```
f = open("refs.bib",'r')
s = f.readline()
while s!='':
    if s.find('@')>-1:print s.split('{')[1][:-2],
    s = f.readline()
```

[1–143]

References

[1] S. Gollapudi and R. Panigrahy, "The power of two min-hashes for similarity search among hierarchical data objects," in *Proceedings of the twenty-seventh ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, PODS '08, (New York, NY, USA), pp. 211–220, ACM, 2008.

[2] A. Vardy and Y. Be'ery, "Maximum likelihood decoding of the leech lattice," *Information Theory, IEEE Transactions on*, vol. 39, pp. 1435–1444, jul 1993.

[3] D. G. Lowe, "Object recognition from local scale-invariant features," in *Computer Vision, 1999. The Proceedings of the Seventh IEEE International Conference on*, vol. 2, pp. 1150–1157, 1999.

[4] N. V. I. D. I. A. Corporation, "Nvidia cuda c programming guide," 2010. Version 3.2.

[5] F.-W. Sun and H. C. A. van Tilborg, "The leech lattice, the octacode, and decoding algorithms," *Information Theory, IEEE Transactions on*, vol. 41, pp. 1097–1106, jul 1995.

[6] A. R. Calderbank, "The art of signaling: fifty years of coding theory," *Information Theory, IEEE Transactions on*, vol. 44, pp. 2561–2595, oct 1998.

[7] O. Amrani, Y. Be'ery, A. Vardy, F.-W. Sun, and H. C. A. van Tilborg, "The leech lattice and the golay code: bounded-distance decoding and multilevel constructions," *Information Theory, IEEE Transactions on*, vol. 40, pp. 1030–1043, jul 1994.

[8] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *Int. J. Comput. Vision*, vol. 60, pp. 91–110, November 2004.

[9] A. Andoni, *Nearest Neighbor Search: the Old, the New, and the Impossible*. PhD thesis, MASSACHUSETTS INSTITUTE OF TECHNOLOGY, September 2009.

[10] R. Panigrahy, "Entropy based nearest neighbor search in high dimensions," in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, SODA '06, (New York, NY, USA), pp. 1186–1195, ACM, 2006.

[11] V. Tarokh and I. F. Blake, "Trellis complexity versus the coding gain of lattices. i," *Information Theory, IEEE Transactions on*, vol. 42, pp. 1796–1807, nov 1996.

[12] V. Tarokh and I. F. Blake, "Trellis complexity versus the coding gain of lattices. i," *Information Theory, IEEE Transactions on*, vol. 42, pp. 1796–1807, nov 1996.

[13] F.-w. Sun, *Decoding Techniques and Modulations Scheme for Band-Limited Communication Channels*. PhD thesis, Technical University of Eidenhoven, February 1994.

[14] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," *Commun. ACM*, vol. 51, pp. 117–122, January 2008.

[15] G. Griffin, A. Holub, and P. Perona, "Caltech-256 object category dataset," Tech. Rep. 7694, 2007.

[16] O. Amrani and Y. Beery, "Efficient bounded-distance decoding of the hexacode and associated decoders for the leech lattice and the golay code," *Communications, IEEE Transactions on*, vol. 44, pp. 534–537, may 1996.

[17] B. Georgescu, I. Shimshoni, and P. Meer, "Mean shift based clustering in high dimensions: a texture classification example," in *Computer Vision, 2003. Proceedings. Ninth IEEE International Conference on*, pp. 456–463, oct. 2003.

[18] M. Fashing and C. Tomasi, "Mean shift is a bound optimization," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 27, pp. 471–474, march 2005.

[19] A. Andoni and P. Indyk, "Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions," in *Foundations of Computer Science, 2006. FOCS '06. 47th Annual IEEE Symposium on*, pp. 459–468, oct. 2006.

[20] L. Paulevé, H. Jégou, and L. Amsaleg, "Locality sensitive hashing: A comparison of hash function types and querying mechanisms," *Pattern Recognition Letters*, vol. 31, no. 11, pp. 1348–1358, 2010.

[21] J. Leech, "Notes on sphere packings," *Canadian Journal of Mathematics*, 1967.

