

Social Media Analysis Course Syllabus

Prerequisites: None.

Time: Spring 2014, Mondays and Wednesdays at 4-5:30pm

Instructors: Professors Kristina Lerman (lerman@isi.edu) and Zornitsa Kozareva (kozareva@isi.edu)

Office: ISI 932, ISI 935

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.

This course will examine topics in social media analysis, including influence and centrality in social networks, information diffusion, sentiment analysis, and show how AI, social network analysis, and linguistic and statistical methods can be used to study these topics. While there are no prerequisites, we 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.

Topics and Readings

- **January 13**
 - **Topic: Course Introduction: (Professors Lerman & Kozareva)**

- **January 15**
 - **Topic: Sentiment Analysis on Twitter (Professor Kozareva)**
 - **Readings:**
 - 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.
 - A Agarwal, B Xie, I Vovsha, O Rambow and R Passonneau. Sentiment Analysis of Twitter Data. In *Proceedings of the Workshop on Languages in Social Media*. 2011
 - I Hendrickx, Z. Kozareva, P Nakov, D Ó Séaghdha, S Szpakowicz and T Veale, SemEval-2013 Task 4: Free Paraphrases of Noun Compounds. in *Proceedings of the SemEval-2013 Workshop (SemEval 2013)*, Atlanta, GA

- **January 20**
 - **MLK Holiday**

- **January 22**
 - **Topic: Sentiment Analysis on YouTube and Movie Reviews (Professor Kozareva)**
 - **Readings:**
 1. B O'Connor, R Balasubramanyan, B R. Routledge, and N A. Smith. From Tweets to Polls: Linking Text Sentiment to Public Opinion Time Series. In *Proceedings of the International AAAI Conference on Weblogs and Social Media (ICWSM 2010)*, pages 122–129, Washington, DC, May 2010
 2. 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)*
 3. LP Morency, R Mihalcea and P Doshi. Towards Multimodal Sentiment Analysis: Harvesting Opinions from the Web. *Proceedings of 13th International Conference on Multimodal Interaction ICMI-2011*, Alicante, Spain
 - **Homework #1 assigned – due February 7**

- **January 27**
 - **Topic: Online social networks (Professor Lerman)**
 - **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)”
<http://www.ams.org/samplings/feature-column/fc-2012-08>

3. N O. Hodas, F Kooti, K Lerman “Friendship paradox redux: Your friends are more interesting than you.” In ICWSM 2013.
4. [optional] L Backstrom, P Boldi, M Rosa, J Ugander, S Vigna. “Four Degrees of Separation,” 2012

- **January 29**

- **Topic: Structure of the Web (Professor Kozareva)**

- **Readings:**

1. 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”
<http://www.cs.cornell.edu/home/kleinber/networks-book/networks-book-ch13.pdf>
2. A Broder, R Kumar, F Maghoul, P Raghavan, S Rajagopalan, R Stata, A Tomkins and J Wiener. Graph structure in the web. *Proceedings of the 9th international World Wide Web conference on Computer networks: the international journal of computer and telecommunications networking*. 2000

2. February 3

- **Topic: Influence and centrality in social networks (Professor Lerman)**

- **Readings:**

1. Freeman, L. 1979 “Centrality in Social Networks: Conceptual Clarification”, *Social Networks* 1, No. 3.
2. Bonacich, P. 1987 “Power and Centrality, a family of measures” *The American Journal of Sociology*, Vol. 92, No. 5.
3. M. Franceschetti 2011 “PageRank: standing on the shoulders of giants” *Commun. ACM*, Vol. 54, pp. 92-101.
4. (optional) Lerman, K and Ghosh, R. 2011 Parametrized Centrality for Network Analysis. *Physical Review E*.

3. February 5

- **Topic: Influence and centrality in social networks (Professor Lerman)**

- **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.

4. February 10

- **Topic: Information diffusion (Professor Lerman)**

- **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.

5. February 12

- **Topic: Information diffusion (Professor Lerman)**
- **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.
- **Homework #2 – due February 28**

6. February 17

- **Topic: Wikipedia knowledge extraction and similarity (Professor Kozareva)**
- **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.

7. February 19

- **Topic: Search query logs (Professor Kozareva)**
- **Readings:**
 1. 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.
 2. M Strohmaier, P Prettenhofer and M Kröll. Acquiring knowledge about human goals from search query logs. *Information Processing and Management*. 2011.

8. February 24

- **President's Day**

9. February 26

- **Topic: Web-based text mining (Professor Kozareva)**
- **Readings:**
 1. Z. Kozareva and E. Hovy, A Semi-Supervised Method to Learn and Construct Taxonomies using the Web. in Proceedings of Conference on Empirical Methods in Natural Language Processing (EMNLP 2010), Boston
 2. Z. Kozareva and E Hovy, Insights from Network Structure for Text Mining. in Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies (ACL-HLT 2011), Portland.

