

Letter from the Special Issue Editor

Over the past decade, social media has emerged as a dominant means through which people communicate. Even if we restrict ourselves to Twitter, it is estimated that people send about 400M tweets on a daily basis covering a variety of topics and opinions. While, individually, such tweets might or might not be very informative, many studies have now clearly established that collectively, such data contain a wealth of information that can be leveraged in various application contexts to bring new capabilities, use-cases, and value propositions. It is getting to be well recognized that technologies for collecting, monitoring and analyzing social media can bring transformative changes to variety of real-world domains. Already, many organizations in service-oriented industry such as hospitals and customer care monitor social media to determine public opinions about their services. Similar strategies are used by product companies (e.g., electronics, cell phones) to determine the public opinion about their products and by political parties and policy makers to assess the sentiment of the community. Department of homeland security in USA has a social media program that in addition to understanding public satisfaction with their services also uses social media as a sensor to detect emerging needs and events during crisis. Social media analysis has always been important to internet companies to understand user profiles in order to bring new customization and/or targeted advertising. Social media (specially when mixed with mobile computing and location based services) is a major driver for startup activity in the Silicon Valley and other IT hubs around the world. There are many new proposals, ideas, products that are attempting to seamlessly integrate social media with pervasive computing technologies including localization and sensing to bring an immersive experience and capabilities to users.

While social media monitoring and analysis offers numerous opportunities, it also poses a large number of technical hurdles. Challenges arise at every technological layer – consider, for instance, building a system or a capability that relies on social media. Only a very small part of the gargantuan amount of information may be relevant to the end-goal leading to the challenge of effective acquisition, filtering, and ranking of social media data. Another challenge arises due to relatively short form of the messages such as tweets and Facebook posts that are seldom well structured or grammatically correct limiting the effectiveness of standard NLP and information extraction mechanisms. Yet another layer of complexity arises due to the “big and fast” nature of such data – the amount of such data and the velocity at which it arrives (coupled with the near real-time need for analysis for certain applications) poses a significant challenge in building infrastructures that can scale. No doubt, social media monitoring and analysis is the driving force behind a large amount of research and innovations at all technology levels – infrastructure level where researchers are exploring hardware and software infrastructures that can support complex social media analysis, at the representation and analysis level, where researchers are exploring mechanisms that can provide valuable insights from social media data, and application level where researchers are exploring diverse applications and new uses of social media. This special issue consists of a set of articles from leading researchers exploring social media at different technology levels.

The special issue is roughly divided into four parts. We begin the bulletin by two articles that highlight experiences and challenges in exploiting social media analysis in the context of concrete applications from researchers who have built significant such systems.

In the article entitled “Social Media Analytics: The Kosmix Story”, the authors provide a glimpse of the “insider story” in building a large real-time social knowledge base entitled Social Genome in a commercial setting as part of Kosmix which was a bay area startup that was later acquired by Walmart. In the article entitled “The Architectural Implications of Social Media Analytics in Support of Crisis Informatics Research”, the authors highlight the software architecture and challenges in building large-scale systems for event monitoring on twitter to support crisis management. The article further address the key lessons learnt and the implication of social media to crisis informatics in the future.

The second set of papers describe variety of ideas and issues related to social media analytics.

“Social Media Analytics Research in MSR Search Labs” focuses on ongoing and current research within Microsoft Search Labs on modeling how information spreads and propagates through social networks and how people assimilate the information and form relationships. In the article entitled “Geospatial Footprints in Social Media: Towards GeoSocial Intelligence”, the authors focus on how geotagged social media collected through smart devices opens new opportunities for developing new geo-social systems which can help uncover how ideas flow from people to people and how people organize. The next two articles in the bulletin focus on the important problem of event identification in social media which is at the heart of much of the use-cases for social media. In the article entitled “Effective and Efficient Event Identification in Social Media”, the authors discuss the limitations of current solutions and describe new approaches to improving detection by increasing the set of features used for clustering as well as using a more informed event model that accounts for time decay. In the article entitled “Event Detection from Social Media Data”, the authors propose using concepts from emotional theories combined with Spatio-Temporal information to build a robust and scalable event detector.

The next set of papers deal with issues related to efficient processing of large social media data. In “Large Scale Tensor Decomposition: Algorithmic Developments and Applications”, the authors summarize recent algorithmic developments in scaling tensor decomposition to big data using map/reduce framework. Such tensor-based analysis is a core technique for analyzing social media data for interesting patterns and anomalies. In “Summarization via Pattern Utility and Ranking: a Novel Framework for Social Media Data Analytics”, the authors describe a new dynamic pattern driven approach to summarizing social networks and topologies that enables efficient processing of user-specific and topic-specific temporal analysis.

The final set of papers in the bulletin deal with new / novel emerging applications and new research opportunities for social media analytics. In “Some Research Opportunities on Twitter Advertisement”, the authors revisit the issue of social advertising which is omnipresent in social media. The authors discuss the new advertising opportunities introduced by Twitter to promote advertisements to targeted individuals and identify the research challenges/opportunities such a model promotes. “S3: A framework for Efficient Social Media Search in the Cyber Physical Systems” describes a new direction of research that seamlessly integrates sensors and cyber physical systems with social media. In particular, the authors describe a framework entitled S3 that supports social media search when queries may be a result of integrating the physical world with communication and computational devices as in a cyber physical system. In the article entitled “Building Social Life Networks”, the authors discuss a novel concept of social life networks that connect people with essential life resources. They identify key challenges in building such networks (viz., algebraic framework for situation modeling and recognition, context determination) and briefly describe their experience in building such systems.

As is often the case with Data Engineering Bulletins, the range of articles vary in the level of depth and treatment of the subject – while some papers focus more on vision and challenges that lie ahead, others describe technically mature approaches based on significant prior work by the authors. Irrespective of the nature of the papers, collectively they provide a good view of the state-of-the-art thoughts and research in the area of social media analytics.

Finally, I would like to acknowledge the generous help by Mehdi Sadri in following up with the authors and compiling the papers into the bulletin. Without his help, production of the bulletin would have been significantly more difficult.

Sharad Mehrotra
University of California, Irvine