[22] . J. Forney, G. D., "A bounded-distance decoding algorithm for the leech lattice, with generalizations," *Information Theory, IEEE Transactions on*, vol. 35, pp. 906–909, jul 1989.

[23] . A. L. . S. C. . G. P. Jégou, H., "Query-adaptive locality sensitive hashing," 2008.

[24] H. Samet, *Foundations of Multidimensional and Metric Data Structures*. Morgan Kaufmann, 2006.

[25] J. Conway and N. Sloane, *Sphere Packings Lattices and Groups Third Edition*. New York: Springer, 1998.

[26] Q. Lv, W. Josephson, Z. Wang, M. Charikar, and K. Li, "Multi-probe lsh: efficient indexing for high-dimensional similarity search," in *Proceedings of the 33rd international conference on Very large data bases*, VLDB '07, pp. 950–961, VLDB Endowment, 2007.

[27] S. Har-Peled, "A replacement for voronoi diagrams of near linear size," in *Foundations of Computer Science, 2001. Proceedings. 42nd IEEE Symposium on*, pp. 94–103, oct. 2001.

[28] D. Forney, "Chapter 14: Introduction to lattice and trellis codes." Lecture: Principles of Digital Communication, 2005.

[29] R. Weber, H.-J. Schek, and S. Blott, "A quantitative analysis and performance study for similarity-search methods in high-dimensional spaces," in *Proceedings of the 24rd International Conference on Very Large Data Bases*, VLDB '98, (San Francisco, CA, USA), pp. 194–205, Morgan Kaufmann Publishers Inc., 1998.

[30] P. Indyk and R. Motwani, "Approximate nearest neighbors: towards removing the curse of dimensionality," in *Proceedings of the thirtieth annual ACM symposium on Theory of computing*, STOC '98, (New York, NY, USA), pp. 604–613, ACM, 1998.

[31] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni, "Locality-sensitive hashing scheme based on p-stable distributions," in *Proceedings of the twentieth annual symposium on Computational geometry*, SCG '04, (New York, NY, USA), pp. 253–262, ACM, 2004.

[32] P. V. Huffman W., *Fundamentals of Error-Correcting Codes*. Cambridge University Press, 1 ed., 2003.

[33] S. B. Wicker, *Error control systems for digital communication and storage*. Upper Saddle River, NJ, USA: Prentice-Hall, Inc., 1995.

[34] P. Woelfel, "Asymmetric balanced allocation with simple hash functions," in *Proceedings of the seventeenth annual ACM-SIAM symposium on Discrete algorithm*, SODA '06, (New York, NY, USA), pp. 424–433, ACM, 2006.

[35] A. Ostlin and R. Pagh, "Uniform hashing in constant time and linear space," in *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, STOC '03, (New York, NY, USA), pp. 622–628, ACM, 2003.

[36] H. Cohn and A. Kumar, "Optimality and uniqueness of the leech lattice among lattices," tech. rep., 2004.

[37] M. Muja and D. G. Lowe, "Fast approximate nearest neighbors with automatic algorithm configuration," in *International Conference on Computer Vision Theory and Application VISSAPP'09)*, pp. 331–340, INSTICC Press, 2009.

[38] A. Thyssen, *ImageMagick v6 Examples – Reference Index*, 2010.

[39] R. T. Curtis, "A new combinatorial approach to m24," *Mathematical Proceedings of the Cambridge Philosophical Society*, vol. 79, no. 01, pp. 25–42, 1976.

[40] G. Shakhnarovich, P. A. Viola, and T. Darrell, "Fast pose estimation with parameter-sensitive hashing," in *ICCV*, 9th IEEE International Conference on Computer Vision (ICCV 2003), 14-17 October 2003, Nice, France, pp. 750–759, IEEE Computer Society, 2003.

[41] . J. Forney, G. D. and G. Ungerboeck, "Modulation and coding for linear gaussian channels," *Information Theory, IEEE Transactions on*, vol. 44, pp. 2384–2415, oct 1998.

[42] R. W. Irving, P. Leather, and D. Gusfield, "An efficient algorithm for the optimal stable marriage," *J. ACM*, vol. 34, pp. 532–543, July 1987.

