

9. REFERENCES

- [1] English stop words.
<http://www.textfixer.com/resources/common-english-words.txt>.
- [2] T. Broderick, N. Boyd, A. Wibisono, A. C. Wilson, and M. Jordan. Streaming variational bayes. In C. Burges, L. Bottou, M. Welling, Z. Ghahramani, and K. Weinberger, editors, *Advances in Neural Information Processing Systems 26*, pages 1727–1735. 2013.
- [3] P. G. Constantine and D. F. Gleich. Tall and skinny qr factorizations in mapreduce architectures. In *Proceedings of the Second International Workshop on MapReduce and Its Applications*, MapReduce '11, pages 43–50, New York, NY, USA, 2011. ACM.
- [4] J. Fox. Robust regression: Appendix to an r and s-plus companion to applied regression, 2002.
- [5] L. Getoor and B. Taskar. *Introduction to Statistical Relational Learning*. The MIT Press, 2007.
- [6] B. Hall, A. Jaffe, and M. Trajtenberg. The nber patent citations data file: Lessons, insights and methodological tools, 2001.
- [7] K. S. Jones. A statistical interpretation of term specificity and its application in retrieval. *Journal of Documentation*, 28:11–21, 1972.
- [8] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1, 2007.
- [9] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over time: Densification laws, shrinking diameters and possible explanations. In *Proceedings of the Eleventh ACM SIGKDD International Conference on Knowledge Discovery in Data Mining*, KDD '05, pages 177–187, New York, NY, USA, 2005. ACM.
- [10] S. A. Macskassy and F. Provost. A simple relational classifier. In *MRDM-KDD*, 2003.
- [11] G. S. Manku, S. Rajagopalan, and B. G. Lindsay. Approximate medians and other quantiles in one pass and with limited memory. In *Proceedings of the 1998 ACM SIGMOD International Conference on Management of Data*, SIGMOD '98, pages 426–435, New York, NY, USA, 1998. ACM.
- [12] L. McDowell and D. W. Aha. Semi-supervised collective classification via hybrid label regularization. In *ICML*. icml.cc / Omnipress, 2012.
- [13] K. Nigam, A. K. McCallum, S. Thrun, and T. Mitchell. Text classification from labeled and unlabeled documents using em. *Mach. Learn.*, 39(2-3):103–134, May 2000.
- [14] J. J. Pfeiffer III, J. Neville, and P. Bennett. Composite likelihood data augmentation for within-network statistical relational learning. In *Proceedings of the 14th IEEE International Conference on Data Mining (ICDM 2014)*, 2014.
- [15] M. Richardson and P. Domingos. Markov logic networks. *Mach. Learn.*, 62(1-2):107–136, Feb. 2006.
- [16] B. Taskar, P. Abbeel, and D. Koller. Discriminative probabilistic models for relational data. In *Proceedings of the Eighteenth Conference on Uncertainty in Artificial Intelligence*, UAI'02, pages 485–492, San Francisco, CA, USA, 2002. Morgan Kaufmann Publishers Inc.
- [17] Y. Tsuruoka and J. Tsujii. Training a naive bayes classifier via the em algorithm with a class distribution constraint. In *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, CONLL '03, pages 127–134, Stroudsburg, PA, USA, 2003. Association for Computational Linguistics.
- [18] R. Xiang and J. Neville. Pseudolikelihood em for within-network relational learning. In *ICDM*, 2008.
- [19] R. Xiang and J. Neville. Understanding propagation error and its effect on collective classification. In D. J. Cook, J. Pei, W. W. 0010, O. R. Zařane, and X. Wu, editors, *ICDM*, pages 834–843. IEEE, 2011.
- [20] X. Zhu, Z. Ghahramani, and J. Lafferty. Semi-supervised learning using gaussian fields and harmonic functions. In *IN ICML*, pages 912–919, 2003.