Introduction

Sorin Adam Matei, Elisa Bertino, and Martha Russell

As engagement with social media has become a dominant information acquisition and dissemination experience, the nature of the collection, production, and consumption of information has also changed. One of the most significant changes is the lowering of cost and technological barriers for sharing knowledge or opinions. User generated content dominates social media. This challenges traditional methods of collecting, disseminating and evaluating information. As much of the information exchanged on social media is often created or vetted by individuals or corporations whose identities, motives, or abilities are poorly or often simply unknown, we need new tools, theories, and practical strategies for evaluating the quality of the content and the credibility of its authors. Modelling the provenance and impact of authorship on social media is of crucial importance for explaining the emergence and impact of human motivations on social media content generation. Research on presenting, visualizing and explaining the social context of any given user in a social medium information exchange is equally important. In brief, researchers and practitioners need to create theories, methods and tools that make the authorship and dissemination process more transparent. We need new ways to understand at a glance, who, in what context, and if possible why creates or disseminates specific units of content.

The significance of this task cannot be underestimated, especially we consider the following facts:

S.A. Matei (✉)
Brian Lamb School of Communication, Purdue University, West Lafayette, IN, USA
e-mail: smatei@purdue.edu

E. Bertino
Purdue University, West Lafayette, IN, USA
e-mail: bertino@purdue.edu

M. Russell
Stanford University, Stanford, CA, USA
e-mail: martha.russel@stanford.edu

S.A. Matei et al. (eds.), *Transparency in Social Media*,
Computational Social Sciences, DOI 10.1007/978-3-319-18552-1_1

1. Between seven to nine Google searches for common terms return as one of the first page results a Wikipedia page (Miller, 2012). As a consequence, Wikipedia is the sixth most visited site in the world, attracting about 300 million readers every day (Alexa.com). This is almost 15 % of the entire world Internet population, estimated at around two billion people.
2. Facebook is the most visited site in the world, having more than a billion users, or half or the world Internet population (Alexa.com). Facebook is not, however, just for amusement. One third of US adults get their daily news from Facebook, either as links to articles or as original content (Pew Internet).
3. Twitter, on the other hand, is becoming a leading source for breaking news on a par with traditional newswire services. A study indicated that of 27 seven major breaking stories within a 6 week period in 2011, Twitter beat the newswires eight times and was just as fast in four other instances, showing that it has become just as powerful a source of information as traditional media (Petrovic et al., 2013).

The contributors to this volume present in a synergistic manner some of the significant theoretical and practical contributions in the area of social media reputation and authorship measurement, visualization, and modelling. The book justifies and propose several significant contributions to a future agenda for understanding the requirements for making more social media authorship more transparent. Building on work presented in a previous volume, *Roles, Trust, and Reputation in Social Media Knowledge Markets* (Bertino & Matei, 2014), the present volume discusses new tools, applications, services, and algorithms that are needed for the future of content in a real-time publishing world. These insights may help people who interact with or create content through social media or may assist in analyzing audience attitudes, perceptions and behavior in informal social media or in formal organizational structures. The contributors propose and answer forward looking questions about tools, applications, and services that visually or editorially mediate social interactions in social media. In addition, the volume includes several chapters that analyze the higher order ethical, critical thinking, and philosophical principles that may be used to ground social media authorship or use on valid moral and socially consequential foundations. Together, the perspectives presented in this volume may help us understand how social media content is created and how its impact can be evaluated. The chapters demonstrate thought leadership through new ways of constructing social media experiences and making traces of social interaction visible.

The substantive goal of this synergetic volume is to help researchers and practitioners design services, tools, or methods of analysis that encourage a more transparent process of interaction and communication on social media. Knowing who has added what content and with what authority to a specific online social media project can help the user community better understand, evaluate and make decisions and, ultimately, act on the basis of such information. As mentioned, research and development that support this goals need not ignore the ethical dimension of the problem, which is discussed in the last section of the volume.

The chapters are both retrospective and prospective. Scholars and practitioners look back at the work they have conducted so far, sharing with the readers some of

the lessons they have learned from their own work. They also discuss the areas that they find understudied or that promise the greatest intellectual or practical payoff in the future. Many of the contributions place the discussion in the context of social network analysis or "big data" research.

