CS599 : Special Topics : Social Media Analysis

Course Syllabus

Prerequisites: None.
Time: Fall 2014, Mondays and Wednesdays at 4-5:20pm, KAP146
Instructors: Professors Kristina Lerman (lerman@isi.edu)
Office: ISI 932

Course Introduction

The phenomenal growth of social media has transformed the social, political, and technological landscape. 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, Facebook, and YouTube, large numbers of people publish rich content, annotate it with descriptive metadata, communicate and collaborate 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. By exposing individual and collective behavior, social media delivers large quantities of social data for analysis, offering new research opportunities and challenges.

This course will examine topics in social data analysis, including influence and centrality, information diffusion, sentiment analysis, modeling collective dynamics and show how AI, social network analysis, and statistical methods can be used to study these topics. While there are no prerequisites, I expect students to be proficient in programming, algorithms and data structures, and have taken college level or above courses in linear algebra and statistics. AI and machine learning coursework is a plus.

Course Requirements

There are no required textbooks. 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 and presentations (30% of the grade) and weekly quizzes (30% of the grade). An integral part of the course is the class project (40% of the grade) using real-world social media data.

Statement for Students with Disabilities

Any student requesting academic accommodations based on a disability is required to register with Disability Services and Programs (DSP) each semester. A letter of verification for approved accommodations can be obtained from DSP. Please be sure the letter is delivered to me (or to TA) as early in the semester as possible. DSP is located in STU 301 and is open 8:30 a.m.–5:00 p.m., Monday through Friday. The phone number for DSP is (213) 740-0776.

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/student-affairs/SJACS/.

Emergency Preparedness/Course Continuity in a Crisis
In case of a declared emergency if travel to campus is not feasible, USC executive leadership will announce an electronic way for instructors to teach students in their residence halls or homes using a combination of Blackboard, teleconferencing, and other technologies.

Topics and Readings

- **Week 1: August 25**
  - **Topic: Course Introduction**

- **Week 1: August 27**
  - **Topic: Phenomenology of social media**
  - **Readings**:
    1. Lerman, K. (2007) Social Information Processing in Social News Aggregation *IEEE Internet Computing: special issue on Social Search*, 11(6):16--28. 2007.
    2. Wilkinson, D. 2008 "Strong regularities in online peer production" In *EC '08: Proceedings of the 9th ACM conference on Electronic commerce,* pp. 302-309.
    3. A. Anagnostopoulos, R. Kumar, M. Mahdian, 2008 "Influence and correlation in social networks", In *Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pp. 7-15.

- **Week 2: September 1**
  - **Labor Day**

- **Week 2: September 3**
  - **Topic: Network Analysis Basics**
  - **Readings:**
    1. A. L. Barabasi *Network Science*, Chapters 2 and 4.
    2. D. Austin, "It's a small world afterall" http://www.ams.org/samplings/feature-column/fc-2012-08
    3. [optional] D Easley and J Kleinberg. *Networks, Crowds, and Markets: Reasoning about a Highly Connected World*. Cambridge University Press, 2010. Chapter 13 on "The Structure of the Web"
    4. [optional] L Backstrom, P Boldi, M Rosa, J Ugander, S Vigna. "Four Degrees of Separation," 2012

- **Week 3: September 8**
  - **Topic: Topic Analysis Basics**
  - **Readings:**
    1. Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM*, 55(4):77-84. http://www.cs.princeton.edu/~blei/papers/Blei2012.pdf
    2. Yehuda Koren, Robert Bell and Chris Volinsky. Matrix Factorization Techniques For Recommender Systems. In Journal of Computer, 2009.
    3. [optional] http://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/

- **Week 3: September 10**
  - **Topic: Sentiment Analysis**
  - **Readings:**

1. S. O. Sood and L. Vasserman. "ESSE: Exploring Mood on the Web", In *ICWSM* 2009.
2. S. Golder and M. Macy, Diurnal and Seasonal Mood Vary with Work, Sleep, and Daylength Across Diverse Cultures, *Science* Vol. 333 no. 6051 pp. 1878-1881, 2011.
3. A Pak and P Paroubek. Twitter as a corpus for sentiment analysis and opinion mining. *Proceedings of International Conference on Language Resources and Evaluation (LREC-2010)*, Valletta, Malta, May 17-23, 2010.
4. [optional] B Pang and L Lee. Seeing stars: Exploiting class relationships for sentiment categorization with respect to rating scales. Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics (ACL 2005)

- **Quiz 1**

- **Week 4: September 15**
  - **Topic: Influence and Centrality in Social Networks**
  - **Readings:**
    1. Bonacich, P. 1987 "Power and Centrality, a family of measures" *The American Journal of Sociology*, Vol. 92, No. 5.
    2. M. Franceschetti 2011 "PageRank: standing on the shoulders of giants" *Commun. ACM*, Vol. 54, pp. 92-101.
    3. [optional] Freeman, L. 1979 "Centrality in Social Networks: Conceptual Clarification", *Social Networks* 1, No. 3.

