**COMPREHENSIVE REVIEW READING LIST for Fall 2016 and Spring 2017**

**Database/Data Mining**

1. Michael Stonebraker Joseph M. Hellerstein:”What Goes Around Comes Around” Readings in Database Systems, Fourth Edition Morgan Kaufmann 2005.
2. Domingos, Pedro. "A few useful things to know about machine learning." *Communications of the ACM* 55.10 (2012): 78-87.
3. Fabrizio Sebastiani. “Machine learning in automated text categorization. ACM Comput. Surv. 34(1): 1-47 (2002).
4. Ankit Toshniwal, Siddarth Taneja, Amit Shukla, Karthik Ramasamy, Jignesh M. Patel, Sanjeev Kulkarni, Jason Jackson et al. "Storm@ twitter." In Proceedings of the 2014 ACM SIGMOD international conference on Management of data, pp. 147-156. ACM, 2014.
5. Chen, M.-S., Han, J. and Yu, P. S. “Data Mining: An Overview from a Database Perspective,” IEEE Transactions on Knowledge and Data Engineering, 8(6):866-883, 1996.
6. Gengm L. and Hamilton, H. J. "Interestingness Measures for Data Mining: A Survey", ACM Computing Surveys, 38(2), 2006.
7. Ulrike Luxburg. 2007. A tutorial on spectral clustering. Statistics and Computing 17, 4 (December 2007), 395-416.
8. Zheleva, Elena, and Lise Getoor. "Privacy in social networks: A survey." *Social Network Data Analytics*. Springer US, 2011. 277-306. http://www.cise.ufl.edu/~mythai/courses/2012/cis6930/Notes/privacy\_survey\_chapter.pdf
9. Cynthia Dwork. 2006. Differential privacy. In Proceedings of the 33rd international conference on Automata, Languages and Programming - Volume Part II (ICALP'06), Michele Bugliesi, Bart Preneel, Vladimiro Sassone, and Ingo Wegener (Eds.), Vol. Part II. Springer-Verlag, Berlin, Heidelberg, 1-12. DOI=10.1007/11787006\_1 <http://dx.doi.org/10.1007/11787006_1>
10. Nikita Spirin and Jiawei Han. 2012. Survey on web spam detection: principles and algorithms. *SIGKDD Explorations Newsletter* 13, 2 (May 2012), 50-64. DOI=10.1145/2207243.2207252 http://doi.acm.org/10.1145/2207243.2207252
11. Yi Mao, Wenlin Chen, Yixin Chen, Chenyang Lu, Marin Kollef, and Thomas C. Bailey. "An Integrated Data Mining Approach to Real-time Clinical Monitoring and Deterioration Warning". In KDD'12.
12. L. Sweeney, k-anonymity: a model for protecting privacy. International Journal on Uncertainty, Fuzziness and Knowledge-based Systems, 10 (5), 2002; 557-570
13. Jeffrey Dean and Sanjay Ghemawat. "MapReduce: Simplified Data Processing on Large Clusters". Communications of the ACM, vol. 51, no. 1 (2008), pp. 107-113
14. C. L. Philip Chen, Chun-Yang Zhang: Data-intensive applications, challenges, techniques and technologies: A survey on Big Data. Inf. Sci. 275: 314-347 (2014), http://www.sciencedirect.com/science/article/pii/S0020025514000346
15. Xindong Wu, Xingquan Zhu, Gong-Qing Wu, and Wei Ding. "Data mining with big data." Knowledge and Data Engineering, IEEE Transactions on 26, no. 1 (2014): 97-107.
16. Matei Zaharia, Mosharaf Chowdhury, Tathagata Das, Ankur Dave, Justin Ma, Murphy McCauley, Michael J. Franklin, Scott Shenker, and Ion Stoica. "Resilient distributed datasets: A fault-tolerant abstraction for in-memory cluster computing." In Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation, pp. 2-2. USENIX Association, 2012.
17. David M. Blei, Andrew Y. Ng, and Michael I. Jordan. "Latent Dirichlet Allocation". Journal of Machine Learning Research (3): 993-1022 (2003).
18. A. Sperotto, G. Schaffrath, R. Sadre, C. Morariu, A. Pras, B. Stiller. "An Overview of IP Flow-Based Intrusion Detection" IEEE Communications Surveys & Tutorials, 12(3), 2010.
19. A. Halevy, A. Rajaraman, J. Ordille. "Data Integration: The Teenage Years". Proceedings of the 32nd International Conference on Very Large Data Bases (VLDB), September 2006, Seoul, Korea, pp. 9-16.
20. J. Bleiholder, and F. Naumann. "Data fusion". /ACM Computing Surveys/ 41, 1 (Dec. 2008), 1-41. DOI= <http://doi.acm.org/10.1145/1456650.1456651>
21. B. Haslhofer, W. Klass. "A Survey of Techniques for Achieving Metadata Interoperability," ACM Computing Surveys, 42(2), 2010.
22. P. Bisht, P. Madhusudan, V. Venkatakrishnan. "CANDID: Dynamic Candidate Evaluations for Automatic Prevention of SQL Injection Attacks," ACM Transactions on Information and System Security 13(2), Feb. 2010.
23. Martin Ester, Hans-Peter Kriegel, Jörg Sander (University of Munich), Algorithms and Applications for Spatial Data Mining, Published in Geographic Data Mining and Knowledge Discovery, Research Monographs in GIS, Taylor and Francis, 2001. <http://www.dbs.informatik.uni-muenchen.de/Publikationen/Papers/Chapter7.revised.pdf>
24. Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. ACM Comput. Surv. 41, 3, Article 15 (July 2009), 58 pages. DOI=<http://doi.acm.org/10.1145/1541880.1541882>
25. Daniel B. Neill, Andrew W. Moore, Maheshkumar Sabhnani, and Kenny Daniel. 2005. Detection of emerging space-time clusters. In Proceedings of the eleventh ACM SIGKDD international conference on Knowledge discovery in data mining (KDD '05). ACM, New York, NY, USA, 218-227. DOI=10.1145/1081870.1081897 http://doi.acm.org/10.1145/1081870.1081897
26. Saha Budhaditya, Duc-Son Pham, Mihai Lazarescu, and Svetha Venkatesh. 2009. Effective Anomaly Detection in Sensor Networks Data Streams. In Proceedings of the 2009 Ninth IEEE International Conference on Data Mining (ICDM '09). IEEE Computer Society, Washington, DC, USA, 722-727. DOI=10.1109/ICDM.2009.110 <http://dx.doi.org/10.1109/ICDM.2009.110>
27. Jure Leskovec, Jon Kleinberg, and Christos Faloutsos. "Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations". In KDD'05, 177-187.
28. A. Langville and C. Meyer, “A Survey of Eigenvector Methods for Web Information Retrieval,” SIAM Review, vol. 47, no. 1, pp. 135–161, 2005.
29. Plantié, Michel, and Michel Crampes. "Survey on social community detection." Social Media Retrieval. Springer London, 2013. 65-85.
30. Jiliang Tang, Yi Chang, Huan Liu. "Mining Social Media with Social Theories: A Survey". SIGKDD Explorations, December 2013, Volume 15, Issue 2.