Social Media Analytics

An Interdisciplinary Approach and Its Implications for Information Systems

Social Media Analytics is an emerging interdisciplinary research field that aims on combining, extending, and adapting methods for analysis of social media data. On the one hand it can support IS and other research disciplines to answer their research questions and on the other hand it helps to provide architectural designs as well as solution frameworks for new social media-based applications and information systems. The authors suggest that IS should contribute to this field and help to develop and process an interdisciplinary research agenda.

DOI 10.1007/s12599-014-0315-7

The Authors

Stefan Stieglitz (✉)
Department of Information Systems
University of Muenster
Leonardo Campus 11
48149 Muenster
Germany
stefan.stieglitz@uni-muenster.de

Linh Dang-Xuan
University of Muenster
Muenster
Germany

Axel Bruns
Queensland University of Technology
Kelvin Grove
Australia

Christoph Neuberger
Ludwig-Maximilians-University
Munich
Munich
Germany

Received: 2013-03-01
Accepted: 2013-10-30
Accepted after two revisions by Professors Bichler, Hess, Krishnan, and Loos.
Published online: 2014-02-15

This article is also available in German in print and via http://www.wirtschaftsinformatik.de: Stieglitz S, Dang-Xuan L, Bruns A, Neuberger C (2014) Social Media Analytics. Ein interdisziplinärer Ansatz und seine Implikationen für die Wirtschaftsinformatik. WIRTSCHAFTSINFORMATIK. doi:10.1007/s11576-014-0407-5.

1 The Era of Big Data and Social Media Analytics

In recent years, social media have experienced tremendous growth in their user base. For example, there are more than one billion members belonging to the Facebook network (Facebook 2013), while Twitter now has more than 280 million monthly active users (GlobalWebIndex 2013). There are a large number of different social media applications or platforms which in general can be categorized as weblogs, microblogs, social network sites, location-based social networks, discussion forums, wikis, podcast networks, picture and video sharing platforms, ratings and reviews communities, social bookmarking sites, and avatar-based virtual reality spaces. In a broader sense, social media refers to "a conversational, distributed mode of content generation, dissemination, and communication among communities" (Zeng et al. 2010, p. 13).

The mainstream adoption of social media applications has caused a paradigm shift in how people communicate, collaborate, create, and consume information. In particular, the process of information consumption and dissemination is closely interrelated with the process of generating and sharing information (Zeng et al. 2010). Furthermore, while of course huge societal power differentials persist, the curation and diffusion of publicly available information is no longer as easily controlled by a small number of institutional "gatekeepers".

In addition to the personal, everyday uses that are the source of social media's mass adoption, social media have also been increasingly used as communication channels in business, political, and other contexts. For example, companies have started to adopt internal as well as external (public) social media platforms for a number of purposes. While the use of internal social media applications should improve communication and collaboration among employees, knowledge management, and product/service innovation, various companies have started to establish social media-based networks with business partners and have also begun to engage in public social media activities for the purposes of marketing, public relations (PR), customer relationships, reputation management, and recruitment. In the political domain, social media are believed to have the potential to increase political participation by citizens and voters (Wattal et al. 2010). While Twitter is an ideal public platform for disseminating political information and opinions quickly and widely

(Stieglitz and Dang-Xuan 2013a, 2013b; Bruns and Highfield 2013), political actors (e.g., politicians, parties, foundations, etc.) have also begun to use Facebook to enter into dialogues with citizens and to encourage more political discussions.

Academic research from various disciplines of the social and even the natural and applied sciences has recently devoted more attention to social media. Social networks and social media have also become an important domain in information systems (IS) research. This recent interest in "Big Social Data" (Manovich 2012; Burgess and Bruns 2012) has been driven in part by facilitated access to large-scale empirical datasets from popular online social networking platforms such as Twitter, Facebook, and LinkedIn, as well as from other platforms that facilitate mass collaboration and self-organization such as weblogs, wikis, and user tagging systems. As boyd and Crawford (2012) point out, "the era of Big Data is underway. Computer scientists, physicists, economists, mathematicians, political scientists, bio-informaticists, sociologists, and other scholars are clamoring for access to the massive quantities of information produced by and about people, things, and their interactions" (boyd and Crawford 2012, p. 663). Defined as a cultural, technological, and scholarly phenomenon that rests on the interplay of technology, analysis, and methodology, Big Data is "less about data that is big than it is about a capacity to search, aggregate, and cross-reference large data sets" (boyd and Crawford 2012, p. 663). From a research perspective, social media can be understood as a kind of living lab, which enables academics to collect large amounts of data generated in a real-world environment.

