and platforms.

The work done so far leaves much to be desired. Tools were discussed only superficially, with the emphasis on methodological issues as well as the requirements, hopes and needs for the development of such tools. In the literature, ethical considerations pertaining to the studies conducted were mentioned in the introduction or discussion rather than reflected upon more deeply, e.g. in the methods section. The results of monitoring were linked to strategy options in only one article, and that on the topic of reputation crises. The relation with strategy is especially important for crisis situations, which require quick decision-making and action. When time is short and emotive reactions may spur people to behave in unpredictable ways, strategic thinking can help to mitigate harm.

Factors relevant for the monitoring of social media in crisis situations:

- Textual analysis is especially useful in seeking to understand citizen viewpoints
- Monitoring follows distinct stages; however, connecting reporting to decision making on communication strategies needs to be added to these
- Computer-assisted methods are needed to cover a large body of messages in real time, but researcher-driven work also continues to be needed
- Many of the tools available are not open enough to check precise procedures
- A combination of social media is aimed at, including e.g. Twitter
- It should be remembered that big samples are not always of good quality
- Monitoring in real time needs specialist manpower around the clock and communication expertise to interpret the results.

Participation in social media interaction begins with monitoring. To do this, authorities need to build accounts, profiles and – more importantly – gain followers before a crisis occurs. This means creating a basis, comprising preparedness, joint communication strategies with

other response organizations and, preferably, a multi-channel approach using different (social) media linked to a dedicated crisis website.

Here the emphasis was on monitoring as a start activity. How to utilise the results of monitoring for communication strategy-making remains a challenge for future studies.

Acknowledgements

“The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under Grant Agreement n° 261693“.

References

- Ayres, L. (2008). Thematic coding and analysis. In L. M. Given (Ed.), *The SAGE Encyclopedia of Qualitative research methods* (pp. 268-269). Thousand Oaks, CA: SAGE Publications.
- Bengston, D. N., Fan, D. P., Reed, P., & Goldhor-Wilcock, A. (2009). Rapid issue tracking: A method for taking the pulse of the public discussion of environmental policy. *Environmental Communication, 3*(3), 367-385.
- Boyd, D., & Crawford, K. (2012). Critical questions for big data. *Information, Communication & Society, 15*(5), 662-679.
- Branthwaite, A., & Patterson, S. (2011). The power of qualitative research in the era of social media. *Qualitative Market Research: An International Journal, 14*(4), 430-440.
- Bruns, A., Burgess, J., Highfield, T., Kirchhoff, L., & Nicolai, T. (2011). Mapping the Australian networked public sphere. *Social Science Computer Review, 29*(3), 277-287.
- Bruns, A., & Liang, Y. E. (2012). Tools and methods for capturing Twitter data during natural disasters. *First Monday, 17*(4-2).
- Campbell, C., Pitt, L. F., Parent, M., & Berthon, P. (2011). Tracking back-talk in consumer-generated advertising. *Journal of Advertising Research, 51*(1), 224-238.
- Chou, W. S., Hunt, Y., Folkers, A., & Augustson, E. (2011). Cancer survivorship in the age of YouTube and social media: A narrative analysis. *Journal of Medical Internet Research, 13*(1), e7.
- Coombs, W. T., & Holladay, J. S. (2012). The paracrisis: The challenges created by publicly managing crisis prevention. *Public Relations Review, 38*(3), 408-415.
- Deluca, P., Davey, Z., Corazza, O., Di Furia, L., Farre, M., Flesland, L. H., Mannonen, M., Majava, A., Peltoniemi, T., Pasinetti, M., Pezolesi, C., Scherbaum, N., Siemann, H., Skutle, A., Torrens, M., van der Kreeft, P., Iversen, E. & Schifano, F. (2012). Identifying emerging trends in recreational drug use; outcomes from the psychonaut web mapping project. *Progress in Neuro-Psychopharmacology & Biological Psychiatry, 39*(2), 221-226.
- Elo, S., & Kyngäs, H. (2008). The qualitative content analysis process. *Journal of Advanced Nursing, 62*(1), 107-115.

