Information Diffusion In Social Networks: Observing and Influencing Societal Interests*

Divyakant Agrawal Ceren Budak Amr El Abbadi
Department of Computer Science UCSB
Santa Barbara, CA 93106-5110, USA
Phone: +01-805-893-4385, Fax: +01-805-893-8553
{agrawal, cbudak, amr}@cs.ucsb.edu

1. INTRODUCTION

With hundreds of millions of users worldwide, social networks provide great opportunities for social connection, learning, political and social change, as well as individual entertainment and enhancement in a wide variety of forms. Because many social interactions currently take place in online networks, social scientists have access to unprecedented amounts of information about social interaction. Prior to online networks, these investigations required resource-intensive activities such as random trials, surveys, and manual data collection to gather even small data sets. Now, massive amounts of information about social networks and social interactions are recorded. This wealth of data can allow social scientists to study social interactions on a scale and at a level of detail that has never before been possible. In addition to providing a platform for scientists to observe social interactions in large scale, online social networks are also changing the very nature of social interactions. People now have ready access to almost inconceivably vast information repositories that are increasingly portable, accessible, and interactive in both delivery and formation. Basic human activities have changed as a result, and new possibilities have emerged. For instance, the process by which people locate, organize, and coordinate groups of individuals with shared interests, the number and nature of information and news sources available, and the ability to solicit and share opinions and ideas across various topics have all undergone dramatic change with the rise of social networks.

Social networks have already emerged as a significant medium for the widespread distribution of news and instructions in mass convergence events such as the 2008 U.S. Presidential Election [35], the 2009 Presidential election in Iran [31], and emergencies like the landfall of Hurricanes Ike and Gustav in the fall of 2008 [35]. Use of social networks such as Facebook and Twitter has also been noted as providing great ease during the recent demonstrations in Middle East [63]. In light of these notable outcomes, understanding information diffusion over online social networks is a critical research goal. This greater understanding can be achieved through data analysis, the development of reliable models that can predict outcomes of social processes, and ultimately the creation of applications that can shape the outcome of these processes. In this tutorial, we aim to provide an overview of such recent research based on a wide variety of techniques such as optimization algorithms, data mining, data streams covering a large number of problems such as influence spread maximization, misinformation limitation and study of trends in online social networks.

VLDB '11, August 29- September 3, 2011, Seattle, WA

2. TUTORIAL OUTLINE

Our tutorial is intended to span three hours, and is divided into four broad subsections. We now provide a high-level summary of the goals of these subsections where we give a broad survey of the recent research in the area, identifying different techniques and indicate open problems and challenges associated with such research.

2.1 Characterizing Social Networks

A clear understanding of information diffusion in online social networks can not be achieved without a clear understanding of significance and characteristics of social networks. Therefore, we start our tutorial with an overview of social networks identifying its significance and characteristics. We start our tutorial with a discussion on significance of online social networks in today's world [25, 31, 34, 35, 45, 63, 64, 74]. This discussion lays the groundwork for answering the following questions: How are people using social networks today? What is the use of social networks beyond their ability to connect people online? Are online social networks redefining social interactions as we know them? How can research in social networks facilitate a more effective use of social networks?

Given the motivation provided by many examples of successful uses of social networks, next we provide a survey of early measurement studies in social networks focusing on important characteristics that distinguish social networks from other networks [16, 51, 52, 73]. A few of such commonly known characteristics are power-law degree distribution, high clustering and positive assortativity [71]. A better understanding of these characteristics clarifies which studies from the literature applies to social networks and the new solutions needed. At the end of this section, our goal is to provide a clear picture of characteristics of social networks and theirs possible uses. Next we will focus on one of the most important uses of social networks, namely information diffusion.