[43] K. Terasawa and Y. Tanaka, "Spherical lsh for approximate nearest neighbor search on unit hypersphere," in *WADS*, pp. 27–38, 2007.

[44] V. Garcia, E. Debreuve, and M. Barlaud, "Fast k nearest neighbor search using gpu," in *CVPR Workshop on Computer Vision on GPU*, (Anchorage, Alaska, USA), June 2008.

[45] F. Li and R. Klette, "A variant of adaptive mean shift-based clustering," in *Proceedings of the 15th international conference on Advances in neuro-information processing - Volume Part I*, ICONIP'08, (Berlin, Heidelberg), pp. 1002–1009, Springer-Verlag, 2009.

[46] D. R. Karger, "Minimum cuts in near-linear time," *J. ACM*, vol. 47, pp. 46–76, January 2000.

[47] A. Rajaraman and J. D. Ullman, *Mining of Massive Datasets*. Cambridge University Press, 2011.

[48] M. d. Berg, M. v. Kreveld, and M. O. . . . . . e. a.], *Computational geometry : algorithms and applications*. Berlin: Springer, 2000.

[49] K. Beyer, J. Goldstein, R. Ramakrishnan, and U. Shaft, "When is "nearest neighbor" meaningful?," in *In Int. Conf. on Database Theory*, pp. 217–235, 1999.

[50] P. Harish and P. J. Narayanan, "Accelerating large graph algorithms on the gpu using cuda," in *Proceedings of the 14th international conference on High performance computing*, HiPC'07, (Berlin, Heidelberg), pp. 197–208, Springer-Verlag, 2007.

[51] K. Zhou, Q. Hou, R. Wang, and B. Guo, "Real-time kd-tree construction on graphics hardware," *ACM Trans. Graph.*, vol. 27, pp. 126–1, dec 2008.

[52] J. Freeman, "Parallel algorithms for depth-first search," Tech. Rep. MS-CIS-91-71, October 1991.

[53] N. Bell and M. Garland, "Efficient sparse matrix-vector multiplication on cuda," Tech. Rep. NVR-2008-004, dec 2008.

[54] D. M. Mount and S. A. , "Ann: A library for approximate nearest neighbor searching," 1998.

[55] S. Arya and H.-Y. A. Fu, "Expected-case complexity of approximate nearest neighbor searching," in *Proceedings of the eleventh annual ACM-SIAM symposium on Discrete algorithms*, SODA '00, (Philadelphia, PA, USA), pp. 379–388, Society for Industrial and Applied Mathematics, 2000.

[56] S. Kaski, "Dimensionality reduction by random mapping: Fast similarity computation for clustering," 1998.

[57] R. Hecht-Nielsen, "Context vectors: general purpose approximate meaning representations self-organized from raw data," *Computational intelligence: Imitating life*, pp. 43–56, 1994.

[58] D. Comaniciu and P. Meer, "Mean shift: a robust approach toward feature space analysis," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 24, pp. 603–619, may 2002.

[59] E. Bingham and H. Mannila, "Random projection in dimensionality reduction: Applications to image and text data," in *in Knowledge Discovery and Data Mining*, pp. 245–250, ACM Press, 2001.

[60] C. Boutsidis, A. Zouzias, and P. Drineas, "Random projections for k-means clustering," *CoRR*, vol. abs/1011.4632, 2010.

[61] l. Carraher, "A parallel algorithm for query adaptive, locality sensitive hash search," 2012.

[62] W. Carraher, "A gpgpu algorithm for c-approximate r-nearest neighbor search in high dimensions," 2012.

[63] B. M. Marlin, D. C. Kale, R. G. Khemani, and R. C. Wetzel, "Unsupervised pattern discovery in electronic health care data using probabilistic clustering models," in *Proceedings of the 2nd ACM SIGHIT International Health Informatics Symposium*, IHI '12, (New York, NY, USA), pp. 389–398, ACM, 2012.

[64] A. Vattani, "k-means requires exponentially many iterations even in the plane," in *Proceedings of the 25th annual symposium on Computational geometry*, SCG '09, (New York, NY, USA), pp. 324–332, ACM, 2009.