There is a significant interest in analyzing Big Social Data from social media not only for research but also for practical purposes. For example, in analyzing social media data, companies see the opportunity for targeting advertising, PR, social customer relationship management (CRM), and business intelligence (BI). In particular, the primary interest behind corporate activities in social media is how to effectively use them as an additional channel for marketing. Further, B2B companies have also started to use social media analytics to identify new potential customers.

Recently, political institutions have also shown an interest in monitoring the public opinion on policies and political positions, detecting trending political topics, and managing their own reputation in the social web. Public officials could potentially use social media to identify situational information created by citizens in times of natural disasters (Bruns and Burgess 2012; Bruns and Liang 2012). Furthermore, based on social media data, health organizations could establish an early warning system for disease outbreaks that should help provide timely response measures. Also, individuals and consumers seek to make use of information and opinions from diverse sources in order to make more informed decisions. An additional use case for research can be found in conducting intercultural studies: by analyzing social media content, academics are able to directly compare how people in different countries or cultures react to certain global events, for example.

This interest has two main drivers. First, due to technological advances there is a new possibility of continuous, automated (real-time) monitoring, and analytics of social media content and interactions. Second, there is a change in public participation that leads to an increased complexity of the communication environment (e.g., increasing quantity and heterogeneity of communicators, unbounded communication, higher level of information diffusion with respect to range, scale, and speed, in particular due to the rapid development of mobile devices). In this sense, geo-data are another highly promising data source which can be effectively included in social media analytics procedures.

Indeed, recent studies and surveys have revealed an emerging need to continuously collect, monitor, analyze, summarize, and visualize relevant information from social interactions and user-generated content in various domains such as business, public administration, politics, or consumer decision-making (e.g., Zeng et al. 2010; Kavanaugh et al. 2011; Stieglitz et al. 2012). These activities, however, are considered difficult tasks due to the large number of different social media platforms as well as the vast amount, dynamics, and complexity of social media data. More specifically, social media communication generates an enriched and dynamic set of data and metadata, which have not been treated systematically in the data- and text-mining literature (Zeng et al. 2010). Examples include user-expressed subjective opinions, views, emotions, evaluation, and attitudes, the manifestations of which are pictures, videos, ratings, tags, user profiles, and other spatial, temporal, and attention-related data (i.e., data related to "likes", comments, retweets, mentions, etc.). In the last few years, however, a new interdisciplinary research area called "social media analytics" (SMA) has been established to address exactly these issues. Its primary goal is to develop and evaluate scientific methods as well as technical frameworks and software tools for tracking, modeling, analyzing, and mining large-scale social media data for various purposes. In a business setting, SMA might be considered a subset of BI that is concerned with methodologies, processes, architectures, and technologies that transform raw data from social media into meaningful and useful information for business purposes.

Recently, IS research has also shown an interest in studying the design, implementation, use, and management of social media as socio-technical systems, as well as their practical impacts on employees, enterprises, and society. In this paper, we propose SMA as a new research field combining knowledge from multiple disciplines to provide IS research with methodological foundations for gathering, modeling, analyzing, and mining large-scale social media data in order to make business, economic, social, and technical claims from both research and practical perspectives. Furthermore, we argue that SMA can help IS research develop decision-making or decision-aiding frameworks for enterprises that employ social media. Since such frameworks would require performance measures for social media use, SMA must also tackle the issue of measurement, which has been challenging until now. Finally, based on insights and practical implications derived from analyzing social media data, SMA should provide architectural designs and solution frameworks for new social media-related applications and information systems.

The remainder of this paper is structured as follows. First, we outline different problems and challenges of SMA research. We then present the most important interdisciplinary methods for SMA and point out the need for an interdisciplinary research agenda and co-operation to meet the challenges of SMA research. Finally, we provide some examples of initial systematic works in this emerging research area.