2.2 Diffusion of Information or Opinions

In this part of the tutorial, we outline the techniques used in optimizing or facilitating information diffusion in social networks. There are a large number of problems that relate to diffusion of information in social networks. Unsurprisingly, all such problems, as different as they are in terms of the motivation behind them, rely on sub-problems and therefore techniques that are inherently very similar. We identify two example problem definitions through which a broad survey of techniques in recent research is provided. Namely, we study the problem of (i) maximizing spread of influence and (ii) minimizing spread of misinformation in social networks. The goal is to provide a summary of sub-problems that are common to the two and other similar problems. We outline a list of interesting techniques for each sub-problem. We begin with motivating the study of diffusion of information in social networks and outline some of the useful applications from research as well as the industry that are a byproduct of such studies. Next we delve into more detail about the sub-problems; (i) Diffusion model formation, (ii) formalization and optimization, (iii) large-scale data analysis, (iv) technique validation.

Diffusion Model Formation. Central to optimization problems relating to information diffusion is the problem of identifying the right diffusion model since use of an inaccurate model would result in optimization problems that are optimizing a non-existent problem. Therefore it is an important research goal to increase our understanding of information diffusion processes. To this end, we provide a survey of available models and address following questions: Can one find *one* model that fits most to all human interactions? Do different social networks require use of a different model? What are the necessary and sufficient parameters of an accurate model? How can we validate use of a specific model? How can one obtain data about the parameters? Given the intricacy of human interactions, finding the right diffusion model is still an open problem, even in the presence of large data sets available today. A large body of recent research that focus on diffusion of information assume a different model [12,21,28,39,47,48,50]. In addition to the recent research in information diffusion models, there is a large number of studies that model the interactions between the structure of a network and its role in disseminating diseases [3,20], ideas [32], data, adoption of technologies [5] or even its role in segregation of its members [61]. As part of this tutorial, we will give an overview of such studies as well as studies validating some of these models on real data.

Formalization and Optimization. Our first goal in this part of the tutorial is to provide a summary of different formulations for the problem of maximizing spread of influence in social network. In the second part, we will focus on the problem of limiting spread of misinformation. Choosing the right strategy for for facilitating the dissemination of a piece of information in an effective and efficient manner is a problem of great importance and recent studies in online social networks provide a great insight to this problem. Depending on the specific goal at hand, the exact definition of *influence* or the information diffusion model, a good strategy requires a different problem formalization and ultimately a different solution. For instance; Are the nodes of the network of same importance or adoption of a specific group more important? Is the timing of adoption of importance? An intensively studied problem formulation is: Given a number k, which set S with $|S| \leq k$ will eventually infect as many nodes of graph G as possible? This problem is also referred to as the identification of *influential users* or *opinion leaders* in a social network. We start this part of the tutorial giving a summary of a number of recent research [17, 18, 22, 39, 58, 68] on this optimization problem which was first introduced by Domingos and Richardson motivated by viral marketing [22,58]. This specific formulation is one of the many in the literature [1,4,15,17,18,39,40,44,49,68,69]. Next we give a summary of a number of such studies and discuss the advantages and disadvantages associated with each technique . We delve into more detail of one such data-centric study that identifies different types of *influencers* [11]. This work identifies 4 types of different actors based on the theories introduced in the popular book "The Tipping Point" by Malcolm Gladwell [27] and analyzes their correlation with successful diffusion scenarios in the blogosphere. This technique is orthogonal to the first set of studies we will discuss, in that it relies on data from real social networks and diffusion scenarios rather than depending on assumed models.

A characteristic common to the studies discussed so far in the tutorial is the assumption that information cascades of campaigns happen in isolation. Next we discuss a group of problem formulations that capture the notion of competing campaigns in a social network [8, 12, 14, 23, 41]. While the ease of information propagation in social networks can be very beneficial, it can also have disruptive effects. A number of examples of this sort are the spread of misinformation on swine flu in Twitter [53], exaggerated reports on a bomb attack in Grand Central and celebrities that are falsely claimed as being dead [33]. In the light of the large number of misinformation campaigns in social networks as well as the possible advantages of limiting such campaigns, we believe research on this topic is of great influence. Therefore, we explore the research in this area and discuss the desired characteristics of possible solutions. We specifically focus on the study that addresses the problem of *influence limitation* [12] where a "bad" campaign starts propagating from a certain node in the network and use the notion of limiting campaigns to counteract the effect of misinformation. This work also studies the influence limitation problem in the presence of missing data where the current states of nodes in the network are only known with a certain probability and show that prediction in this setting is a supermodular problem.