- Eysenbach, G. (2009). Infodemiology and infoveillance: Framework for an emerging set of public health informatics methods to analyze search, communication and publication behavior on the internet. *Journal of Medical Internet Research*, 11(1), e11.
- Gallaugh, J., & Ransbotham, S. (2010). Social media and customer dialog management at Starbucks. *Mis Quarterly Executive*, 9(4), 197-212.
- Highfield, T., Kirchhoff, L., & Nicolai, T. (2011). Challenges of tracking topical discussion networks online. *Social Science Computer Review*, 29(3), 340-353.
- Hipperson, T. (2010). The changing face of data insight - and its relationship to brand marketing. *Journal of Database Marketing & Customer Strategy Management*, 17(3-4), 262-266.
- Jesson, J. K., Lacey, F. M., & Matheson, L. (2011). *Doing your literature review: Traditional and systematic techniques*. London: Sage Publications.
- Kavanaugh, A. L., Fox, E. A., Sheetz, S. D., Yang, S., Lin, T. L., Shoemaker, D. J., et al. (2012). Social media use by government: From the routine to the critical. *Government Information Quarterly*, 29(4), 480-491.
- Keelan, J., Pavri, V., Balakrishnan, R., & Wilson, K. (2010). An analysis of the human papilloma virus vaccine debate on MySpace blogs. *Vaccine*, 28(6), 1535-1540.
- Laakso, E., Armstrong, K., & Usher, W. (2012). Cyber-management of people with chronic disease: A potential solution to eHealth challenges. *Health Education Journal*, 71(4), 483-490.
- Lindsay, B. R. (2010). Social media and disasters: Current uses, future options and policy considerations. *Journal of Current Issues in Media & Telecommunications*, 2(4), 287-297.
- Rappaport, S. D. (2010). Listening solutions. *Journal of Advertising Research*, 50(2), 197-213.
- Reynolds, B., & Seeger, M. W. (2005). Crisis and emergency risk communication as an integrative model. *Journal of Health Communication*, 10(1), 43-55.
- Riessman, C. K., & Quinney, L. (2005). Narrative in social work: A critical review. *Qualitative Social Work*, 4(4), 391-412.
- Robinson, S. J., & Newstetter, W. C. (2003). Uncertain science and certain deadlines: CDC responses to the media during the anthrax attacks of 2001. *Journal of Health Communication*, 8, 17-34.

- Rubin, G. J., Amlôt, R., & Page, L. (2011). The London polonium incident: Lessons in risk communications. *Health Physics*, 101(5), 545-550.
- Schwandt, T. A. (2007). THEMATIC ANALYSIS. In T. A. Schwandt (Ed.), *The SAGE dictionary of qualitative inquiry*. (3rd ed., p. 292). SAGE Publications, Inc. Retrieved from <http://srmo.sagepub.com/view/the-sage-dictionary-of-qualitative-inquiry/SAGE.xml>
- Sheppard, B. (2011). Mitigating terror and avoidance behavior through the risk perception matrix to augment resilience. *Journal of Homeland Security and Emergency Management*, 8(1), 26.
- Sobkowicz, P., Kaschesky, M., & Bouchard, G. (2012). Opinion mining in social media: Modeling, simulating, and forecasting political opinions in the web. *Government Information Quarterly*, 29(4), 470-479.
- Sundaram, H., Xie, L., De Choudhury, M., Lin, Y., & Natsev, A. (2012). Multimedia semantics: Interactions between content and community. *Proceedings of the IEEE*, 100(9), 2737-2758.
- Sutton, J. N. (2009). Social media monitoring and the democratic national convention: New tasks and emergent processes. *Journal of Homeland Security and Emergency Management*, 6(1), 1-20.
- Yang, M., Kiang, M., Ku, Y., Chiu, C., & Li, Y. (2011). Social media analytics for radical opinion mining in hate group web forums. *Journal of Homeland Security and Emergency Management*, 8(1), 1-17.
- Zailskaite-Jakste, L., & Kuvykaite, R. (2012). Implementation of communication in social media by promoting studies at higher education institutions. *Engineering Economics*, 23(2), 174-188.
- Zhang, M., Jansen, B. J., & Chowdhury, A. (2011). Business engagement on twitter: A path analysis. *Electronic Markets*, 21(3), 161-175.