[65] K. R. Sivaramakrishnan, K. Karthik, and C. Bhattacharyya, "Kernels for large margin time-series classification," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, pp. 2746–2751, aug. 2007.

[66] T. W. Liao, "Clustering of time series data–-a survey," *Pattern Recognition*, vol. 38, no. 11, pp. 1857–1874, 2005.

[67] L. Liu, W. Zuo, D. Zhang, and D. Zhang, "Learning with multiple gaussian distance kernels for time series classification," in *Advanced Computer Control (ICACC), 2011 3rd International Conference on*, pp. 624–628, jan. 2011.

[68] P. Sung, Z. Syed, and J. Guttag, "Quantifying morphology changes in time series data with skew," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference on*, pp. 477–480, april 2009.

[69] D. Dilts, J. Khamalah, and A. T. I. T. L. E. Plotkin, "Using cluster analysis for medical resource decision making," *Med Decis Making*, vol. 15, no. 4, pp. 333–47.

[70] M. A. Carreira-Perpinan, "Gaussian mean-shift is an em algorithm," *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, vol. 29, pp. 767–776, may 2007.

[71] A. S. Foundation, *Apache Mahout:: Scalable machine-learning and data-mining library*, April 2011.

[72] T. W. Liao, "Clustering of time series data––a survey," *Pattern Recognition*, vol. 38, pp. 1857–1874, 2005.

[73] M. Saeed, M. Villarroel, A. T. Reisner, G. Clifford, L.-W. Lehman, G. Moody, T. Heldt, T. H. Kyaw, B. Moody, and R. G. Mark, "Multiparameter intelligent monitoring in intensive care ii (mimic-ii): A public-access intensive care unit database," *Critical Care Medicine*, vol. 39, pp. 952–960, May 2011.

[74] T. Hofmann, "Unsupervised learning by probabilistic latent semantic analysis," *Mach. Learn.*, vol. 42, pp. 177–196, jan 2001.

[75] Y. Yang and K. Chen, "Temporal data clustering via weighted clustering ensemble with different representations," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 23, pp. 307–320, feb. 2011.

[76] F. Li and R. Klette, "A variant of adaptive mean shift-based clustering," in *ICONIP (1)*, pp. 1002–1009, 2008.

[77] M. Arias, A. Troncoso, and J. Riquelme, "Advances in intelligent systems and computing: Application to atmospheric pollutants," in *Soft Computing Models in Industrial and Environmental Applications* (V. Snášel, A. Abraham, and E. S. Corchado, eds.), vol. 188, ch. A Kernel for Time Series Classification, pp. 417–426, Springer Berlin Heidelberg, 2013.

[78] K. R. Sivaramakrishnan, K. Karthik, and C. Bhattacharyya, "Kernels for large margin time-series classification," in *Neural Networks, 2007. IJCNN 2007. International Joint Conference on*, pp. 2746–2751, aug. 2007.

[79] G. D. H. E. Gymrek M, McGuire AL and E. Y, "Identifying personal genomes by surname inference," *Science*, no. 339, pp. 321–324, 2013.

[80] X. Ding, L. Zhang, Z. Wan, and M. Gu, "A brief survey on de-anonymization attacks in online social networks," in *Computational Aspects of Social Networks (CASoN), 2010 International Conference on*, pp. 611–615, 2010.

[81] A. Narayanan and V. Shmatikov, "Robust de-anonymization of large sparse datasets," in *Security and Privacy, 2008. SP 2008. IEEE Symposium on*, pp. 111–125, 2008.

[82] M. Azzouzi and I. T. Nabney, "Analysing time series structure with hidden markov models," in *Neural Networks for Signal Processing VIII, 1998. Proceedings of the 1998 IEEE Signal Processing Society Workshop*, pp. 402–408, 1998.

[83] P. K. Dash, M. Nayak, M. R. Senapati, and I. W. C. Lee, "Mining for similarities in time series data using wavelet-based feature vectors and neural networks," *Engineering Applications of Artificial Intelligence*, vol. 20, no. 2, pp. 185–201, 2007. ¡ce:title¿Special Issue on Applications of Artificial Intelligence in Process Systems Engineering¡/ce:title¿.