Large-Scale Data Analysis. No matter which technique is used in studying information diffusion, a theoretical approach that relies on optimization algorithms and modeling or a data-centric method, large-scale data analysis is a significant aspect of study as well as being a significant challenge. For the latter, large-scale data analysis *is* the method, while for the former it is needed for mining values of parameters of the model in use. In this part of the tutorial, we will give an overview of various data analysis techniques used in the literature and discuss which of those are applicable to study of social networks given the very large scale of its nature. With the increase of studies in social networks, there are a number of data sets available to researchers [24, 43, 46, 54]. As obtaining data and analyzing it to achieve a form from which one can draw conclusions is an important subproblem for studying information diffusion in social networks, we discuss the feasibility of a common repository of data for researchers in the area and discuss possible advantages and disadvantages associated with such repository.

Technique Validation. Next we focus on techniques to validate the techniques introduced in the modeling and optimization of information diffusion. Without a proper validation method, it is impossible to see the greater impact of the theoretical research in social networks on real social networks. A good validation involves testing the accuracy of such methods on real social networks, i.e. influence spread acquired in real social networks, as well as their efficiency, i.e. the resources needed or the complexity of the algorithm used. We focus on the challenges of technique validation.

2.3 Information Trends

As we will spend the first part of the tutorial to focus on recent research on information diffusion that focuses on diffusion of different ideas in isolation, for completeness purposes we will use the second part of the tutorial focusing on the interplay between different information campaigns and *trends* that result from such interplay. Social networks provide large-scale information infrastructures for people to discuss and exchange ideas about different topics. Detecting trends of such topics is of significant interest for many reasons. For one, it can be used to detect emergent or suspicious behavior in the network. They can also be viewed as a reflection of societal concerns or even as a consensus of collective decision making. Understanding how a community *decides* that a topic is trendy can help us better understand how ad-hoc communities are formed and how decisions are made in such communities. In general, constructing "useful" trend definitions and providing scalable solutions for them will contribute towards a better understanding of human interactions in the context of social media. With these notable outcomes in mind, we will provide a summary of research on trend analysis in social networks. We will focus on the following questions: What are the nature of trends in social networks? What are important dimensions of such trends? How can one efficiently analyze trends in social networks given the large scale of social networks?

Trends in social networks have recently been a major focus of interest among researchers studying them from perspectives such as temporal [48] and geographical dimensions [59,60,65] or sentiment embedded in the shared information [9, 10]. A similar interest can be observed in industry. For instance, Twitter trends [66] have been a testament to societal concerns, to such an extent that when there was interest in Wikileaks and the hashtag #wikileaks did not appear in the trends list in Twitter, there was substantial discussion upon which Twitter had to make an official announcement stating they have not excluded #wikileaks from the trending topics [70]. The large number of companies reporting trends in Twitter is another testament to the importance of trends [36, 37, 66]. We will give a summary of such studies from both research and industry.

In today's social networks where users are highly influenced by their friends, trend definitions that reach beyond simple heavy-hitters approaches to integrate the importance of such flow of influence can be of great benefit. Since information diffusion on a social network is a substantial part of the process that creates the information trends, properties that are defined in this context are of significant interest. We delve into one recent study that analyzes social networks from the perspective of the structural properties of the graph that creates such trends. [13]. This work considers the problem of identifying the number of connected (and disconnected) pairs of users in a social network that are discussing a specific topic named *correlated trends* (and uncorrelated trends). Through this study we will focus on methods of validation of *value proposition* of new techniques and problem formulations. Since large-scale data analysis an important subproblem of information trend analysis, we will focus on challenges associated with large-scale data analysis as well as possible solutions.