Synthesized thematically, the present volume explores the following core issues:

1. How do author feedback and incentive structures influence participation, value creation and reputation of social communities and social media content creation in various contexts—commerce, education, entertainment, government?
2. In what ways does curation infrastructure influence content creation—e.g., crowdsourcing—and sharing. How do the participants' perception of the factors influence content credibility, risk and trustworthiness? What types of statistical strategies or procedures are needed to better understand how social media roles emerge, function, generate valuable content, accrue trust and inspire credibility?
3. What kinds of tools, especially net-centric statistical analysis aimed at large social media datasets, can be adapted to make social media interactions more transparent to social science researchers or avid content authors?
4. What new approaches are needed to explore security and user identity in social media contexts?
5. What ethical and philosophical dimensions are involved in social media authorship and analysis processes?

The volume is structured in four sections. The first section introduces the main themes of the volume by reviewing the KredibleNet workshop on Reputation, Trust, and Authority funded by a National Science Foundation grant and held at Stanford University in 2013. The workshop proposed an agenda for future research and development related to social media knowledge production. Researchers and practitioners discussed the ways in which online information is transformed into "truth"—validated, repudiated, credentialed, measured, weighed and processed—as well as influenced, manipulated and controlled. The participants examined online behavior in order to reveal the organic emergence and evolution of social roles, hierarchy and elites online. Participants also identified procedures and features instrumental for stimulating, managing and otherwise controlling this behavior. Researchers discussed ways by which social media content is leveraged for insight into public opinion and sentiment, the flow of information, and the rise to prominence of relevant topics and issues. Analysis of social media also exposed the vulnerability and malleability of information and facts. Presentations investigated the potential of online crowdsourcing for accomplishing challenging tasks—including visualizing innovation, organizing people, aggregating information and data, sourcing high quality content, managing complex projects—as well as for seeding misinformation.

The second section proposes a frameworks for understanding the various paths by which authors emerge and content is created on social media, and how the trustworthiness of this content can be assessed. A set of case studies, methodological chapters, and tool presentations discuss the theoretical principles and methodological approaches needed for explaining authoring and trust mechanisms

involved in social media. Aleksandar Ignajatovic, Mohsen Rezvani, and Elisa Bertino discuss the principles required by aggregating robust data from inconsistent information. The chapter also introduces a methodology to protect against the malicious information sources that collude in order to perform information deception attacks. James Cavarlee proposes a new methodology for detecting crowdsourcing of unprincipled or socially deleterious tasks, such as spreading malicious URLs in social media, deploying artificial grassroots campaigns (astroturf), spreading rumors and misinformation, and manipulating search engines. His chapter is an important contribution to better understanding the authenticity and validity of user generated content.

User generated social media content represents, at the same time, a very valuable source of information, especially in time of emergencies. Social media flows can be used as sui generis social monitoring tools and just in time information dissemination channels. The chapter authored by the VACCINE team (Jiawei Zhang, Shehzad Afzal, Junghoon Chae, Guizhen Wang, Dennis Thom, Sorin A. Matei, Niklas Elmqvist, and David S. Ebert) is dedicated to a tool suite that identifies influential users in social networks, detects anomalous information diffusion patterns, and locates their geo-spatial coverage or impact. The chapter describes a visual analytics framework that can handle these issues based on dynamic social network analysis. It also discusses a visual analytics approach that allows decision makers to analyze large volumes of social media data to detect and examine abnormal events within location-based social networks.

The next two chapters are dedicated to alternative methodologies and tools for detecting specific social structures of interaction on social media. Sorin Adam Matei, Robert Bruno, and Pamela Morris discuss a methodology derived from social entropy theory and a visualization tool for wiki spaces that can detect optimal social structuration in online collaborative groups. The tool can be used to create self-monitoring, motivation and moderation mechanisms to enhance collaboration and learning in online spaces. Marc Smith, Itai Himmelboim, Ben Shneiderman, and Lee Rainie discuss the applicability of network analysis and of a dedicated tool, accessible to the lay public, NodeXL (a spreadsheet plugin), to categorize the basic structures of communication and interaction on Twitter. The six patterns they uncover reveal the important role played by self-expression, on the one hand, and polarization, on the other, in shaping Twitter interactions. Their research can also start a broader conversation about what is and what is not possible in such social interactional spaces.

