References

[1] D. Achlioptas. Database-friendly random projections. In *ACM Principles of Database Systems*, pages 274–281, 2001.

[2] P. K. Agarwal, G. Cormode, Z. Huang, J. Phillips, Z. Wei, and K. Yi. Mergeable summaries. *ACM Transactions on Database Systems*, 38(4), 2013.

[3] P. K. Agarwal, S. Har-Peled, and K. R. Varadarajan. Approximating extent measures of points. *Journal of the ACM*, 51:606–635, 2004.

[4] P. K. Agarwal and R. Sharathkumar. Streaming algorithms for extent problems in high dimensions. In *ACM-SIAM Symposium on Discrete Algorithms*, 2010.

[5] P. K. Agarwal and H. Yu. A space-optimal data-stream algorithm for coresets in the plane. In *Symposium on Computational Geometry*, 2007.

[6] C. C. Aggarwal. On biased reservoir sampling in the presence of stream evolution. In *International Conference on Very Large Data Bases*, pages 607–618, 2006.

[7] K. J. Ahn, S. Guha, and A. McGregor. Analyzing graph structure via linear measurements. In *ACM-SIAM Symposium on Discrete Algorithms*, 2012.

[8] N. Ailon and B. Chazelle. Approximate nearest neighbors and the fast Johnson-Lindenstrauss transform. *SIAM J. Comput.*, 39(1):302–322, 2009.

[9] N. Alon, P. Gibbons, Y. Matias, and M. Szegedy. Tracking join and self-join sizes in limited storage. In *ACM Principles of Database Systems*, pages 10–20, 1999.

[10] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. In *ACM Symposium on Theory of Computing*, pages 20–29, 1996.

[11] N. Alon, Y. Matias, and M. Szegedy. The space complexity of approximating the frequency moments. *JCSS: Journal of Computer and System Sciences*, 58:137–147, 1999.

[12] D. Anderson, P. Bevan, K. Lang, E. Liberty, L. Rhodes, and J. Thaler. A high-performance algorithm for identifying frequent items in data streams. In *Proceedings of the 2017 Internet Measurement Conference*, IMC '17, pages 268–282, 2017.

[13] A. Andoni, P. Indyk, and I. Razenshteyn. Approximate nearest neighbor search in high dimensions. https://arxiv.org/abs/1806.09823, 2018.

[14] A. Andoni and H. L. Nguyên. Width of points in the streaming model. *ACM Trans. Algorithms*, 12(1):5:1–5:10, 2016.

[15] B. Aronov, E. Ezra, and M. Sharir. Small-size ε-nets for axis-parallel rectangles and boxes. *SIAM Journal on Computing*, 39(7):3248–3282, 2010.

[16] S. Arya, D. Mount, N. S. Netanyahu, R. Silverman, and A. Y. Wu. An optimal algorithm for approximate nearest neighbor searching fixed dimensions. *Journal of the ACM*, 45(6):891–923, 1998.

[17] N. Bansal. Constructive algorithms for discrepancy minimization. In *IEEE Conference on Foundations of Computer Science*, 2010.

[18] Z. Bar-Yossef, T. Jayram, R. Kumar, D. Sivakumar, and L. Trevisian. Counting distinct elements in a data stream. In *Proceedings of RANDOM 2002*, pages 1–10, 2002.

[19] Z. Bar-Yossef, R. Kumar, and D. Sivakumar. Reductions in streaming algorithms, with an application to counting triangles in graphs. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 623–632, 2002.

[20] N. Barkay, E. Porat, and B. Shalem. Feasible Sampling of Non-strict Turnstile Data Streams. In *Fundamentals of Computation Theory*, Sept. 2013.

[21] R. B. Basat, G. Einziger, R. Friedman, and Y. Kassner. Heavy hitters in streams and sliding windows. In *IEEE INFOCOMM*, 2016.