2.4 Other Research in Social Networks

Even though the focus of this tutorial is on information diffusion, we note that there are a large number of exciting and influential studies relating to other aspects of social networks. In the last section of our tutorial, we will give an overview of such studies. With the prevalence of social networks, researchers have new opportunities to study how groups of people come together and form communities. A better understanding of communities in social networks provide great benefits. Therefore we give an overview of state-of-the-art of research on community formation and evolution [2, 6, 11, 19, 26, 29, 42, 55–57, 62, 75]. Another important area of research in social networks is spam detection and elimination since the decentralized nature of social networks today make them vulnerable to such behavior. We therefore give a summary of recent research in the area of spam detection and elimination [7,13,30,38,67,72]. With this discussion, we further highlight the challenges associated with the rise of online social networks.

3. GOALS OF THE TUTORIAL

3.1 Learning Outcomes

Following are the learning outcomes from this tutorial:

- Overview of significance of social networks as well as their important characteristics.
- State-of-the-art in data analysis and optimization problems in the area of information diffusion in social networks. Summary of current research projects and future research directions.
- Identifying common subproblems in a large number of recent studies.
- Understanding various social interaction/information diffusion models.
- Understanding the advantages and shortcomings of optimization algorithms in the area of information diffusion in social networks compared to data-centric models
- Understanding the appropriate validation methods and using such techniques to navigate through the large number of studies in social networks to reach the specific goal in mind.
- A list of open research challenges in social networks that must be addressed to ensure the applicability of research results to real life and retrieving effective results as well as an overview of other significant sub-categories of studies in social networks in addition to information diffusion.

3.2 Intended Audience

This tutorial is intended to benefit researchers and industry in the broad area of social networks. Our tutorial would benefit both researchers as well as the industry interested in understanding information diffusion in social networks since a survey of the current research and applications is essential for building new models and algorithms as well as systems.

4. BIOGRAPHICAL SKETCHES

Divyakant Agrawal is currently a Professor of the Department of Computer Science at the University of California, Santa Barbara. He was a visiting researcher at IBM Almaden Research Center, Visiting Senior Research Scientist position at NEC Computing and Communication Research Laboratories. From January 2006 through December 2007, Prof. Agrawal served as VP of Data Solutions and Advertising Systems at the Internet Search Company ASK.com. He has served in the program committees of many leading conferences and is the PC Chair of SIGMOD 2010. He currently serves on the editorial boards of the VLDB journal and the Proceedings of the VLDB. Prof. Agrawal's research expertise is in the areas of database systems, distributed computing, data warehousing, and large-scale information systems.

Ceren Budak is currently a PhD Candidate at the Department of Computer Science, University of California Santa Barbara. Her research interests lie in the area of information diffusion in social networks.

Amr El Abbadi is currently a Professor and Chair of the Department of Computer Science at the University of California, Santa

Barbara. He has held visiting professor positions at the University of Campinas in Brazil, IBM Almaden Research Center, the Swedish Institute of Computer Science in Stockholm, and at IRISA at the University of Rennes in France. He was Vice Chair of ICDCS 1999, Vice Chair for ICDE 2002, the Americas Program Chair for VLDB 2000, and is the Program Chair for SIGSPATIAL 2010. He served as a board member of the VLDB Endowment from 2002 to 2008. In 2007, Prof. El Abbadi received the UCSB Senate Oustanding Mentorship Award for his excellence in mentoring graduate students. His research interests lie in the broad area of scalable database and distributed systems.

5. REFERENCES

[1] A. G. 0002, F. Bonchi, and L. V. S. Lakshmanan. Discovering leaders from community actions. In *CIKM*, pages 499–508, 2008.

[2] L. Backstrom, D. P. Huttenlocher, J. M. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD*, pages 44–54, 2006.

[3] N. Bailey. *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975.

[4] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *WSDM*, pages 65–74, 2011.