2 Challenges of Social Media Analytics

As an emerging research field, SMA is still in an early stage of development and therefore faces a number of research challenges. Similar to social network analysis, which is an interdisciplinary academic field mostly borrowing theories from other disciplines such as network science, sociology, statistics, and graph theory, SMA lacks a theoretical core and requires interdisciplinary research co-operation. Recently, there has been an increase in research attention directed at social media or social networks from research communities in all major disciplines of the social sciences (e.g., sociology, media and communication studies, business studies, economics, political science, and social psychology) as well as the natural and applied sciences (e.g., computer science, information systems, linguistics, statistics, and physics). Each discipline has its own research perspectives and objectives. For example, while computer science aims at developing efficient algorithms and tools for analyzing, mining, and predicting changes in the structures and processes of social networks, political science seeks to examine the impacts of social media use on political participation. Meanwhile, social science and business researchers are using social media as a sensor network and a laboratory for natural experimentation, providing valuable indicators and helping to test hypotheses about social interactions as well as their economic, political, and societal implications (Zeng et al. 2010). Therefore, SMA research must necessarily be of a multidisciplinary nature. However, the level of interdisciplinary research co-operation still tends to be low. In many cases, research methodologies and questions from more technical domains, such as computer and network science, are taking a dominant role compared to research that employs the methods and addresses issues that are central to other fields. Furthermore, a well-defined SMA research agenda is still missing. More systematic research along with specific, well-evaluated results is needed.

From a methodological perspective, SMA faces a number of challenges relating to the nature of social media data, their collection, and existing analysis and mining methods. First, social media data are generated in very large quantities and are highly dynamic and complex in their nature. Therefore, they cannot be processed easily using traditional data processing applications or database management tools as well as desktop statistics and visualization packages. Furthermore, they exhibit both structured and unstructured characteristics. While structured data (or meta-data) comprise user profile characteristics, spatial, temporal, and thematic data as well as attention-related data (e.g., number of "likes", comments, retweets, mentions etc.), unstructured data include user-generated textual content ranging from relatively context-sparse microblogs through Facebook comments to context-rich blogs and audiovisual materials. This information overload represents a significant challenge that calls for large computing capacities and sophisticated sampling, extraction, and analysis methods. In particular, deriving meaningful insights from such large and dynamic information sources poses significant challenges for data mining. Another important issue relates to the accuracy and objectivity of social media data. As boyd and Crawford (2012) point out, large data sets from Internet sources are often unreliable because of their potential incompleteness and inconsistency, in particular, when multiple data sets are used together (e.g. social media content and location-based data). Regardless of their size, data sets are always subject to limitations and biases. Without an understanding of those biases and limitations, misinterpretation is likely to happen.

Regarding data collection, it is a challenge to gather data or meta-data from a large number of different social media platforms using different methods of access. Many social media platforms do not provide convenient standardized ways for gaining access to data, such as application programming interfaces (APIs). Individual parsing or scraping techniques are required in such cases. Even when APIs are provided, they are often subject to obscure restrictions of data access, and to further changes as the business models of the platforms change. Moreover, different data formats have to be preprocessed in order to conduct further analyses. In addition, the fast-changing nature of the designs and concepts of social media platforms as well as unanticipated usage conventions by users (e.g., user-created innovations such as Twitter hashtags and retweets) represent another challenge. Finally, privacy issues are always present when data are collected. Researchers and other parties interested in gathering data may face questions such as whether it is ethical to collect, process, use, and report on social media data even if these are actually "public" in principle. As researchers point out, little is understood about the ethical implications of Big Data in general and of social media data in particular (boyd and Crawford 2012).

Regarding data analysis and mining, there is a lack of systematic interdisciplinary methodological frameworks that summarize the scientific analysis methods from multiple disciplines for diverse analytical purposes. Also, many methods are still in an early stage of development. In particular, existing computational methods and algorithms exhibit a number of weaknesses when it comes to eliciting semantics from large-scale dynamic social media data and making predictions based on those insights. For example, social media textual content tends to be context-rich and informal in nature, including emoticons, abbreviations, amplifications, slang, sarcasm, and irony, which exacerbate the challenges of effective text mining. Furthermore, it is still a challenge to effectively discover relevant user communities from large-scale and dynamic data. Another example concerns the detection of trends or memes from social media data. Recent advances in computer science and statistics may have introduced a variety of methods and algorithms to predict emerging topics. However, they still lack accuracy.

3 Social Media Analytics and Interdisciplinary Methods

Recent advances in different disciplines, particularly computer science, statistics, network analysis, computational linguistics, etc., have provided a variety of tracking, modeling, analysis, and mining techniques to tackle challenges related to social media data. In the context of SMA, three main analysis methods almost always find their application: (1) text analysis/mining, (2) social network analysis, and (3) trend analysis.