[22] R. Ben Basat, G. Einziger, and R. Friedman. Fast flow volume estimation. In *Proceedings of the International Conference on Distributed Computing and Networking (ICDCN)*, pages 44:1–44:10, 2018.

[23] R. Ben-Basat, G. Einziger, R. Friedman, and Y. Kassner. Optimal elephant flow detection. In *IEEE Conference on Computer Communications (INFOCOM)*, pages 1–9, 2017.

[24] J. L. Bentley and J. B. Saxe. Decomposable searching problems I: Static-to-dynamic transformation. *Journal of Algorithms*, 1:301–358, 1980.

[25] R. Berinde, G. Cormode, P. Indyk, and M. Strauss. Space-optimal heavy hitters with strong error bounds. In *ACM Principles of Database Systems*, 2009.

[26] K. S. Beyer, P. J. Haas, B. Reinwald, Y. Sismanis, and R. Gemulla. On synopses for distinct-value estimation under multiset operations. In *ACM SIGMOD International Conference on Management of Data*, pages 199–210, 2007.

[27] G. Bianchi, K. Duffy, D. J. Leith, and V. Shneer. Modeling conservative updates in multi-hash approximate count sketches. In *24th International Teletraffic Congress, ITC*, pages 1–8, 2012.

[28] L. Bledaite. Count-min sketches in real data applications. https://skillsmatter.com/skillscasts/6844-count-min-sketch-in-real-data-applications, 2015.

[29] B. Bloom. Space/time trade-offs in hash coding with allowable errors. *Communications of the ACM*, 13(7):422–426, July 1970.

[30] A. Blumer, A. Ehrenfeucht, D. Haussler, and M. Warmuth. Learnability and the Vapnik-Chervonenkis dimension. *Journal of the ACM*, 36:929–965, 1989.

[31] B. Bollobás. *Extremal Graph Theory*. Academic Press, 1978.

[32] P. Bose, E. Kranakis, P. Morin, and Y. Tang. Bounds for frequency estimation of packet streams. In *SIROCCO*, 2003.

[33] B. Boyer and J. Moore. A fast majority vote algorithm. Technical Report ICSCA-CMP-32, Institute for Computer Science, University of Texas, Feb. 1981.

[34] V. Braverman, G. Frahling, H. Lang, C. Sohler, and L. F. Yang. Clustering high dimensional dynamic data streams. In *Proceedings of the 34th International Conference on Machine Learning, ICML 2017, Sydney, NSW, Australia, 6-11 August 2017*, pages 576–585, 2017.

[35] V. Braverman and R. Ostrovsky. Smooth histograms for sliding windows. In *IEEE Conference on Foundations of Computer Science*, pages 283–293, 2007.

[36] V. Braverman and R. Ostrovsky. Zero-one frequency laws. In *ACM Symposium on Theory of Computing*, pages 281–290, 2010.

[37] V. Braverman, R. Ostrovsky, and D. Vilenchik. How hard is counting triangles in the streaming model? In *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 244–254, 2013.

[38] A. Z. Broder and M. Mitzenmacher. Network applications of Bloom filters: A survey. *Internet Mathematics*, 1(4), 2004.

[39] M. Bădoiu and K. L. Clarkson. Smaller core-sets for balls. In *ACM-SIAM Symposium on Discrete Algorithms*, 2003.

[40] M. Bădoiu and K. L. Clarkson. Optimal core-sets for balls. *Computational Geometry: Theory and Applications*, 40(1):14–22, 2008.

[41] M. Bădoiu, S. Har-Peled, and P. Indyk. Approximate clustering via core-sets. In *ACM Symposium on Theory of Computing*, 2002.

[42] L. S. Buriol, G. Frahling, S. Leonardi, A. Marchetti-Spaccamela, and C. Sohler. Counting triangles in data streams. In *ACM Principles of Database Systems*, 2006.

[43] D. Cai, M. Mitzenmacher, and R. P. Adams. A bayesian nonparametric view on count-min sketch. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 8782–8791, 2018.