[5] F. M. Bass. A new product growth for model consumer durables. *Manage. Sci.*, 50:1825–1832, December 2004.

[6] J. Baumes, M. Goldberg, and M. Magdon-Ismail. Efficient identification of overlapping communities. In *In IEEE International Conference on Intelligence and Security Informatics (ISI)*, pages 27–36, 2005.

[7] F. Benevenuto, G. Magno, T. Rodrigues, and V. Almeida. Detecting spammers on twitter. In *Proceedings of the 7th Annual Collaboration, Electronic messaging, Anti-Abuse and Spam Conference (CEAS)*, 2010.

[8] S. Bharathi, D. Kempe, and M. Salek. Competitive influence maximization in social networks. In *WINE*, pages 306–311, 2007.

[9] J. Bollen, H. Mao, and X.-J. Zeng. Twitter mood predicts the stock market. *CoRR*, abs/1010.3003, 2010.

[10] J. Bollen, A. Pepe, and H. Mao. Modeling public mood and emotion: Twitter sentiment and socio-economic phenomena. *CoRR*, abs/0911.1583, 2009.

[11] C. Budak, D. Agrawal, and A. El Abbadi. Where the blogs tip: connectors, mavens, salesmen and translators of the blogosphere. In *SIGKDD Workshop on Social Media Analytics*, 2010.

[12] C. Budak, D. Agrawal, and A. El Abbadi. Limiting the spread of misinformation in social networks. In *WWW*, 2011.

[13] C. Budak, D. Agrawal, and A. El Abbadi. Structural trend analysis for online social networks. Technical Report UCSB/CS-2011-04, UCSB, 2011.

[14] T. Carnes, C. Nagarajan, S. M. Wild, and A. van Zuylen. Maximizing influence in a competitive social network: a follower's perspective. In *ICEC '07*, pages 351–360, New York, NY, USA, 2007. ACM.

[15] M. Cha, H. Haddadi, F. Benevenuto, and P. K. Gummadi. Measuring user influence in twitter: The million follower fallacy. In *ICWSM*, 2010.

[16] M. Cha, A. Mislove, B. Adams, and K. P. Gummadi. Characterizing social cascades in flickr. In *Proceedings of the first workshop on Online social networks*, WOSP '08, pages 13–18, New York, NY, USA, 2008. ACM.

[17] W. Chen, C. Wang, and Y. Wang. Scalable influence maximization for prevalent viral marketing in large-scale social networks. In *KDD*, pages 1029–1038, 2010.

[18] W. Chen, Y.Wang, and S. Yang. Efficient influence maximization in social networks. In *Proceedings of the 15th ACM International Conference on Knowledge Discovery and Data Mining*, pages 199–208, 2009.

[19] A. Clauset. Finding local community structure in networks. *Physical Review E (Statistical, Nonlinear, and Soft Matter Physics)*, 72(2):026132, 2005.

[20] O. Diekmann and J. Heesterbeek. *Mathematical epidemiology of infectious diseases: model building, analysis, and interpretation*. Wiley, 2000.

[21] P. S. Dodds and D. J. Watts. A generalized model of social and biological contagion. *Journal of Theoretical Biology*, 232:587–604, 2005.

[22] P. Domingos and M. Richardson. Mining the network value of customers. In *Proceedings of the 7th ACM International Conference on Knowledge Discovery and Data Mining*, pages 57–66, 2001.

[23] P. Dubey, R. Garg, and B. D. Meyer. Competing for customers in a social network: The quasi-linear case. In *WINE*, pages 162–173, 2006.

[24] C. for Computational Analysis of Social and O. S. (CASOS). Casos networks. `http://www.casos.cs.cmu.edu/computational_tools/data2.php`.

[25] S. Garriss, M. Kaminsky, M. J. Freedman, B. Karp, D. Mazires, and H. Yu. Re: Reliable email. In *In Proc. NSDI*, pages 297–310, 2006.