Text analysis/mining is a research technique within the field of content analysis that supports researchers in making replicable and valid inferences from texts to the contexts of their use (Krippendorff 2004). Automated quantitative methods of text analysis are required because of the massively growing amount of social media data. Based on these methods a broad variety of questions can be answered, among which are the classification of texts (i.e., sentiment analysis) and the identification and modeling of recurring topics (Krippendorff 2004). Recently a number of novel approaches made their way into the social sciences, most notably text classification based on unsupervised and supervised learning (Sebastiani 2002; Liu 2011). Supervised text classification, which is based on statistical algorithms from machine learning (e.g., support vector machine (SVM) or naive Bayesian classifier), has the potential to become a standard method for automated text mining. In addition, documents might be clustered based on unsupervised learning, using techniques such as hierarchical and k-means clustering. Regarding topic modeling, recent advances in natural language processing provide more sophisticated statistical models for discovering abstract topics that occur in documents, as well as for predicting emerging topics.

One important subfield of text analysis/mining is sentiment analysis or opinion mining, which has emerged as a distinct method to study people's opinions in terms of views, attitudes, appraisals, and emotions towards entities, individuals, issues, events, topics, and their attributes in a more thorough way (Pang and Lee 2008; Liu 2011). Until now, it has been difficult for people to find relevant sites, extract related sentences with opinions, read them, summarize them, and organize them into usable forms. Automated opinion discovery and summarization systems are thus needed, which can be accomplished by sentiment analysis (Liu 2011). Basically, sentiment analysis can be performed based on two different approaches. The first one is the traditional dictionary-based classification of sentiment orientation including polarity (positive and negative) and strength: i.e., dictionaries of words, each annotated with their sentiment orientation, are used to extract sentiment from text. Recently, sentiment analysis has begun to make use of another approach based on machine learning, where the classification of sentiment can be formulated as a learning problem with three classes: positive, negative, and neutral. Here, sentiment classification can be performed based on supervised or unsupervised learning (Liu 2011). Despite many advantages of automated approaches, sentiment analysis still faces many problems regarding the typically informal nature of social media textual content, which often contains emoticons, acronyms, amplifications, slang, and sarcasm or irony. In addition, different contexts or domains might exacerbate the accurate classification of texts in general. Therefore, manual text analysis is nevertheless needed to test the findings of automated analysis and to develop a more fine-grained picture, as this would provide a set of practices that enables human coders to define reproducible categories for qualitative features of text more reliably (Krippendorff 2004).

The second main analysis method is social network analysis (SNA), which studies the relationships between persons, organizations, interest groups, states, etc., by analyzing the structure of their connections (Scott and Carrington 2011). In an SMA context, SNA may help identify influential users or opinion leaders, and relevant user communities in social media. There are a number of different measures for the influence of an actor in a network. SNA thereby provides different metrics for the concept of centrality and prestige that can be applied to measure influence (e.g., degree, betweenness or eigenvector centrality; degree, proximity or rank prestige) (Wasserman and Faust 1994; Scott and Carrington 2011). Regarding the detection of relevant communities, SNA might also be useful with different community detection methods and algorithms (e.g., graph theoretical approaches such as the Girvan–Newman algorithm (Girvan and Newman 2002) or other clustering methods such as hierarchical, k-means and fuzzy c-means clustering (Prabhu et al. 2010)). However, SNA still faces challenges in discovering changing clusters in large-scale and dynamic data.

Trend analysis is the third main analysis method that makes use of recent advances in computer science and statistics to predict emerging topics. Many trend-detecting algorithms are based on so-called hidden Markov models where observations of topics are trained by such models which in turn are saved in a library for the topic's prediction. Topics with a similar life cycle are recorded and share the same model (e.g., Zeng et al. 2007; Liu and Guo 2011). Budak et al. (2011) propose two trend definitions called coordinated and uncoordinated trends that detect topics that are popular among highly clustered and distributed users, respectively. In contrast to this approach, Kasiviswanathan et al. (2011) propose a dictionary learning-based framework for detecting emerging topics in social media and related streams. The overall framework consists of two stages: (1) determining novel documents in the stream and (2) subsequently identifying each cluster structure among the novel documents. Mathioudakis et al. (2010) suggest another method for early online identification of items that attract a lot of attention in social media. They model social media activity using a stochastic model that intuitively captures the concept of attention gathering information items.

In practice, there are many analytical issues that are covered not by a single, but rather by multiple analysis approaches, which require a combination of different analytical methods. It is therefore useful to have systematic frameworks that are able to serve as methodological guidelines for different analytical purposes. However, such frameworks outlining scientific methods from multiple disciplines for different analysis purposes are still lacking. We therefore propose that SMA research should develop such frameworks, which should be tailored to different contexts of analysis (e.g., enterprise, politics, personal), since analysis purposes do vary accordingly. Moreover, as described above, given their weaknesses SMA research should aim to continuously improve computational analysis methods and algorithms as well as the development of new and innovative methods.