[84] K. B. PRATT and E. FINK, "Search for patterns in compressed time series," *International Journal of Image and Graphics*, vol. 02, no. 01, pp. 89–106, 2002.

[85] F. Korn, H. V. Jagadish, and C. Faloutsos, "Efficiently supporting ad hoc queries in large datasets of time sequences," *SIGMOD Rec.*, vol. 26, pp. 289–300, jun 1997.

[86] Y.-L. Wu, D. Agrawal, and A. El Abbadi, "A comparison of dft and dwt based similarity search in time-series databases," in *Proceedings of the ninth international conference on Information and knowledge management*, CIKM '00, (New York, NY, USA), pp. 488–495, ACM, 2000.

[87] T.-c. Fu, "A review on time series data mining," *Eng. Appl. Artif. Intell.*, vol. 24, pp. 164–181, feb 2011.

[88] S. S. Vempala, *The Random Projection Method*. DIMACS Series, American Mathematical Society, 2004.

[89] X. Z. Fern and C. E. Brodley, "Random projection for high dimensional data clustering: A cluster ensemble approach,"

[90] T. Urruty, C. Djeraba, and D. Simovici, "Lecture notes in computer science," in *Advances in Data Mining. Theoretical Aspects and Applications* (P. Perner, ed.), vol. 4597, ch. Clustering by Random Projections, pp. 107–119, Springer Berlin Heidelberg, 2007.

[91] S. Dasgupta, "Experiments with random projection," in *Proceedings of the Sixteenth conference on Uncertainty in artificial intelligence*, UAI'00, (San Francisco, CA, USA), pp. 143–151, Morgan Kaufmann Publishers Inc., 2000.

[92] D.-Q. Zhang and S.-C. Chen, "Clustering incomplete data using kernel-based fuzzy c-means algorithm," *Neural Processing Letters*, vol. 18, no. 3, pp. 155–162, 2003.

[93] R. J. Hathaway and J. C. Bezdek, "Fuzzy c-means clustering of incomplete data," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 31, no. 5, pp. 735–744, 2001.

[94] J. C. Bezdek, R. Ehrlich, and W. Full, "Fcm: The fuzzy c-means clustering algorithm," *Computers & Geosciences*, vol. 10, no. 2–3, pp. 191–203, 1984.

[95] R. E. Ladner and M. J. Fischer, "Parallel prefix computation," *J. ACM*, vol. 27, pp. 831–838, oct 1980.

[96] A. Vardy, "Even more efficient bounded-distance decoding of the hexacode, the golay code, and the leech lattice," *Information Theory, IEEE Transactions on*, vol. 41, no. 5, pp. 1495–1499, 1995.

[97] M. Chowdhury, M. Zaharia, and I. Stoica, "Performance and scalability of broadcast in spark,"

[98] K. Shvachko, H. Kuang, S. Radia, and R. Chansler, "The hadoop distributed file system," in *Proceedings of the 2010 IEEE 26th Symposium on Mass Storage Systems and Technologies (MSST)*, MSST '10, (Washington, DC, USA), pp. 1–10, IEEE Computer Society, 2010.

[99] J. Dean and S. Ghemawat, "Mapreduce: Simplified data processing on large clusters," *Commun. ACM*, vol. 51, pp. 107–113, jan 2008.

[100] D. Achlioptas, "Database-friendly random projections," in *Proceedings of the twentieth ACM SIGMOD-SIGACT-SIGART symposium on Principles of database systems*, pp. 274–281, 2001.

[101] M. Al-Razgan and C. Domeniconi, "Weighted clustering ensembles," in *Proceedings of the SIAM international conference on data mining*, pp. 258–269, 2006.

[102] R. Avogadri and G. Valentini, "Fuzzy ensemble clustering based on random projections for dna microarray data analysis," *Artificial Intelligence in Medicine*, vol. 45, no. 2, pp. 173–183, 2009.

[103] I. Florescu, A. Molyboha, and A. Myasnikov, "Scaling and convergence of projection sampling," 2009.

[104] M. Ester, H.-P. Kriegel, J. Sander, and X. Xu, "A density-based algorithm for discovering clusters in large spatial databases with noise.," in *KDD*, vol. 96, pp. 226–231, 1996.