- **Week 4: September 17**
  - **Topic: Influence and Centrality in Social Networks**
  - **Readings**:

    1. E Bakshy, J. M. Hofman, W. A. Mason, D. J. Watts. 2011 "Everyone's an influencer: quantifying influence on Twitter" In *Proceedings of Int. Conf. on Web Search and Data Mining (WSDM)*
    2. Cha, M., Haddadi, 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)*.
    3. 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.
    - **Quiz 2**

- **Week 5: September 22**
  - **Topic: Wikipedia knowledge extraction**
  - **Readings:**
    1. F Suchanek, G Kasneci and G Weikum. YAGO: A Large Ontology from Wikipedia and WordNet. *Journal of Web Semantics: Science, Services and Agents on the World Wide Web*, Volume 6, Issue 3, 2008

2. E Gabrilovich and S Markovitch. Computing semantic relatedness using Wikipedia-based explicit semantic analysis. *Proceedings of the 20th International Joint Conference on Artificial Intelligence*, 2007.
3. Arazy, O. F. E. R., Morgan, W., and Patterson, R. (2006). Wisdom of the crowds: Decentralized knowledge construction in wikipedia. *Social Science Research Network Working Paper Series*.

- **Week 5: September 24**
  - **Topic: Search query logs**
  - **Readings:**
    1. Kuthuria, A., Jansen, B.J., Hafernik, C., Spink, A. (2010) Classifying user intent of web queries using k-means clustering. Journal of Internet Research: Electronic Networking Applications and Policy. 20(5), 563-581.
    2. M Pasca (2007) Weakly-supervised discovery of named entities using web search queries. Proceedings of the 16th ACM conference on Conference on information and knowledge management (CIKM'07) November 6–8, 2007, Lisbon, Portugal.
    3. M Strohmaier, P Prettenhofer and M Kröll. Acquiring knowledge about human goals from search query logs. Information Processing and Management. 2011.
    - **Quiz 3**

- **Week 6: September 29**
  - **Topic: Information diffusion**
  - **Readings:**
    1. Goel, S, Watts, D and Goldstein, D.G. "The structure of online diffusion networks", In *Proc. Electronic Commerce* 2012.
    2. J Borge-Holthoefer, R Banos, S Gonzalez-Bailon, and Y Moreno, "Cascading Behavior in Complex socio-technical networks", Journal of Complex Networks, 2013
    3. Y Wang, D Chakrabarti, C Wang, C Faloutsos, "Epidemic Spreading in Real Networks: An Eigenvalue Viewpoint", In *Proc SRDS* 2003.

- **Week 6: October 1**
  - **Topic: Information diffusion**
  - **Readings:**
    1. 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)*
    2. 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*.
    3. N. Hodas and K. Lerman, "How limited visibility and divided attention constrain social contagion." In *Proc. Social Computing*, 2012.

- **Quiz 4**

- **Week 7: October 6**
  - **Topic: Social ties and information diffusion**
  - **Readings:**
    1. M Granovetter, "The Strength of weak ties" *American Journal of Sociology*, Vol. 78, No. 6. (1973)
    2. J. P. Onnela, J. Saramäki, J. Hyvönen, G. Szabó, D. Lazer, K. Kaski, J. Kertész, A. L. Barabási, "Structure and tie strength in mobile communication networks", *Proceedings of the National Academy of Sciences*, Vol. 104, No. 18. (01 May 2007).
    3. Bakshy, E et al. The role of social networks in information diffusion", in *WWW*, 2012.

- **Week 7: October 8**
  - **Topic: Social ties and link prediction**
  - **Readings:**
    1. D Liben-Nowell & J Kleinberg, "The link prediction problem for social networks." *Journal of the American Society for Information Science and Technology*, Vol. 58, No. 7. (May 2007), pp. 1019-1031.
    2. L Lu and T Zhou, "Link prediction in complex networks: a survey", *Physica A* 390(6):11501170 (2011)
    3. Backstrom, L. and Kleinberg, J. (2013). Romantic partnerships and the dispersion of social ties: A network analysis of relationship status on facebook. In *Proceedings of the 17th ACM conference on Computer supported cooperative work & social computing - CSCW '14*, CSCW '14, pages 831-841, New York, NY, USA. ACM Press.
  - **Project proposals due**
  - **Quiz 5**