[44] J. L. Carter and M. N. Wegman. Universal classes of hash functions. *Journal of Computer and System Sciences*, 18(2):143–154, 1979.

[45] J. Chambers, C. Mallows, and B. Stuck. A method for simulating stable random variables. *Journal of the American Statistical Association*, 71(354):340–344, 1976.

[46] T. M. Chan. Faster core-set constructions and data-stream algorithms in fixed dimensions. *Computational Geometry: Theory and Applications*, 35:20–35, 2006.

[47] T. M. Chan and V. Pathak. Streaming and dynamic algorithms for minimum enclosing balls in high dimensions. In *International Symposium on Algorithms and Data Structures*, 2011.

[48] K. Chandra. View counting at reddit. https://redditblog.com/2017/05/24/view-counting-at-reddit/, 2017.

[49] M. Charikar, K. Chen, and M. Farach-Colton. Finding frequent items in data streams. In *Procedings of the International Colloquium on Automata, Languages and Programming (ICALP)*, 2002.

[50] K. Chen and S. Rao. An improved frequent items algorithm with applications to web caching. Technical Report UCB/CSD-05-1383, EECS Department, University of California, Berkeley, 2005.

[51] K. L. Clarkson and D. P. Woodruff. Numerical linear algebra in the streaming model. In *ACM Symposium on Theory of Computing*, pages 205–214, 2009.

[52] E. Cohen. Size-estimation framework with applications to transitive closure and reachability. *J. Comput. Syst. Sci.*, 55(3):441–453, 1997.

[53] E. Cohen. All-distances sketches, revisited: HIP estimators for massive graphs analysis. *IEEE Transactions on Knowledge and Data Engineering*, 27(9):2320–2334, 2015.

[54] E. Cohen, N. Duffield, H. Kaplan, C. Lund, and M. Thorup. Efficient stream sampling for variance-optimal estimation of subset sums. *SIAM Journal on Computing*, 40(5):1402–1431, 2011.

[55] E. Cohen and M. Strauss. Maintaining time-decaying stream aggregates. In *ACM Principles of Database Systems*, 2003.

[56] S. Cohen and Y. Matias. Spectral Bloom filters. In *ACM SIGMOD International Conference on Management of Data*, 2003.

[57] J. Considine, M. Hadjieleftheriou, F. Li, J. W. Byers, and G. Kollios. Robust approximate aggregation in sensor data management systems. *ACM Transactions on Database Systems*, 34(1):6:1–6:35, 2009.

[58] D. Coppersmith and R. Kumar. An improved data stream algorithm for frequency moments. In *ACM-SIAM Symposium on Discrete Algorithms*, 2004.

[59] T. H. Cormen, C. E. Leiserson, and R. L. Rivest. *Introduction to Algorithms*. MIT Press, 1990.

[60] G. Cormode, M. Datar, P. Indyk, and S. Muthukrishnan. Comparing data streams using Hamming norms. *IEEE Transactions on Knowledge and Data Engineering*, 15(3):529–541, 2003.

[61] G. Cormode and D. Firmani. On unifying the space of ℓ_0-sampling algorithms. In *Algorithm Engineering and Experiments*, 2013.

[62] G. Cormode and M. Garofalakis. Sketching streams through the net: Distributed approximate query tracking. In *International Conference on Very Large Data Bases*, 2005.

[63] G. Cormode, M. Garofalakis, and D. Sacharidis. Fast approximate wavelet tracking on streams. In *International Conference on Extending Database Technology*, pages 4–22, 2006.

[64] G. Cormode and M. Hadjieleftheriou. Finding frequent items in data streams. In *International Conference on Very Large Data Bases*, 2008.

[65] G. Cormode and H. Jowhari. A second look at counting triangles in graph streams (corrected). *Theor. Comput. Sci.*, 683:22–30, 2017.

[66] G. Cormode and H. Jowhari. l_p samplers and their applications: a survey. *ACM Computing Surveys*, 2019.

