

Social Media Analysis

Proposed Syllabus

Prerequisites: CSCI 561 (Introduction to AI)

Time: Spring 2012, Mondays and Wednesdays at 3:30-5pm

Instructors: Professors Kristina Lerman (lerman@isi.edu) and Andrew Gordon (gordon@ict.usc.edu)

Office: ISI 932

Course Introduction

The biggest story of the last few years has been the phenomenal growth of social media, and the technological, social and political transformations that accompanied it. Social media sparked an information revolution by putting knowledge production and communication tools in the hands of the masses. Today on sites such as Twitter, Flickr, and YouTube, large numbers of users publish rich content, annotate it with descriptive metadata, and engage in discussions and collaborations with others. Social media promises to transform how we create and use knowledge, respond to disasters, monitor environment, manage resources, and interact with the world and one another. Social media offers new research opportunities and challenges. We will examine several topics, including social network analysis, information flow and learning from tagging, and show how AI, linguistic and statistical methods were developed to study these topics.

Course Requirements

There are no required text books. The reading material is based on recently published technical papers available via the ACM/IEEE/Springer digital libraries. All USC students have automatic access to these digital archives.

Grading

The class will run as a seminar course with student participation (30% of the grade) and weekly quizzes (30% of the grade). Instructors will introduce each topic after Week 3, and students will choose a research paper from the syllabus and present in-depth report about it. This will count for the participation grade. An integral part of the course is the class project (40% of the grade) using real-world social media data sets.

Topics and Readings

Week 1: Introduction: Social media, computational sociology and data mining?

- Kwak, H., Lee, C., Park, H., and Moon, S. 2010 What is Twitter, a Social Network or a News Media? In *Proceedings of 19th World-Wide Web (WWW) Conference*
- Innovative applications: Collaborative mapping (VGI), Ushahidi

Week 2: Influence

- Cha, M., Haddadiy, H., Benevenuto, F., and Gummadi, K.P. 2010 Measuring User Influence in Twitter: The Million Follower Fallacy, In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*.
- Ghosh, R., and Lerman, K. 2010. Predicting Influential Users in Online Social Networks. In *Proceedings of KDD workshop on Social Network Analysis (SNA-KDD)*, July.
- Lerman, K and Ghosh, R. 2011 Parametrized Centrality for Network Analysis.

Week 3: Social tagging and folksonomies

- Golder, S. and Huberman, B. 2005. The Structure of Collaborative Tagging Systems.
- Marlow, C., Naaman, M., Boyd, and Davis, M. 2006. Ht06, tagging paper, taxonomy, flickr, academic article, toread. In *Proceedings of Hypertext 2006*, New York. ACM, New York: ACM Press.
- Plangprasopchok, A. and Lerman, K. Exploiting Social Annotation for Automatic Resource Discovery, in *Proc. of AAAI Workshop on Information Integration on the Web*, 2007.
- Mika, P. Ontologies are us: a unified model of social networks and semantics. 2007 In *Selected Papers from the International Semantic Web Conference, International Semantic Web Conference (ISWC2005)*, Vol. 5, No. 1pp. 5-15.
- Plangprasopchok, A. Lerman, K., and Getoor, L. 2010 Growing a tree in a forest: constructing folksonomies by integrating structured metadata. In *KDD*
- Schmitz, P. 2006 Inducing Ontologies from Flickr Tags, in *Proc. of WWW Collaborative Web Tagging workshop*.

Week 4: Dynamics of social media

- Gruhl, D., Guha, R., Nowell, D.L. and Tomkins, A. 2004. Information diffusion through blogspace, In *Proceedings of the 13th international conference on World Wide Web*, pp. 491-501.
- J. Leskovec, L. Backstrom, J. Kleinberg. 2009 Meme-tracking and the dynamics of the news cycle. In *Proc. 15th ACM SIGKDD Intl. Conf. on Knowledge Discovery and Data Mining*.
- Kamath, K. Y. and Caverlee, J. 2011 Transient Crowd Discovery on the Real-Time Social Web, In *Proceedings of Web Search and Data Mining Conference*.
- Wu, F. and Huberman, B.A. 2010 Opinion formation under costly expression, *ACM Trans. Intell. Syst. Technol.*, Vol. 1

Week 5: Information cascades and social epidemics

- Romero, D. M., Meeder, B. and Kleinberg, J. 2011. Differences in the Mechanics of Information Diffusion Across Topics: Idioms, Political Hashtags, and Complex Contagion on Twitter, In *Proceedings of World Wide Web Conference*.
- Lerman, K and Ghosh, R. 2010 "Information Contagion: an Empirical Study of the Spread of News on Digg and Twitter Social Networks", In *Proceedings of 4th International Conference on Weblogs and Social Media (ICWSM)*
- Ver Steeg, G., Lerman, K and Ghosh, R. 2011 "What stops social epidemics", in *Proc. 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*
- Liben-Nowell, D. and Kleinberg, J. 2008 "Tracing information flow on a global scale using Internet chain-letter data" *Proceedings of the National Academy of Sciences*, Vol. 105, No. 12, pp. 4633-4638.