10. March 3

- **Topic: Social ties and information diffusion (Professor Lerman)**
- **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.
 4. (optional) S. Aral and M van Alstyne, "[The Diversity-bandwidth tradeoff](#)", *American Journal of Sociology*, Vol. 117, No. 1. (2011),

11. March 5

- **Topic: Social ties and link prediction (Professor Lerman)**
- **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. Lerman, K et al. "Using Proximity to Predict Activity in Social Networks" <http://arxiv.org/abs/1112.2755>.
- **Project proposals due**

12. March 10

- **Topic: Social Multimedia Analysis: Photos (Professor Kozareva)**
- **Readings:**
 1. M Mitchell, J Dodge, A Goyal, K Yamaguchi, K Stratos, X Han, A Mensch, AC Berg, T L. Berg, H Daumé III. Midge: Generating Image Descriptions From Computer Vision Detections. In Proceedings of the Conference of the

European Chapter of the Association for Computational Linguistics (EACL), 2012

2. X Jin, A Gallagher, L Cao, J Luo, and J Han. The wisdom of social multimedia: Using Flickr for prediction and forecast. In *ACM Multimedia*, October 25-29, 2010, Firenze, Italy
3. S Agarwal, N Snavely, I Simon, S M. Seitz and R Szeliski. Building Rome in a Day. *Communications of the ACM* 54(10):105-112.

13.March 12

- **Topic: Social Multimedia Analysis: Videos (Professor Kozareva)**
- **Readings:**
 1. M Cha, H Kwak, P Rodriguez, YY Ahn and S Moon. I Tube, You Tube, Everybody Tubes: Analyzing the World's Largest User Generated Content Video System. In *Proceedings of Usenix/ACM SIGCOMM Internet Measurement Conference (IMC)*, San Diego, CA, 2007
 2. JI Biel and D Gatica-Perez. VlogSense: Conversational Behavior and Social Attention in YouTube. In *Proceedings of ACM Transactions on Multimedia Computing, Communications, and Applications, Special Issue on Social Media*, 2011

14.March 17, 19

- **Spring break!**

15.March 24

- **Topic: Mining geo-spatial data (Professor Lerman)**
- **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.

16.March 26

- **Topic: Geospatial social networks (Professor Lerman)**
- **Readings:**
 1. 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)*
 2. 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)*

3. 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*.

17. March 31

- **Topic: Crowdsourcing with AMT (Professor Kozareva)**
- **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 Empirical 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

18. April 2

- **Topic: Collaborative Filtering on LinkedIn and Netflix (Professor Kozareva)**
- **Readings:**
 1. Svetlin Alex Bostandjiev, John O'Donovan, Tobias Hollerer. LinkedVis: Exploring Social and Semantic Career Recommendations. IUI '13 Proceedings of the 2013 international conference on Intelligent user interfaces, 2013 <http://www.igvita.com/2006/10/29/dissecting-the-netflix-dataset/>
 2. Yehuda Koren, Robert Bell and Chris Volinsky. Matrix Factorization Techniques For Recommender Systems. In *Journal of Computer*, 2009.
- **Midterm progress reports due**

19. April 7

- **Topic: Social tagging and folksonomies (Professor Lerman)**
- **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. Plangprasopchok, A. and Lerman, K. Exploiting Social Annotation for Automatic Resource Discovery, in *Proc. of AAAI Workshop on Information Integration on the Web*, 2007.
 4. (optional) 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.

20. April 9

- **Topic: Social tagging an folksonomies (Professor Lerman)**
- **Readings:**

1. 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. 5-15.
2. Schmitz, P. 2006 Inducing Ontologies from Flickr Tags, in *Proc. of WWW Collaborative Web Tagging workshop*.
3. Plangprasopchok, A. Lerman, K., and Getoor, L. 2010 [Growing a tree in a forest: constructing folksonomies by integrating structured metadata](#). In *KDD'10*.

21. April 14

- **Topic: Bullying Detection on Twitter, YouTube and Formspring (Professor Kozareva)**
- **Readings:**
 1. JM Xu, KS Jun, X Zhu and A Bellmore. Learning from bullying traces in social media. *Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (ACL 2012)*. 2012. <http://dl.acm.org/citation.cfm?id=2382029.2382139>
 2. JM Xu, B Burchfiel, X Zhu and A Bellmore. An Examination of Regret in Bullying Tweets. *Proceedings of the 2013 Conference of North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL HLT 2013)* <http://aclweb.org/anthology/N/N13/N13-1082.pdf>
 3. K Dinakar, B Jones, C Havasi and H Lieberman. Common Sense Reasoning for Detection, Prevention, and Mitigation of Cyberbullying. *Journal of ACM Transactions on Interactive Intelligent Systems*, Vol. 2, No. 3, Article 18, 2012 <http://web.media.mit.edu/~lieber/Publications/Bullying-TiiS.pdf>

22. April 16

- **Topic: Reasoning with social media (Professor Kozareva)**
- **Readings:**
 1. A Gordon, C Bejan and K Sagae. Commonsense Causal Reasoning Using Millions of Personal Stories. Twenty-Fifth Conference on Artificial Intelligence (AAAI-11), San Francisco, CA.
 2. A Ritter, C Cherry and W Dolan. Data-Driven Response Generation in Social Media. *Proceedings of the Conference on Empirical Methods in Natural Language Processing (EMNLP-2011)* Edinburgh, Scotland

23. April 21

- **Topic: Spam in social media (Professor Lerman)**
- **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)*.

24. April 23

- **Topic: Predicting the future with social media (Professor Lerman)**

- **Readings:**

1. 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
2. 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)
3. D. Gayo-Avello, "I wanted to predict elections on Twitter, but all I got was this lousy paper."

25. April 28

- **Class presentations (Professors Lerman and Kozareva)**

26. April 30

- **Class presentations (Professors Lerman and Kozareva)**