[67] G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava. Space- and time-efficient deterministic algorithms for biased quantiles over data streams. In *ACM Principles of Database Systems*, 2006.

[68] G. Cormode, F. Korn, and S. Tirthapura. Exponentially decayed aggregates on data streams. In *IEEE International Conference on Data Engineering*, 2008.

[69] G. Cormode and S. Muthukrishnan. Improved data stream summary: The Count-Min sketch and its applications. Technical Report 2003-20, DIMACS, 2003.

[70] G. Cormode and S. Muthukrishnan. What's hot and what's not: Tracking most frequent items dynamically. In *ACM Principles of Database Systems*, pages 296–306, 2003.

[71] G. Cormode and S. Muthukrishnan. What's new: Finding significant differences in network data streams. In *Proceedings of IEEE Infocom*, 2004.

[72] G. Cormode and S. Muthukrishnan. An improved data stream summary: The Count-Min sketch and its applications. *Journal of Algorithms*, 55(1):58–75, 2005.

[73] G. Cormode and S. Muthukrishnan. Space efficient mining of multigraph streams. In *ACM Principles of Database Systems*, 2005.

[74] G. Cormode, S. Muthukrishnan, and I. Rozenbaum. Summarizing and mining inverse distributions on data streams via dynamic inverse sampling. In *International Conference on Very Large Data Bases*, 2005.

[75] G. Cormode, S. Tirthapura, and B. Xu. Time-decaying sketches for sensor data aggregation. In *ACM Conference on Principles of Distributed Computing (PODC)*, 2007.

[76] S. Das, S. Antony, D. Agrawal, and A. E. Abbadi. Cots: A scalable framework for parallelizing frequency counting over data streams. In *IEEE International Conference on Data Engineering*, 2009.

[77] A. Dasgupta, K. J. Lang, L. Rhodes, and J. Thaler. A framework for estimating stream expression cardinalities. In *International Conference on Database Theory*, pages 6:1–6:17, 2016.

[78] S. Dasgupta and A. Gupta. An elementary proof of a theorem of johnson and lindenstrauss. *Random Structures and Algorithms*, 22(1):60–65, 2003.

[79] M. Datar, A. Gionis, P. Indyk, and R. Motwani. Maintaining stream statistics over sliding windows. In *ACM-SIAM Symposium on Discrete Algorithms*, 2002.

[80] M. Datar, N. Immorlica, P. Indyk, and V. S. Mirrokni. Locality-sensitive hashing scheme based on p-stable distributions. In *Symposium on Computational Geometry*, 2004.

[81] E. Demaine, A. López-Ortiz, and J. I. Munro. Frequency estimation of internet packet streams with limited space. In *European Symposium on Algorithms (ESA)*, 2002.

[82] F. Deng and D. Rafiei. New estimation algorithms for streaming data: Count-Min can do more. Unpublished manuscript.

[83] M. Dietzfelbinger, A. Goerdt, M. Mitzenmacher, A. Montanari, R. Pagh, and M. Rink. Tight thresholds for cuckoo hashing via XORSAT. In *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 213–225, 2010.

[84] A. Dobra and F. Rusu. Sketches for size of join estimation. In *ACM Transactions on Database Systems*, 2008.

[85] D. Donoho. Compressed sensing. *IEEE Transactions on on Information Theory*, 52(4):1289–1306, April 2006.

[86] P. Drineas, M. Magdon-Ismail, M. W. Mahoney, and D. P. Woodruff. Fast approximation of matrix coherence and statistical leverage. *Journal of Machine Learning Research*, 13:3475–3506, 2012.

[87] N. Duffield, C. Lund, and M. Thorrup. Priority sampling for estimation of arbitrary subset sums. *Journal of the ACM*, 54(6), 2007.

[88] N. Duffield, C. Lund, and M. Thorup. Estimating flow distributions from sampled flow statistics. In *Proceedings of ACM SIGCOMM*, 2003.