[105] R. T. Ng and J. Han, "Clarans: A method for clustering objects for spatial data mining," *Knowledge and Data Engineering, IEEE Transactions on*, vol. 14, no. 5, pp. 1003–1016, 2002.

[106] C. C. Aggarwal, J. L. Wolf, P. S. Yu, C. Procopiuc, and J. S. Park, "Fast algorithms for projected clustering," in *ACM SIGMOD Record*, vol. 28, pp. 61–72, 1999.

[107] R. Agrawal, J. Gehrke, D. Gunopulos, and P. Raghavan, *Automatic subspace clustering of high dimensional data for data mining applications*, vol. 27. ACM, 1998.

[108] S. Dasgupta, *The hardness of k-means clustering*. Department of Computer Science and Engineering, University of California, San Diego, 2008.

[109] M. Mahajan, P. Nimbhorkar, and K. Varadarajan, "The planar k-means problem is np-hard," in *WALCOM: Algorithms and Computation* (S. Das and R. Uehara, eds.), vol. 5431, ch. Lecture Notes in Computer Science, pp. 274–285, Springer Berlin Heidelberg, 2009.

[110] T. Urruty, C. Djeraba, and D. A. Simovici, "Clustering by random projections," in *Advances in Data Mining. Theoretical Aspects and Applications*, pp. 107–119, Springer, 2007.

[111] "Privacy and progress in whole genome sequencing," report to the president, Presidential Commission for the Study of Bioethical Issues, Washington, October 2012.

[112] R. Kohavi and D. Sommerfield, "Feature subset selection using the wrapper method: Overfitting and dynamic search space topology.," in *KDD*, pp. 192–197, 1995.

[113] P. S. Bradley, O. L. Mangasarian, and W. N. Street, "Clustering via concave minimization," *Advances in neural information processing systems*, pp. 368–374, 1997.

[114] G. Reeves and M. Gastpar, "Sampling bounds for sparse support recovery in the presence of noise," in *Information Theory, 2008. ISIT 2008. IEEE International Symposium on*, pp. 2187–2191, July 2008.

[115] R. Xu and . I. I. Wunsch, D., "Survey of clustering algorithms," *Neural Networks, IEEE Transactions on*, vol. 16, pp. 645–678, May 2005.

[116] L. A. Carraher, P. A. Wilsey, and F. S. Annexstein, "A gpgpu algorithm for c-approximate r-nearest neighbor search in high dimensions," in *Parallel and Distributed Processing Symposium Workshops PhD Forum (IPDPSW), 2013 IEEE 27th International*, pp. 2079–2088, May 2013.

[117] J. Bourgain, "On lipschitz embedding of finite metric spaces in hilbert space," *Israel Journal of Mathematics*, vol. 52, no. 1-2, pp. 46–52, 1985.

[118] J. Magidson and J. Vermunt, "Latent class models for clustering: A comparison with k-means," *Canadian Journal of Marketing Research*, vol. 20, no. 1, pp. 36–43, 2002.

[119] U. V. Luxburg and S. Ben-david, "Towards a statistical theory of clustering," in *In PASCAL workshop on Statistics and Optimization of Clustering*, 2005.

[120] I. Dhillon and D. Modha, "A data-clustering algorithm on distributed memory multiprocessors," in *Large-Scale Parallel Data Mining* (M. Zaki and C.-T. Ho, eds.), vol. 1759, ch. Lecture Notes in Computer Science, pp. 245–260, Springer Berlin Heidelberg, 2000.

[121] G. M. Amdahl, "Validity of the single processor approach to achieving large scale computing capabilities," in *Proceedings of the April 18-20, 1967, Spring Joint Computer Conference*, AFIPS '67 (Spring), (New York, NY, USA), pp. 483–485, ACM, 1967.

[122] B.-H. Park and H. Kargupta, "Distributed data mining: Algorithms, systems, and applications," pp. 341–358, 2002.

[123] T. Tamura, M. Oguchi, and M. Kitsuregawa, "Parallel database processing on a 100 node pc cluster: Cases for decision support query processing and data mining," *SC Conference*, vol. 0, p. 49, 1997.