One possibility for developing new methods may be executed (at least in part) according to the guidelines for design science in IS research, which have been proposed by Hevner et al. (2004). Rooted in the engineering and artificial sciences, design science is a problem-solving paradigm that seeks to create new and innovative (IT) artifacts – which are broadly defined as constructs (vocabulary and symbols), models (abstractions and representations), methods (algorithms and practices), and instantiations (implemented and prototype systems) (Hevner et al. 2004). In our case, the first step would be the creation of a purposeful method for a specified problem (domain). The method must be then evaluated with respect to its utility, quality, and efficacy. It must contribute to research and practice by being innovative: solving an unsolved problem, or a known problem more effectively and efficiently. Given the multidisciplinary nature of

SMA, results have to be communicated effectively across different disciplines (see Hevner et al. 2004). In our opinion, it is equally important to improve automated instruments by comparing results with findings from manual content analyses, as these are said to be more accurate. One example for this is the manual identification of the characteristics of social media communication (e.g., usage of acronyms, emoticons, colloquial language) in order to train automated algorithms. Also, it has to be considered that social media content in different contexts (e.g., politics, business, sports, entertainment) will require different methods which consider case-specific conditions. Machine learning approaches are a good way of facing this challenge. However, besides addressing different domains, methods also need to be able to process different types of data such as images, videos, and geo-data. In order to solve these problems, interdisciplinary research projects are needed.

4 Future Research Directions and Interdisciplinary Co-operation

To tackle the research challenges described above, we suggest some future research directions that should be implemented on an interdisciplinary basis. First, SMA research should primarily be concerned with developing and evaluating scientific methods, technical frameworks, and software tools and platforms for tracking, modeling, analyzing, and mining large-scale social media data from an interdisciplinary perspective. In particular, machine learning-based classification of social media textual content and recognition of social network patterns need to be improved. In this regard, disciplines including computer science, artificial intelligence, natural language processing, statistics, and network science should be the main contributors. Importantly, SMA research should take the highly dynamic properties of social media data as well as the fast-changing nature of the designs and concepts of social media applications into account. In doing so, SMA research provides other disciplines with methodological foundations for their social media-related research. In particular, IS research that focuses on the design, implementation, use, and management of social media as socio-technical systems as well as on their practical impacts may greatly benefit from SMA.

Based on these methods, frameworks, and toolsets, SMA research must – as another direction – address the development of corresponding data-driven and dynamic decision-making or decision-aiding frameworks, particularly in the business context. For example, SMA research can help extend BI by adding a "social" component, which can be referred to as "social BI". More specifically, as business-related decision-making frameworks require well-articulated and clearly defined performance measures, SMA research should elaborate on the issue of measurement, which is highly relevant in IS research. Until now, it has been challenging to quantify performance measures, not least due to a broad spectrum of different social media applications, among other things. The consequence of this measurement problem is the difficulty for enterprises to determine social media return-on-investment. For example, there are still open questions on how to measure the effectiveness of public relations and advertising activities through social media, which is relevant for decisions on the appropriate targeting and budget allocation for such activities, as well as on the selection of the instruments.

Another research stream should deal with the provision of architectural designs and solution frameworks for existing and new social media applications, based on research findings from the study of the interaction between the design and concept of social media applications and the eventual adoption and usage behavior, as well as from other insights and practical implications derived from analyzing social media data. IS research may also benefit from this research stream when it comes to designing and implementing new social media-based applications and information systems.

Finally, SMA should not be restricted to analyzing social media only. Rather, it also should address the interplay or interdependence between social and traditional media, which also includes other (non-social media) online content and activities. For example, an (online) advertising campaign may stimulate consumers' production of user-generated content and social media activities, and vice versa.

In this contribution, we call for a significant increase in the level of interdisciplinary research co-operation. This co-operation should not be limited only to scattered collaborations between individual researchers, but must extend to large-scale, coordinated research activities across all major disciplines that are stakeholders in social media research, including sociology, media and communication studies, business studies, economics, political science, and social psychology, as well as computer science, information systems, linguistics, and statistics. This effort must aim to generate significant advancements in the scientific methods for analyzing social media, as well as to answer research questions from across the different disciplines. For example, in communication studies, theories about public sphere and media effects such as agenda-setting, information diffusion, opinion leadership, the spiral of silence, deliberation, fragmentation, and polarization hypotheses, the digital divide, and inequalities in the distribution of attention and influence should be adapted to the special conditions of the Internet and tested. More generally, theories from the social sciences should be better connected to the methods of the applied sciences such as computer science, information systems, and statistics, and vice versa.