[89] M. Durand and P. Flajolet. Loglog counting of large cardinalities (extended abstract). In *European Symposium on Algorithms (ESA)*, pages 605–617, 2003.

[90] R. Durstenfeld. Algorithm 235: Random permutation. *Communications of the ACM*, 7(7):420, 1964.

[91] O. Egecioglu and B. Kalantari. Approximating the diameter of a set of points in the Euclidean space. *Information Processing Letters (IPL)*, 32:205–211, 1989.

[92] G. Einziger and R. Friedman. A formal analysis of conservative update based approximate counting. In *International Conference on Computing, Networking and Communications, ICNC*, pages 255–259, 2015.

[93] M. Elkin. Streaming and fully dynamic centralized algorithms for constructing and maintaining sparse spanners. *ACM Transactions on Algorithms*, 7(2):20, 2011.

[94] D. Eppstein and M. T. Goodrich. Straggler identification in round-trip data streams via newton's identities and invertible bloom filters. *IEEE Transactions on Knowledge and Data Engineering*, 23(2):297–306, 2011.

[95] Ú. Erlingsson, V. Pihur, and A. Korolova. RAPPOR: randomized aggregatable privacy-preserving ordinal response. In *Computer and Communications Security*, pages 1054–1067, 2014.

[96] C. Estan and G. Varghese. New directions in traffic measurement and accounting. In *Proceedings of ACM SIGCOMM*, volume 32, 4 of *Computer Communication Review*, pages 323–338, 2002.

[97] L. Fan, P. Cao, J. Almeida, and A. Broder. Summary cache: A scalable wide-area web cache sharing protocol. In *IEEE INFOCOMM*, 1998.

[98] J. Feigenbaum, S. Kannan, A. McGregor, S. Suri, and J. Zhang. Graph distances in the streaming model: The value of space. In *ACM-SIAM Symposium on Discrete Algorithms*, 2005.

[99] D. Felber and R. Ostrovsky. A randomized online quantile summary in $O(1/\epsilon \log(1/\epsilon))$ words. In *APPROX-RANDOM*, 2015.

[100] P. Flajolet. Approximate counting: A detailed analysis. *BIT*, 25:113–134, 1985.

[101] P. Flajolet, E. Fusy, O. Gandouet, and F. Meunier. Hyperloglog: The analysis of a near-optimal cardinality estimation algorithm. In *Analysis of Algorithms (AOFA)*, pages 127–146, 2007.

[102] P. Flajolet and G. N. Martin. Probabilistic counting. In *IEEE Conference on Foundations of Computer Science*, pages 76–82, 1983. Journal version in *Journal of Computer and System Sciences*, 31:182–209, 1985.

[103] P. Flajolet and G. N. Martin. Probabilistic counting algorithms for database applications. *Journal of Computer and System Sciences*, 31:182–209, 1985.

[104] G. Frahling, P. Indyk, and C. Sohler. Sampling in dynamic data streams and applications. In *Symposium on Computational Geometry*, June 2005.

[105] S. Ganguly. Counting distinct items over update streams. In *International Sympoisum on Algorithms and Computation (ISAAC)*, pages 505–514, 2005.

[106] M. Ghashami, E. Liberty, J. M. Phillips, and D. P. Woodruff. Frequent directions: Simple and deterministic matrix sketching. *SIAM Journal on Computing*, 45(5):1762–1792, 2016.

[107] M. Ghashami and J. M. Phillips. Relative errors for deterministic low-rank matrix approximations. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 707–717, 2014.

[108] P. Giannopoulos, C. Knauer, M. Wahlstrom, and D. Werner. Hardness of discrepancy computation and ε-net verification in high dimension. *Journal of Complexity*, 28(2):162–176, 2012.

[109] P. Gibbons and S. Tirthapura. Estimating simple functions on the union of data streams. In *ACM Symposium on Parallel Algorithms and Architectures (SPAA)*, pages 281–290, 2001.