- **Week 8: October 13**
  - **Topic: Social Spam and Malicious Behavior**
  - **Readings:**
    1. Grier, C., Thomas, K., Paxson, V., 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.
    2. B Markines, C Cattuto, F Menczer, 2009. "Social spam detection" In *Proceedings of the 5th International Workshop on Adversarial Information Retrieval on the Web*, pp. 41-48.
    3. Ghosh, R.; Surachawala, T.; and Lerman, K. 2011. "Entropy-based Classification of 'Retweeting' Activity on Twitter." In *Proceedings of KDD workshop on Social Network Analysis (SNA-KDD).*

- **Week 8: October 15**
  - **Topic: Social Spam and Malicious Behavior**

- **Readings:**
  1. Budak, C., Agrawal, D., and El Abbadi, A. (2011). Limiting the spread of misinformation in social networks. In Proceedings of the 20th International Conference on World Wide Web, WWW '11, pages 665-674, New York, NY, USA. ACM.
  2. Ratkiewicz, J., Conover, M., Meiss, M., Gonçalves, B., Patil, S., Flammini, A., and Menczer, F. (2011). Detecting and tracking the spread of astroturf memes in microblog streams. In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*.
  3. Ferrara, E., Varol, O., Davis, C., Menczer, F., and Flammini, A. (2014). The rise of social bots. http://arxiv.org/abs/1407.5225
  - **Quiz 6**

- **Week 9: October 20**
  - **Topic: Geospatial social data mining**
  - **Readings:**
  1. T Rattenbury, M Naaman. 2009 "Methods for extracting place semantics from Flickr tags" *ACM Trans. Web*, Vol. 3, No. 1, pp. 1-30.
  2. 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*.
  3. D J. Crandall, L Backstrom, D Huttenlocher, J Kleinberg, 2009 "Mapping the world's photos" In *Proceedings of the 18th international conference on World Wide Web*, pp. 761-770.

- **Week 9: October 22**
  - **Topic: Geospatial social data mining**
  - **Readings:**
    - Bo Han et al, (2014) "Text-based User Twitter Geolocation Prediction." *J. Artificial Intelligence Research* 49 pp 451—500. http://www.jair.org/media/4200/live-4200-7781-jair.pdf
    - 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)*
    - Backstrom, L., Sun, E., Marlow, C. 2010 "Find me if you can: improving geographical prediction with social and spatial proximity." In *Proceedings of the 19th international conference on World Wide Web*.
    - [optional] Scellato, S., Noulas, A., Lambiotte, R., Mascolo, C. 2011 "Socio-spatial Properties of Online Location-based Social Networks" In *Proceedings of the 5th International AAAI Conference on Weblogs and Social Media (ICWSM)*
    - **Quiz 7**

- **Week 10: October 27**

- **Topic: Privacy in a Networked World**
- **Readings**

1. Kosinski, M., Stillwell, D., and Graepel, T. (2013). Private traits and attributes are predictable from digital records of human behavior. *Proceedings of the National Academy of Sciences*, 110(15):5802-5805
2. Golbeck J, Robles C, Turner K (2011) Predicting personality with social media. *Conference on Human Factors in Computing Systems*, pp 253–262
3. Gosling, S. D., Augustine, A. A., Vazire, S., Holtzman, N., and Gaddis, S. (2011). Manifestations of personality in online social networks: self-reported facebook-related behaviors and observable profile information. *Cyberpsychology, behavior and social networking*, 14(9):483-488.
4. [optional] Jennifer Golbeck, "The Curly Fries Conundrum" http://www.ted.com/talks/jennifer_golbeck_the_curly_fry_conundrum_why_social_media_likes_say_more_than_you_might_think

- **Week 10: October 29**
  - **Topic: Health**
  - **Readings:**
    1. M. De Choudhury, S. Counts, E. Horvitz, A. Hoff. Characterizing and Predicting Postpartum Depression from Facebook Data., *ICWSM 2014*
    2. R.W. White, R. Harpaz, N.H. Shah, W. DuMouchel, and E. Horvitz. Toward Enhanced Pharmacovigilance using Patient-Generated Data on the Internet, *Nature CPT*, April 2014.
    - **Quiz 8**

- **Week 11: November 3**
  - **Topic: Politics and Social Media**
  - **Readings:**
    1. MD Conover, E Ferrara, F Menczer, and A Flammini. The Digital Evolution of Occupy Wall Street. PLoS ONE 8(5):e64679, 2013
    2. Conover, M. D., Gonçalves, B., Flammini, A., and Menczer, F. (2012). Partisan asymmetries in online political activity. *EPJ Data Science*, 1(1):6+.
    3. Lietz, H., Wagner, C., Bleier, A., and Strohmaier, M. (2014). When politicians talk: Assessing online conversational practices of political parties on twitter. In *Proceedings of AAAI International Conference on Weblogs and Social Media*.