Considering the implications of SMA for IS, there are two important perspectives which both have an impact on the research agenda. First, SMA itself is a research field which has its own research questions. Examples are:

- Which methods (from various disciplines) can be adapted to SMA?
- How can methods be combined and improved to consider the specific requirements of social media communication?
- What tools might help researchers to gather and analyze data in social media? How can those tools be developed?
- How can methods be adapted to changing platform characteristics and changing communication behaviors?
- How can all relevant data (e.g., about a certain topic) be extracted from the overall network structure?

It has to be remarked that most of those questions cannot be answered generically. For example, the selection of appropriate methods depends on the actual research question as well as on the social media platforms. However, we believe that all such questions are relevant to IS research, and should therefore be addressed by IS researchers.

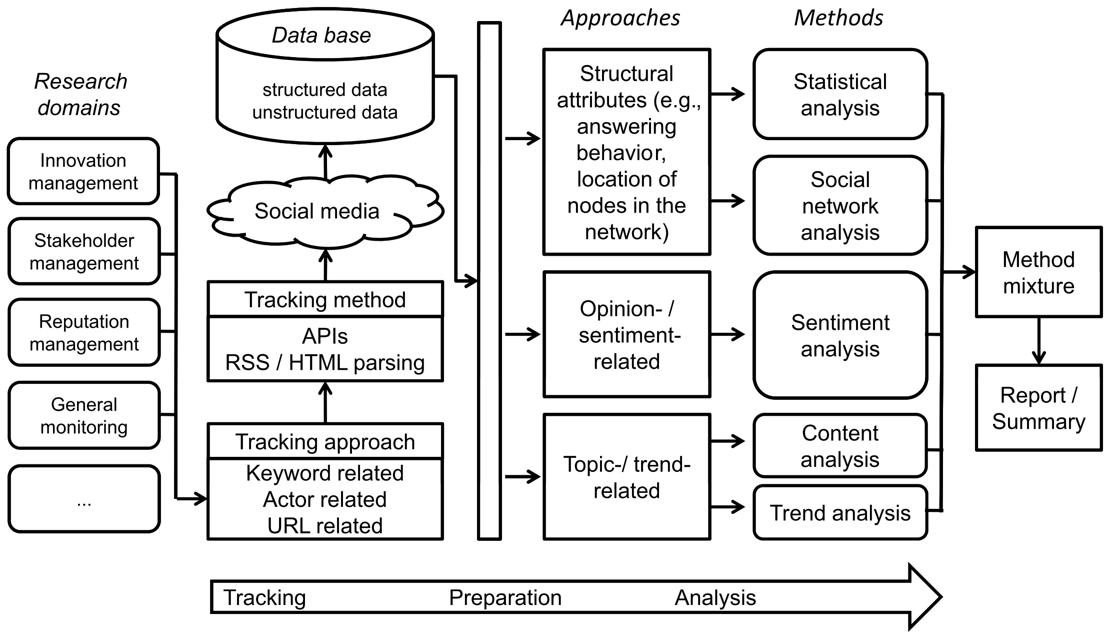

Fig. 1 Social Media Analytics Framework (adapted from Stieglitz and Dang-Xuan 2013a)

As a second perspective, SMA can also be understood to be a methodological approach which might be usefully applied in various domains besides IS (i.e., communication studies, marketing). With regard to IS, SMA might help to address the following exemplary questions of organizational and technical nature:

- How do certain social media features (such as the act of "liking" or retweeting) affect communication behavior?
- How does information spread through a network?
- What differences and similarities exist between internal business networks and public networks?
- How do different social media platforms affect each other?
- How can data from social media analytics be used to support knowledge management and community management?

To answer those or related questions, a systematic research design including appropriate methods is required. However, until now, only some first steps into this direction have been made.

5 Examples of Initial Systematic Works

As mentioned before, SMA can be applied to address different goals, such as to improve innovation management (e.g., by filtering the suggestions and ideas of customers), stakeholder management (e.g., by learning about the concerns of customers), reputation management (e.g., by evaluating the public's opinions about the enterprise), and general monitoring (e.g., to identify overall new developments as soon as possible). In order to support these tasks it is necessary to develop an interdisciplinary, systematic methodological framework that employs a mixture of analysis methods from different disciplines, particularly computer science, mathematics/statistics, network analysis, linguistics, etc.