[110] A. Gilbert, S. Guha, P. Indyk, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Fast, small-space algorithms for approximate histogram maintenance. In *ACM Symposium on Theory of Computing*, pages 389–398, 2002.

[111] A. Gilbert, S. Guha, P. Indyk, S. Muthukrishnan, and M. Strauss. Near-optimal sparse Fourier representation via sampling. In *ACM Symposium on Theory of Computing*, 2002.

[112] A. Gilbert, Y. Kotidis, S. Muthukrishnan, and M. Strauss. Surfing wavelets on streams: One-pass summaries for approximate aggregate queries. In *International Conference on Very Large Data Bases*, pages 79–88, 2001. Journal version in *IEEE Transactions on Knowledge and Data Engineering*, 15(3):541–554, 2003.

[113] A. C. Gilbert and P. Indyk. Sparse recovery using sparse matrices. *Proceedings of the IEEE*, 98(6):937–947, 2010.

[114] A. C. Gilbert, Y. Kotidis, S. Muthukrishnan, and M. J. Strauss. How to summarize the universe: Dynamic maintenance of quantiles. In *In VLDB*, pages 454–465, 2002.

[115] A. Gionis, P. Indyk, and R. Motwani. Similarity search in high dimensions via hashing. In *International Conference on Very Large Data Bases*, pages 518–529, 1999.

[116] A. Goel, P. Indyk, and K. Varadarajan. Reductions among high dimensional proximity problems. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 769–778, 2001.

[117] L. Golab and M. T. Özsu. Issues in data stream management. *SIGMOD Record (ACM Special Interest Group on Management of Data)*, 32(2):5–14, June 2003.

[118] M. T. Goodrich and M. Mitzenmacher. Invertible bloom lookup tables. In *Annual Allerton Conference on Communication, Control, and Computing*, pages 792–799, 2011.

[119] M. Greenwald and S. Khanna. Space-efficient online computation of quantile summaries. In *ACM SIGMOD International Conference on Management of Data*, 2001.

[120] M. Greenwald and S. Khanna. Power-conserving computation of order-statistics over sensor networks. In *ACM Principles of Database Systems*, 2004.

[121] A. Gronemeier and M. Sauerhoff. Applying approximate counting for computing the frequency moments of long data streams. *Theory of Computer Systems*, 44(3):332–348, 2009.

[122] S. Guha. Tight results for clustering and summarizing data streams. In *International Conference on Database Theory*, 2009.

[123] A. Hall, O. Bachmann, R. Büssow, S. Ganceanu, and M. Nunkesser. Processing a trillion cells per mouse click. *PVLDB*, 5(11):1436–1446, 2012.

[124] S. Har-Peled, P. Indyk, and R. Motwani. Approximate nearest neighbor: Towards removing the curse of dimensionality. *Theory of Computing*, 8:321–350, 2012.

[125] H. Hassanieh, P. Indyk, D. Katabi, and E. Price. Simple and practical algorithm for sparse fourier transform. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 1183–1194, 2012.

[126] D. Haussler and E. Welzl. Epsilon-nets and simplex range queries. *Discrete and Computational Geometry*, 2:127–151, 1987.

[127] S. Heule, M. Nunkesser, and A. Hall. Hyperloglog in practice: algorithmic engineering of a state of the art cardinality estimation algorithm. In *International Conference on Extending Database Technology*, pages 683–692, 2013.

[128] Z. Huang, L. Wang, K. Yi, and Y. Liu. Sampling based algorithms for quantile computation in sensor networks. In *ACM SIGMOD International Conference on Management of Data*, 2011.

[129] Z. Huang and K. Yi. The communication complexity of distributed epsilon-approximations. In *IEEE Conference on Foundations of Computer Science*, 2014.

[130] Z. Huang, K. Yi, Y. Liu, and G. Chen. Optimal sampling algorithms for frequency estimation in distributed data. In *IEEE INFOCOMM*, 2011.