[26] R. Ghosh and K. Lerman. Community detection using a measure of global influence. In *Proceedings of the 2nd KDD Workshop on Social Network Analysis (SNAKDD'08)*, 2008.

[27] M. Gladwell. *The Tipping Point: How Little Things Can Make a Big Difference*. Back Bay Books, January 2002.

[28] M. Gomez-Rodriguez, J. Leskovec, and A. Krause. Inferring networks of diffusion and influence. In *KDD*, pages 1019–1028, 2010.

[29] S. Gregory. An algorithm to find overlapping community structure in networks. *Knowledge Discovery in Databases: PKDD 2007*, pages 91–102, 2007.

[30] C. Grier, K. Thomas, V. Paxson, and M. Zhang. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, CCS '10, pages 27–37, New York, NY, USA, 2010. ACM.

[31] L. Grossman. Iran protests: Twitter, the medium of the movement. *Time (online)*, June 2009.

[32] R. Guimera and L. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, 2005.

[33] K. M. Heussner. Enough already! 7 twitter hoaxes and half-truths. *ABC News*, January 2010.

[34] J. Horrigan and L. Rainie. When facing a tough decision, 60 million americans now seek the internet's help: The internet's growing role in life's major moments [retrieved october 13, 2006]. `http://pewresearch.org/obdeck/?ObDeckID=19`, 2006.

[35] A. L. Hughes and L. Palen. Twitter adoption and use in mass

convergence and emergency events. In *Proceedings of the 6th International Information Systems for Crisis Response and Management Conference*, 2009.

[36] Tweetstats. http://tweetstats.com/trends.

[37] Trendistic. http://trendistic.com/.

[38] T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Commun. ACM*, 50:94–100, October 2007.

[39] D. Kempe, J. M. Kleinberg, and É. Tardos. Maximizing the spread of influence through a social network. In *Proceedings of the Ninth ACM International Conference on Knowledge Discovery and Data Mining*, pages 137–146, 2003.

[40] M. Kimura, K. Saito, R. Nakano, and H. Motoda. *Social Computing and Behavioral Modeling*, chapter Finding Influential Nodes in a Social Network from Information Diffusion Data. Springer US, 2009.

[41] J. Kostka, Y. A. Oswald, and R. Wattenhofer. Word of mouth: Rumor dissemination in social networks. In *SIROCCO*, pages 185–196, 2008.

[42] R. Kumar, J. Novak, and A. Tomkins. Structure and evolution of online social networks. In *KDD*, pages 611–617, 2006.

[43] H. Kwak, C. Lee, H. Park, and S. Moon. What is twitter, a social network or a news media? http://an.kaist.ac.kr/traces/WWW2010.html.

[44] H. Kwak, C. Lee, H. Park, and S. Moon. What is Twitter, a social network or a news media? In *WWW '10*, pages 591–600, 2010.

[45] C.-A. La and P. Michiardi. Characterizing user mobility in second life. In *Proceedings of the first workshop on Online social networks*, WOSP '08, pages 79–84, New York, NY, USA, 2008. ACM.

[46] J. Leskovec. Stanford large network dataset collection. http://snap.stanford.edu/data/index.html, 2009.

[47] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *TWEB*, 1(1), 2007.

[48] J. Leskovec, L. Backstrom, and J. Kleinberg. Meme-tracking and the dynamics of the news cycle. In *KDD '09*, pages 497–506, 2009.

[49] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. M. VanBriesen, and N. S. Glance. Cost-effective outbreak detection in networks. In *KDD*, pages 420–429, 2007.

[50] S. Liu, L. Ying, and S. Shakkottai. Influence maximization in social networks: An ising-model-based approach. In *Allerton*, 2010.

[51] A. Mislove, H. S. Koppula, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Growth of the flickr social network. In *Proceedings of the first workshop on Online social networks*, WOSP '08, pages 25–30, New York, NY, USA, 2008. ACM.