Faisal Ahmed and Marina Gavrilova discuss innovative methods for identifying social media authors in conditions of poor provenance documentation. Their methodologies, relying on socio-behavioral metrics captured in time and space, is an important contribution to making social authorship online more transparent and accountable.

Finally, Brian Britt presents a methodology for detecting the emergence of social structures on massive social media content projects, such as Wikipedia. His approach reduces a complex problem to a manageable process, utilizing longitudinal analysis of presence or absence of certain actors in a leading elite of contributors.

The third section continues the conversation about making social media interactions and authorship more transparent, this time through new methods of curation, documentation, and segmentation. Two chapters, one discussing the open collaborative science education platform NanoHub (Michael G. Zentner, Lynn K. Zentner, Dwight McKay, Swaroop Samek, Nathan Denny, Sabine Brunswicker, and Gerhard Klimeck) and Ostinato, a data curation and visualization technique (Jukka Huhtamäki, Martha Russell, Kaisa Still, Neil Rubens), offer two best use scenarios for social media curation and content management. Sorin Adam Matei and Jeremy Foote, on the other hand, explore the manner in which Wikipedia avoids to visualize the social nature of the authorship process on the front page of each article in the name of a minimalist design, which in fact hides a social compact that privileges long tenure users and editors.

Mazdak Nik Bakht and Tamer E. El-Diraby discuss the applicability of social media mining strategies to urban development and planning. They propose analyzing conversation networks on social media as a source of creative ideas regarding project scope, funding, design and operations plans. Such analysis should be influenced by two major factors: domain-relevancy and benchmarking best practices.

The fourth and last section elevates the discussion about authorship, transparency, and credibility to the ethical and philosophical level. Robert Laughlin speaks about the relationship between authorship, credibility, and identity in social media. He suggests that social media anonymity, which is considered by many a given, can lead to social dilemmas, favoring amoral or unethical activities online. He proposes that mandating using real names is not always functional, since the requirement can be so easily circumvented. Rather, credibility can be enhanced online by a method of close reciprocal monitoring and "witnessing."

Michael Steinmann, Julia Shuster, Jeff Collmann, Sorin Adam Matei, Rochelle Tractenberg, Kevin FitzGerald, Greg Morgan, Douglas Richardson use the chapters of the book as a testing ground for a method of exploring ethical concerns in big data analysis. Focusing on privacy, they highlight the fact that all contributions raise and need to tackle privacy concerns, which suggests that big data research should have a more explicit method for dealing with ethical issues. To understand the ethical challenges that can arise from privacy concerns in Big Data, the authors elucidate how privacy in Big Data can be analyzed using two dimensions: (1) different *contexts* in which privacy is relevant and (2) different *principles* that specify the ethical meaning of privacy.

The final chapter of the book is an interview with Howard Rheingold, a pioneer of the idea of "virtual community" and of sociability enhanced by communication technology. The interview starts with a review of the promises, achieved or missed, of the cyberculture and of social media. It continues with a discussion of Rheingold's philosophy and methodology for facilitating credibility and trust online, in which he emphasizes the importance of critical thinking and of developing a system of intellectual and cultural "checks and balances."

The present volume hopes to create a bridgehead in an underexplored territory. Knowledge generating social media is here to stay, yet the manner in which it changes authorship, credibility, and trust is at best imperfectly known. Setting

principles, proposing algorithms or tools, and creating social methods for clarifying or addressing them should be a priority. We hope that the volume may bring some clarity to this process of setting priorities, in addition to being a substantive contributions to the emerging field of computational social science.

Acknowledgments Part of the research reported in this book and the editorial work were supported by the NSF award No. 1244708, KredibleNet.

References

Bertino, E., & Matei, S. (2014). *Roles, trust, and reputation in Social Media Knowledge Markets: Theories and Methods*. New York: Springer Publishing House.

Miller, M. (2012, March 22). 3 more studies examine Wikipedia's page 1. Google rankings [Blog]. Retrieved January 14, 2015, from http://searchenginewatch.com/sew/study/2163432/studies-examine-wikipedia-s-page-google-rankings.

Petrovic, S., Osborne, M., McCreadie, R., Macdonald, C., Ounis, I., & Shrimpton, L. (2013). Can Twitter replace Newswire for breaking news? In *Proceedings of the Seventh International AAAI Conference on Weblogs and Social Media ICWSM*.