[131] R. Y. S. Hung and H. F. Ting. An $\omega(1/\epsilon \log 1/\epsilon)$ space lower bound for finding ϵ-approximate quantiles in a data stream. In *Proceedings of the 4th International Conference on Frontiers in Algorithmics*, pages 89–100, 2010.

[132] P. Indyk. Algorithmic aspects of geometric embeddings (invited tutorial). In *IEEE Conference on Foundations of Computer Science*, pages 10–35, 2001.

[133] P. Indyk. A small approximately min-wise independent family of hash functions. *Journal of Algorithms*, 38(1):84–90, 2001.

[134] P. Indyk and R. Motwani. Approximate nearest neighbors: Towards removing the curse of dimensionality. In *ACM Symposium on Theory of Computing*, pages 604–613, 1998.

[135] N. Ivkin, E. Liberty, K. Lang, Z. Karnin, and V. Braverman. Streaming quantiles algorithms with small space and update time. In *To appear*, 2019.

[136] R. Jayaram and D. P. Woodruff. Perfect lp sampling in a data stream. In *IEEE Conference on Foundations of Computer Science*, 2018.

[137] T. S. Jayram. Information complexity: a tutorial. In *ACM Principles of Database Systems*, pages 159–168, 2010.

[138] T. S. Jayram, R. Kumar, and D. Sivakumar. The one-way communication complexity of gap hamming distance. `http://www.madalgo.au.dk/img/SumSchoo2007_Lecture_20slides/Bibliography/p14_Jayram_07_Manusc_ghd.pdf`, 2007.

[139] T. S. Jayram and D. P. Woodruff. The data stream space complexity of cascaded norms. In *IEEE Conference on Foundations of Computer Science*, pages 765–774, 2009.

[140] T. S. Jayram and D. P. Woodruff. Optimal bounds for johnson-lindenstrauss transforms and streaming problems with low error. In *ACM-SIAM Symposium on Discrete Algorithms*, 2011.

[141] C. Jin, W. Qian, C. Sha, J. X. Yu, and A. Zhou. Dynamically maintaining frequent items over a data stream. In *CIKM*, 2003.

[142] W. Johnson and J. Lindenstrauss. Extensions of Lipshitz mapping into Hilbert space. *Contemporary Mathematics*, 26:189–206, 1984.

[143] H. Jowhari and M. Ghodsi. New streaming algorithms for counting triangles in graphs. In *International Conference on Computing and Combinatorics*, pages 710–716, 2005.

[144] H. Jowhari, M. Saglam, and G. Tardos. Tight bounds for lp samplers, finding duplicates in streams, and related problems. In *ACM Principles of Database Systems*, 2011.

[145] B. Kalyanasundaram and G. Schnitger. The probabilistic communication complexity of set intersection. *SIAM Journal on Discrete Mathematics*, 5(4):545–557, 1992.

[146] D. M. Kane and J. Nelson. Sparser johnson-lindenstrauss transforms. In *ACM-SIAM Symposium on Discrete Algorithms*, 2012.

[147] B. M. Kapron, V. King, and B. Mountjoy. Dynamic graph connectivity in polylogarithmic worst case time. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 1131–1142, 2013.

[148] Z. Karnin, K. Lang, and E. Liberty. Optimal quantile approximation in streams. In *IEEE Conference on Foundations of Computer Science*, 2016.

[149] R. Karp, C. Papadimitriou, and S. Shenker. A simple algorithm for finding frequent elements in sets and bags. *ACM Transactions on Database Systems*, 28:51–55, 2003.

[150] R. M. Karp and M. O. Rabin. Efficient randomized pattern-matching algorithms. *IBM Journal of Research and Development*, 31(2):249–260, 1987.

[151] A. Kirsch and M. Mitzenmacher. Less hashing, same performance: Building a better bloom filter. In *European Symposium on Algorithms (ESA)*, pages 456–467, 2006.