Week 6: Wikipedia analysis

- Gabrilovich, E. and Markovitch, S. 2007. "Computing Semantic Relatedness using Wikipedia-based Explicit Semantic Analysis", *Proceedings of The 20th International Joint Conference on Artificial Intelligence (IJCAI)*, Hyderabad, India

- Suchanek, F., Kasneci, G. and Weikum, G. 2007 YAGO: A Large Ontology from Wikipedia and WordNet, Web Semantics: Science, Services and Agents on the World Wide Web, Volume 6, Issue 3, *World Wide Web Conference 2007 Semantic Web Track*, September 2008, Pages 203-217

Week 7: Content analysis: sentiments and topics

- Pang, B., Lee, L. and Vaithyanathan, S. 2002. Thumbs up? Sentiment Classification using Machine Learning Techniques. In *Proceedings of the ACL-02 conference on Empirical methods in natural language processing*, pages 79-86, Morristown, NJ.
- Turney, P. D. 2002. Thumbs Up or Thumbs Down? Semantic Orientation Applied to Unsupervised Classification of Reviews" In *ACL '02: Proceedings of the 40th Annual Meeting on Association for Computational Linguistics*, pages 417-424, Morristown, NJ.
- Pang, B. and Lee, L. 2005. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. In *ACL '05: Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics*, pages 115-124, Morristown, NJ.

Week 8: Query logs and real time search

- Ginsberg, J., Mohebbi, M.H., Patel, R.S., Brammer, L., Smolinski, M.S. & Brilliant, L. 2009 Detecting influenza epidemics using search engine query data. *Nature* Vol 457, 19.
- Kuthuria, A., Jansen, B.J., Hafernik, C. (2010) K-means Clustering to Determine User Intent of Web Queries. *Journal of Internet Research: Electronic Networking Applications and Policy*. 20(5), 563-581.
- Jansen, B.J., Liu, Z., Weaver, C., Campbell, G. and Gregg, M. (Forthcoming) Real Time Search on the Web: Queries, Topics, and Economic Value. *Information Processing & Management*.

Week 9: Geospatial social networks

- Cheng, Z., Caverlee, J. and Lee, K. You Are Where You Tweet: A Content-Based Approach to Geo-locating Twitter Users. *19th ACM International Conference on Information and Knowledge Management (CIKM)*
- Intagorn, S., Plangprasopchok, A. and Lerman, K. 2010. Harvesting Geospatial Knowledge from Social Metadata. In *Proceedings of 7th International Conference on Information Systems for Crisis Response and Management*.

Week 10: Modeling social media: physics perspective

- Crane, R. and Sornette, D. 2008 "Robust Dynamic Classes Revealed by measuring the response function of a social system." *Proceedings of the National Academy of Sciences*, Vol. 105, No. 41. (14 October 2008), pp. 15649-15653.
- Hogg, T. and Lerman, K. 2009 "Stochastic Models of User-Contributory Web Sites", In *Proceedings of 3rd International Conference on Weblogs and Social Media (ICWSM)*.
- Hogg, T. and Szabo, G. 2009 Diversity of User Activity and Content Quality in Online Communities, In *Proceedings of 3rd International Conference on Weblogs and Social Media (ICWSM)*.

Week 11: Modeling social media: CS perspective

- Meeyoung Cha, Haewoon Kwak, Pablo Rodriguez, Yong-Yeol Ahn, and Sue Moon, 2007 I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System, In Proc. of Usenix/ACM SIGCOMM Internet Measurement Conference (IMC), San Diego, CA.
- Morrency, L.P., YouTube analysis.

Week 12: Predicting the future

- Asur, S. and Huberman, B. A., 2010, Predicting the Future with Social Media.
- Watts, D., 2010, "What can search predict?"
- Szabo, G. and Huberman, B. A., 2009. Predicting the Popularity of Online Content, *Social Science Research Network Working Paper Series*

Week 13: Ethical issues: Privacy

- Zheleva, E. and Getoor, L. 2009 To join or not to join : the illusion of privacy in social networks with mixed public and private user profiles, in *Proc. of International World Wide Web Conference*.

Week 14: Ethical issues: spam and other types of exploitation

- Yardi, S., Romero, D., and Boyd, D., 2010. Detecting spam in a Twitter network. *First Monday* **15**.
- Grier, C., Thomas, K., Paxson, V., and Zhang, M. 2010. @spam: the underground on 140 characters or less. In *Proceedings of the 17th ACM conference on Computer and communications security*, pp. 27-37.

Week 15: Project presentations

Class project

The class project requires students to conduct an independent research project in the context of social media. Students will have access to many publicly available data sets. They start by identifying some broad questions they want to answer. After surveying the relevant literature, they will explore the question on real-world data sets. The end product of the project is a paper that will be orally presented in class.

Statement on Academic Integrity

USC seeks to maintain an optimal learning environment. General principles of academic honesty include the concept of respect for the intellectual property of others, the expectation that individual work will be submitted unless otherwise allowed by an instructor, and the obligations both to protect one's own academic work from misuse by others as well as to avoid using another's work as one's own. All students are expected to understand and abide by these principles. Scampus, the Student Guidebook, contains the Student Conduct Code in Section 11.00, while the recommended sanctions are located in Appendix A:
<http://www.usc.edu/dept/publications/SCAMPUS/gov/>. Students will be referred to the Office of Student Judicial Affairs and Community Standards for further review, should there be any suspicion of academic dishonesty. The Review process can be found at: <http://www.usc.edu/studentaffairs/SJACS/>.