CSE 5095 Research Topics in Big Data Analytics

A List of References

1. **General:** J.S. Vitter, *Algorithms and Data Structures for External Memory*, Foundations and Trends in Theoretical Computer Science 2:4, 2008, ISSN: 1551-305X.

2. **Single Disk Sorting:** D.E. Knuth, *The Art of Programming, Volume 3: Sorting and Searching*, Addison-Wesley, 1973, Section 5.4.

3. **Single and Multiple Disks Selection:** S. Rajasekaran, Selection Algorithms for Parallel Disk Systems, *Journal of Parallel and Distributed Computing* 61(4): 536-544, 2001.

4. **A lower bound with proof for external sorting:** `http://www.imada.sdu.dk/~rolf/Edu/DM808/F08/Handouts/Alower.pdf`

5. **B-trees:** T. Cormen, C. Leiserson, and R.L. Rivest, *Introduction to Algorithms*, MIT Press (any edition).

6. **MST:** L. Arge, G. Brodal, and L. Toma, On external memory MST, SSSP and multi-way planar graph separation, *Proc. 7th Scandinavian Workshop on Algorithm Theory*, volume 1851 of LNCS, pages 433447. Springer, 2000.

7. **MST:** Roman Dementiev, Peter Sanders, Dominik Schultes, Jop F. Sibeyn: Engineering an External Memory Minimum Spanning Tree Algorithm, IFIP TCS 2004: 195-208.

8. **PDM Sorting:** S. Rajasekaran, A Framework for Simple Sorting Algorithms on Parallel Disk Systems, *Theory of Computing Systems* 34, 2001, 101-114.

9. **PDM Sorting:** S. Rajaselaran and S. Sen, Optimal and Practical Algorithms for Sorting on the PDM, *IEEE Transactions on Computers* 57(4), 2008, 547-561.

10. **Suffix trees and suffix arrays:** D. Gusfield, *Algorithms on Strings, Trees and Sequences: Computer Science and Computational Biology*, Cambridge University Press, 1997.

11. **Rules Mining:** P.-N. Tang, M. Steinbach, and V. Kumar, *Introduction to Data Mining*, Addison-Wesley, 2005, Chapter 6, Association Analysis: Basic Concepts and Algorithms.

12. **Randomized Rules Mining:** H. Toivonen, Sampling Large Databases for Association Rules, *Proc. 22nd VLDB Conference*, Mumbai, India, 1996.

13. **Hierarchical Clustering:** S. Rajasekaran, Efficient Parallel Hierarchical Clustering Algorithms, *IEEE Transactions on Parallel and Distributed Systems*, 16(6), 2005, 497-502.

14. **Text Mining:** J. Vyas, R.J. Nowling, T. Meusburger, D. Sargeant, K. Kadaveru, M.R. Gryk, V. Kundeti, S. Rajasekaran and M.R. Schiller, MimoSA: a system for minimotif annotation, *BMC Bioinformatics*, 11:328, 2010.

15. **A proof of Johnson-Lindenstrauss Theorem:** S. Dasgupta and A. Gupta, An Elementary Proof of a Theorem of Johnson and Lindenstrauss, *Random Structures and Algorithms* 22(1):60-65, 2003.

16. **Singular Value Decomposition:** S. Rajasekaran and M. Song, A relaxation scheme for increasing the parallelism in Jacobi-SVD, *Journal of Parallel and Distributed Computing* 68(6): 769-777, 2008.

17. **Learning:** L.G. Valiant, A Theory of the Learnable, *Communications of the ACM*, 27(11), 1984, 1134-1142.

18. **The light bulb problem:** R. Paturi, S. Rajasekaran, and J.H. Reif, The Light Bulb Problem, *Information and Computation* 117, 1995, 187-192.

19. **Light Bulb Algorithm Application:** P. Achlioptas, B. Schölkopf, and K.M. Borgwardt, Two-Locus Association Mapping in Subquadratic Time, Proc. KDD, 2011, 726-734.