The first challenge is to track relevant data according to the defined goal. There are different data-tracking approaches such as the keyword-, actor-, or URL-related approach. Depending on the social media platform(s), APIs, RSS, or HTML parsing can be used to track structured data (e.g., links, follower-followee relationships) or unstructured data (e.g., textual content) in social media. There already exist various tracking tools for popular social media platforms such as Twitter and Facebook. However, it might be necessary to develop or modify a tool which gathers and prepares the necessary data. Following the tracking stage, data need to be preprocessed (e.g., by removing spam manually or based on filters). As a next step, appropriate analysis approaches (e.g., the identification of structural attributes, sentiments, or topic- and trend-related patterns), methods (e.g., statistical analysis such as regression analysis, social network analysis, sentiment analysis, content analysis, or trend analysis), and analysis tools (e.g., Gephi, SentiStrength) have to be selected depending on the research question. Moreover, decisions have to be made regarding static or dynamic data analysis. For example, static data analysis might be usefully applied to identify the co-occurrence of specific words in a data set; on the other hand, dynamic data analysis might be useful to better understand how issues are evolving over time in the social space (**Fig. 1**).

There are various potential research questions which can only be answered by combining two or more methods. One example, requiring social network analysis as well as content analysis, would be the identification of opinion leaders within a network.

Automated methods provide an opportunity for an (almost) complete and continuous monitoring of communication. In the past, social sciences mostly had to be content with case studies or small samples. However, researchers should be aware that the development and implementation of new methods and/or improvement of existing methods is still necessary due to the limitations of computationally-based methods and algorithms. Also, we should bear in mind that social media are only one part of the Internet, and of the public sphere as a whole. Therefore, we also need to analyze the differences and connections between social media and other components of the Internet or offline media.

Besides this framework (Stieglitz and Dang-Xuan 2013a), some recent steps were made towards more systematic research in the field of SMA. In a business setting, Larson and Watson (2011) proposed a "social media ecosystem" framework, which explicates the social media-enabled relationships among stakeholders and suggests how future researchers can address research questions based on this model. Furthermore, it paves the way for developing measures of those firm-customer social media activities that have a critical bearing on firm performance. In another contribution, Rosemann et al. (2012) introduced "social CRM" and "social BI" as emerging fields of research. In addition, they designed a multidimensional data model for the conceptual design of social BI systems and demonstrated its application by developing management reports in a retail scenario. Finally, in the absence of standard metrics for a comparison of communicative patterns across different contexts on Twitter, Bruns and Stieglitz (2013) provided a catalogue of widely applicable, standardized metrics for a more comprehensive analysis of Twitter-based communication, with a particular focus on hashtagged exchanges. Moreover, they pointed out a range of potential uses for such metrics.

6 Conclusion

In this paper, we have introduced SMA as an emerging interdisciplinary research field that, in our view, will significantly influence future social media-related research from different disciplines, and will have a very high practical relevance. Despite a number of challenges outlined above, we argue that SMA can provide other disciplines – including IS – with methodological foundations for their social media-related research. Furthermore, we believe that SMA can help IS research to develop decision-making or decision-aiding frameworks by tackling the issue of social media-related performance measurement as well as to provide architectural designs and solution frameworks for new social media-based applications and information systems. Finally, we call for an interdisciplinary SMA research agenda as well as a significant increase in the level of interdisciplinary research co-operation, which must aim to generate significant advancements in scientific methods for analyzing social media, as well as to answer research questions from different disciplines.

References

boyd d, Crawford K (2012) Critical questions for big data. Information, Communication & Society 15(5):662–679

Bruns A, Burgess J (2012) Local and global responses to disaster: #eqnz and the christchurch earthquake. In: Proc of Australia New Zealand disaster and emergency management conference AST management, Brisbane, pp 86–103

Bruns A, Highfield T (2013) Political networks on twitter: tweeting the Queensland state election. Information, Communication & Society 16(5):667–691

Bruns A, Liang EY (2012) Tools and methods for capturing twitter data during natural disasters. First Monday 17(4)

Bruns A, Stieglitz S (2013) Towards more systematic twitter analysis: metrics for tweeting activities. International Journal of Social Research Methodology 16(2):91–108

Budak C, Agrawal D, El Abbadi A (2011) Structural trend analysis for online social networks. Technical report UCSB/CS-2011-04, UCSB

Burgess J, Bruns A (2012) Twitter archives and the challenges of "Big social data" for media and communication research. M/C Journal 15(5):1–7

Facebook (2013) Facebook key facts. http://newsroom.fb.com/Key-Facts. Accessed 2013-02-19

Girvan M, Newman MEJ (2002) Community structure in social and biological networks. Proceedings of the National Academy of Sciences USA 99:8271–8276

GlobalWebIndex (2013) Twitter now the fastest growing social platform in the world. https://www.globalwebindex.net/twitter-now-the-fastest-growing-social-platform-in-the-world/. Accessed 2013-09-03