[52] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and analysis of online social networks. In *Proceedings of the 7th ACM SIGCOMM conference on Internet measurement*, IMC '07, pages 29–42, New York, NY, USA, 2007. ACM.

[53] E. Morozov. Swine flu: Twitter's power to misinform. *Foreign Policy*, April 2009.

[54] M. Newman. Network data. http://www-personal.umich.edu/~mejn/netdata/.

[55] M. E. J. Newman. Fast algorithm for detecting community structure in networks. *Physical Review E*, 69, September 2003.

[56] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, June 2006.

[57] A. Pothen, H. D. Simon, and K.-P. Liou. Partitioning sparse matrices with eigenvectors of graphs. *SIAM J. Matrix Anal. Appl.*, 11(3):430–452, 1990.

[58] M. Richardson and P. Domingos. Mining knowledge-sharing sites for viral marketing. In *Proceedings of the 8th ACM International Conference on Knowledge Discovery and Data Mining*, pages 61–70, 2002.

[59] T. Sakaki, M. Okazaki, and Y. Matsuo. Earthquake shakes twitter users: real-time event detection by social sensors. In *WWW*, pages 851–860, 2010.

[60] J. Sankaranarayanan, H. Samet, B. E. Teitler, M. D. Lieberman, and J. Sperling. Twitterstand: news in tweets. In *GIS '09*, pages 42–51, 2009.

[61] T. Schelling. *Micromotives and macrobehavior*. WW Norton & Company, 2006.

[62] J. P. Scott. *Social Network Analysis: A Handbook*. SAGE Publications, January 2000.

[63] C. Smith. Egypt's facebook revolution: Wael ghonim thanks the social network. *The Huffington Post*, February 2011.

[64] G. Swamynathan, C. Wilson, B. Boe, K. Almeroth, and B. Y. Zhao. Do social networks improve e-commerce?: a study on social marketplaces. In *Proceedings of the first workshop on Online social networks*, WOSP '08, pages 1–6, New York, NY, USA, 2008. ACM.

[65] B. E. Teitler, M. D. Lieberman, D. Panozzo, J. Sankaranarayanan, H. Samet, and J. Sperling. Newsstand: a new view on news. In *GIS '08*, pages 1–10, 2008.

[66] Twitter. http://www.twitter.com.

[67] A. H. Wang. Don't follow me - spam detection in twitter. In *SECRYPT*, pages 142–151, 2010.

[68] Y. Wang, G. Cong, G. Song, and K. Xie. Community-based greedy algorithm for mining top-k influential nodes in mobile social networks. In *Proceedings of the 16th ACM International Conference on Knowledge Discovery and Data Mining*, 2010.

[69] J. Weng, E.-P. Lim, J. Jiang, and Q. He. Twitterrank: finding topic-sensitive influential twitterers. In *WSDM*, pages 261–270, 2010.

[70] Twitter: We are not keeping wikileaks out of trending topics. http://mashable.com/2010/12/06/wikileaks-twitter-censorship/.

[71] C. Wilson, B. Boe, A. Sala, K. P. Puttaswamy, and B. Y. Zhao. User interactions in social networks and their implications. In *EuroSys*, pages 205–218, 2009.

[72] S. Yardi, D. M. Romero, G. Schoenebeck, and D. Boyd. Detecting spam in a twitter network. *First Monday*, 15(1), 2010.

[73] Y. yeol Ahn, S. Han, H. Kwak, Y. ho Eom, S. Moon, and H. Jeong. Analysis of topological characteristics of huge online social networking services. In *WWW07*, pages 835–844. ACM, 2007.

[74] H. Yu, M. Kaminsky, P. B. Gibbons, and A. Flaxman. Sybilguard: defending against sybil attacks via social networks. *SIGCOMM Comput. Commun. Rev.*, 36:267–278, August 2006.

[75] S. Zhang, R. Wang, and X. Zhang. Identification of overlapping community structure in complex networks using fuzzy c-means clustering. *Physica A: Statistical Mechanics and its Applications*, 374(1):483–490, 2007.