[124] J. Han, M. Kamber, and J. Pei, *Data mining: concepts and techniques*. Morgan kaufmann, 2006.

[125] J. Waldo, G. Wyant, A. Wollrath, and S. Kendall, *A note on distributed computing*. Springer, 1997.

[126] Y. Lindell and B. Pinkas, "Privacy preserving data mining," in *Advances in Cryptology — CRYPTO 2000* (M. Bellare, ed.), vol. 1880, ch. Lecture Notes in Computer Science, pp. 36–54, Springer Berlin Heidelberg, 2000.

[127] R. Grossman and Y. Gu, "Data mining using high performance data clouds: Experimental studies using sector and sphere," in *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, KDD '08, (New York, NY, USA), pp. 920–927, ACM, 2008.

[128] M. J. Zaki and C.-T. Ho, *Large-scale parallel data mining*. No. 1759, Springer, 2000.

[129] N. Ailon and B. Chazelle, "Approximate nearest neighbors and the fast johnson-lindenstrauss transform," in *Proceedings of the thirty-eighth annual ACM symposium on Theory of computing*, pp. 557–563, 2006.

[130] A. Dasgupta, R. Kumar, and T. Sarlos, "A sparse johnson: Lindenstrauss transform," in *Proceedings of the Forty-second ACM Symposium on Theory of Computing*, STOC '10, (New York, NY, USA), pp. 341–350, ACM, 2010.

[131] G. Webb, "Multiboosting: A technique for combining boosting and wagging," *Machine Learning*, vol. 40, no. 2, pp. 159–196, 2000.

[132] S. Fortune, "A sweepline algorithm for voronoi diagrams," in *Proceedings of the Second Annual Symposium on Computational Geometry*, SCG '86, (New York, NY, USA), pp. 313–322, ACM, 1986.

[133] R. Klein, "Abstract voronoi diagrams and their applications," in *Computational Geometry and its Applications* (H. Noltemeier, ed.), vol. 333, ch. Lecture Notes in Computer Science, pp. 148–157, Springer Berlin Heidelberg, 1988.

[134] M. L. Gavrilova, "An explicit solution for computing the euclidean d-dimensional voronoi diagram of spheres in a floating-point arithmetic," in *Computational Science and Its Applications — ICCSA 2003* (V. Kumar, M. Gavrilova, C. Tan, and P. L'Ecuyer, eds.), vol. 2669, ch. Lecture Notes in Computer Science, pp. 827–835, Springer Berlin Heidelberg, 2003.

[135] R. Zass and A. Shashua, "A unifying approach to hard and probabilistic clustering," in *Computer Vision, 2005. ICCV 2005. Tenth IEEE International Conference on*, vol. 1, pp. 294–301, Oct 2005.

[136] E. Agrell, T. Eriksson, A. Vardy, and K. Zeger, "Closest point search in lattices," *Information Theory, IEEE Transactions on*, vol. 48, pp. 2201–2214, Aug 2002.

[137] S. Anders and W. Huber, "Differential expression analysis for sequence count data," *Genome biol*, vol. 11, no. 10, pp. –106, 2010.

[138] T. M. Forum, "Mpi: A message passing interface," 1993.

[139] M. R. Anderberg, "Cluster analysis for applications," tech. rep., 1973.

[140] W. J. Reed and B. D. Hughes, "From gene families and genera to incomes and internet file sizes: Why power laws are so common in nature," *Physical Review E*, vol. 66, no. 6, p. 067103, 2002.

[141] N. Ailon and E. Liberty, "Fast dimension reduction using rademacher series on dual bch codes," *Discrete & Computational Geometry*, vol. 42, no. 4, pp. 615–630, 2009.

[142] N. Ailon and E. Liberty, "An almost optimal unrestricted fast johnson-lindenstrauss transform," *ACM Trans. Algorithms*, vol. 9, pp. 21–1, jun 2013.

[143] E. Liberty, N. Ailon, and A. Singer, "Dense fast random projections and lean walsh transforms," in *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques* (A. Goel, K. Jansen, J. P. Rolim, and R. Rubinfeld, eds.), vol. 5171, ch. Lecture Notes in Computer Science, pp. 512–522, Springer Berlin Heidelberg, 2008.