Hevner AR, March ST, Park J, Ram S (2004) Design science in information systems research. MIS Quarterly 28(1):75–105

Kasiviswanathan SP, Melville P, Banerjee A, Sindhwani V (2011) Emerging topic detection using dictionary learning. In: Proc of CIKM'11

Kavanaugh A, Fox EA, Sheetz S, Yang S, Li LT, Whalen T, Shoemaker D, Natsev P, Xie L (2011) Social media use by government: from the routine to the critical. In: Proc of 12th annual international digital government research conference, College Park, Maryland, pp 121–130

Krippendorff K (2004) Content analysis: an introduction to its methodology. Sage, London

Larson K, Watson RT (2011) The value of social media: toward measuring social media strategies. In: Proc of 32nd international conference on information systems, Shanghai, China

Liu B (2011) Web data mining: exploring hyperlinks, contents, and usage data. Springer, Heidelberg

Abstract

Stefan Stieglitz, Linh Dang-Xuan,
Axel Bruns, Christoph Neuberger

Social Media Analytics

An Interdisciplinary Approach and Its Implications for Information Systems

In this contribution, we introduce "social media analytics" (SMA) as an emerging interdisciplinary research field that, in our view, will have a significant impact on social media-related future research from across different academic disciplines. Despite a number of challenges, we argue that SMA can provide other disciplines – including IS – with methodological foundations for research that focuses on social media. Furthermore, we believe that SMA can help IS research to develop decision-making or decision-aiding frameworks by tackling the issue of social media-related performance measurement, which has been challenging until now. Moreover, SMA can provide architectural designs and solution frameworks for new social media-based applications and information systems. Finally, we call for an interdisciplinary SMA research agenda as well as a significantly increased level of interdisciplinary research co-operation, which must aim to generate significant advancements in scientific methods for analyzing social media, as well as to answer research questions from across different disciplines.

Keywords: Social media analytics, Information systems, Big data, Interdisciplinary methods, Research agenda, Interdisciplinary co-operation

Liu R, Guo W (2011) HMM-based state prediction for Internet hot topic. In: Proc of the IEEE international conference on computer science and automation engineering (CSAE).

Manovich L (2012) Trending: the promises and the challenges of big social data. In: Gold MK (ed) Debates in the digital humanities, University of Minnesota Press, Minneapolis, pp 460–475

Mathioudakis M, Koudas N, Marbach P (2010) Early online identification of attention gathering items in social media. In: Proc of the 3rd ACM international conference on web search and data mining, 4–6 February, New York, USA. 2010

Pang B, Lee L (2008) Opinion mining and sentiment analysis. Foundations and Trends in Information Retrieval 2(1–2):1–135

Prabhu J, Sudharshan M, Saravanan M, Prasad G (2010) Augmenting rapid clustering method for social network analysis. In: Proc of 2012 IEEE/ACM int conf adv soc netw analysis min, pp 407–408

Rosemann M, Eggert M, Voigt M, Beverungen D (2012) Leveraging social network data for analytical CRM strategies – the introduction of social BI. In: Proc of 20th European conference on information systems, Barcelona, Spain

Scott J, Carrington PC (2011) Handbook of social network analysis. Sage, London

Sebastiani F (2002) Machine learning in automated text categorization. ACM Computing Surveys 34(1):1–47

Stieglitz S, Brockmann T, Dang-Xuan L (2012) Usage of social media for political communication. In: Proc of 16th Pacific Asia conference on information systems, Ho Chi Minh City, Vietnam

Stieglitz S, Dang-Xuan L (2013a) Social media and political communication – a social media analytics framework. Social Network Analysis and Mining 3(4):1277–1291

Stieglitz S, Dang-Xuan L (2013b) Emotions and information diffusion in social media – an investigation of sentiment of microblogs and sharing behavior. Journal of Management Information Systems 29(4):217–248

Wattal S, Schuff D, Mandviwalla M, Williams C (2010) Web 2.0 and politics: the 2008 US presidential election and an e-politics research agenda. Management Information Systems Quarterly 34(4):669–688

Wasserman S, Faust K (1994) Social network analysis: methods and applications. Cambridge University Press, New York

Zeng J, Zhang S, Wu C, Xie J (2007) Predictive model for internet public opinion. In: Proceedings of the fourth international conference on fuzzy systems and knowledge discovery (FSKD 2007)

Zeng D, Chen H, Lusch R, Li S (2010) Social media analytics and intelligence. IEEE Intelligent Systems 25(6):13–16