- **Week 11: November 5**
  - **Topic: Predicting the future with social media**
  - **Readings:**
    1. A. Tumasjan, T. O. Sprenger, P. G. Sandner, I. M. Welpe, Predicting Elections with Twitter: What 140 Characters Reveal about Political Sentiment, In *ICWSM*, 2010.

2. 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 457, Feb 19, 2009
3. Lazer, D., Kennedy, R., King, G., and Vespignani, A. (2014). The parable of google flu: Traps in big data analysis. *Science*, 343(6176):1203-1205.
4. [optional] Goel, S., Hofman, J., Lahaie, S., Pennock, D., Watts, D. (2010) "Predicting consumer behavior with Web search." *Proceedings of the National Academies of Science* 107(41)
5. [optional] D. Gayo-Avello, "I wanted to predict elections on Twitter, but all I got was this lousy paper." http://arxiv.org/abs/1204.6441

- **Project mid-term report due**
- **Quiz 9**

- **Week 12: November 10**
  - **Topic: Emotional contagion**
  - **Readings:**
    1. Coviello, L., Sohn, Y., Kramer, A. D. I., Marlow, C., Franceschetti, M., Christakis, N. A., and Fowler, J. H. (2014). Detecting emotional contagion in massive social networks. *PLoS ONE*, 9(3):e90315+
    2. Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D. I., Marlow, C., Settle, J. E., and Fowler, J. H. (2012). A 61-million-person experiment in social influence and political mobilization. *Nature*, 489(7415):295-298.
    3. Kramer, A. D. I., Guillory, J. E., and Hancock, J. T. (2014). Experimental evidence of massive-scale emotional contagion through social networks. Proceedings of the National Academy of Sciences, 111(24):8788-8790.
    4. [Optional] http://www.theguardian.com/commentisfree/2014/jul/07/facebook-study-science-experiment-research
    5. [Optional] http://www.nytimes.com/2014/07/01/opinion/jaron-lanier-on-lack-of-transparency-in-facebook-study.html?_r=1

- **Week 12: November 12**
  - **Topic: Friendship paradox and detection of contagions**
  - **Readings:**
    1. Christakis, N. A. and Fowler, J. H. (2010). Social network sensors for early detection of contagious outbreaks. *PLoS ONE*, 5(9):e12948+.
    2. Garcia-Herranz, M., Egido, E. M., Cebrian, M., Christakis, N. A., and Fowler, J. H. (2012). Using friends as sensors to detect Global-Scale contagious outbreaks.
    3. Hodas, N. O., Kooti, F., and Lerman, K. (2013). Friendship paradox redux: Your friends are more interesting than you. In *Proceedings of 7th International Conference on Weblogs and Social Media*.
  - **Quiz 10**

- **Week 13: November 17**
  - **Topic: Crowdsourcing with Mechanical Turk**

- **Readings:**

1. R Snow, B O'Connor, D Jurafsky and A Ng. (2008) Cheap and Fast - But is it Good? Evaluating non-expert annotations for natural language tasks. Proceedings of the conference on Empiricial Methods in Natural Language Processing (EMNLP-08), Honolulu, HI
2. K Fort, G Adda and K. Bretonnel Cohen. Amazon Mechanical Turk: Gold Mine or Coal Mine? In Journal of Computational Linguistics 27(2):413-420, 2011

- **Week 13: November 19**
  - **Topic: Social tagging and folksonomies**
  - **Readings**

1. Golder, S. and Huberman, B. 2005. The Structure of Collaborative Tagging Systems. *Journal of Information Science,* Vol. 32, No. 2.
2. Chi, E. and Mytkowicz, T. 2008. Understanding the efficiency of social tagging systems using information theory, in *HyperText'08.*
3. 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. 1, pp.
4. [optional] Schmitz, P. 2006 Inducing Ontologies from Flickr Tags, in *Proc. of WWW Collaborative Web Tagging workshop*.

- **Quiz 11**

- **Week 14: November 24 (Or class presentations, depending on enrollement)**
  - **Topic: Social Multimedia Analysis: Videos**
  - **Readings:**

1. Cha, M., Kwak, H., Rodriguez, P., Ahn, Y. and Moon, S. (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.
2. Morsillo, N., Mann, G., and Pal, C. (2010) "YouTube Scale, Large Vocabulary Video Annotation." In Schonfeld, D., Shan, C., Tao, D., and Wang, L. (eds) *Video Search and Mining: Studies in Computational Intelligence* 287, Springer, pp. 357-386.
3. Biel, J. and Gatica-Perez, D. (2011) "VlogSense: Conversational Behavior and Social Attention in YouTube." *ACM Transactions on Multimedia Computing, Communications, and Applications, Special Issue on Social Media* (in press, 2011).

- **Week 14: November 26**
  - **Thanksgiving Holiday**

- **Week 15: December 1**

4. **Class presentations**

- **Week 15: December 3**
    1. **Class presentations**