[152] D. E. Knuth. *The Art of Computer Programming, Vol 2, Fundamental Algorithms*. Addison-Wesley, 2nd edition, 1998.

[153] D. E. Knuth. *The Art of Computer Programming, Vol 2, Seminumerical Algorithms*. Addison-Wesley, 2nd edition, 1998.

[154] G. Kollios, J. Byers, J. Considine, M. Hadjieleftheriou, and F. Li. Robust aggregation in sensor networks. *IEEE Data Engineering Bulletin*, 28(1), Mar. 2005.

[155] J. Komlós, J. Pach, and G. Woeginger. Almost tight bounds for ε-nets. *Discrete and Computational Geometry*, 7:163–173, 1992.

[156] P. Kumar, J. S. B. Mitchell, and E. A. Yildirim. Approximate minimum enclosing balls in high dimensions using core-sets. *ACM Journal of Experimental Algorithmics*, 8, 2003.

[157] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 1997.

[158] E. Kushilevitz, R. Ostrovsky, and Y. Rabani. Efficient search for approximate nearest neighbor in high dimensional spaces. In *ACM Symposium on Theory of Computing*, pages 614–623, 1998.

[159] K. J. Lang. Back to the future: an even more nearly optimal cardinality estimation algorithm. Technical report, ArXiV, 2017.

[160] K. G. Larsen, J. Nelson, H. L. Nguyen, and M. Thorup. Heavy hitters via cluster-preserving clustering. In *IEEE Conference on Foundations of Computer Science*, pages 61–70, 2016.

[161] G. M. Lee, H. Liu, Y. Yoon, and Y. Zhang. Improving sketch reconstruction accuracy using linear least squares method. In *Internet Measurement Conference (IMC)*, 2005.

[162] L. Lee and H. Ting. A simpler and more efficient deterministic scheme for finding frequent items over sliding windows. In *ACM Principles of Database Systems*, 2006.

[163] P. Li. Very sparse stable random projections, estimators and tail bounds for stable random projections. Technical Report cs.DS/0611114, ArXiV, 2006.

[164] Y. Li, P. Long, and A. Srinivasan. Improved bounds on the sample complexity of learning. *Journal of Computer and System Sciences*, 62(3):516–527, 2001.

[165] E. Liberty. Simple and deterministic matrix sketching. In *ACM SIGKDD*, pages 581–588, 2013.

[166] R. J. Lipton. Fingerprinting sets. Technical Report CS-TR-212-89, Princeton, 1989.

[167] Y. Lu, A. Montanari, S. Dharmapurikar, A. Kabbani, and B. Prabhakar. Counter braids: A novel counter architecture for per-flow measurement. In *ACM SIGMETRICS*, 2008.

[168] J. O. Lumbroso. How flajolet processed streams with coin flips. Technical Report 1805.00612, ArXiV, 2018.

[169] A. Manjhi, V. Shkapenyuk, K. Dhamdhere, and C. Olston. Finding (recently) frequent items in distributed data streams. In *IEEE International Conference on Data Engineering*, pages 767–778, 2005.

[170] G. S. Manku, S. Rajagopalan, and B. G. Lindsay. Approximate medians and other quantiles in one pass and with limited memory. In *ACM SIGMOD International Conference on Management of Data*, pages 426–435, 1998.

[171] G. S. Manku, S. Rajagopalan, and B. G. Lindsay. Random sampling techniques for space efficient online computation of order statistics of large datasets. In *ACM SIGMOD International Conference on Management of Data*, pages 251–262, 1999.

[172] J. Matoušek. Tight upper bounds for the discrepancy of halfspaces. *Discrete and Computational Geometry*, 13:593–601, 1995.

[173] A. McGregor, S. Vorotnikova, and H. T. Vu. Better algorithms for counting triangles in data streams. In *ACM Principles of Database Systems*, pages 401–411, 2016.

[174] D. McIlroy. Development of a spelling list. Technical report, Bell Labs, 1982.

[175] A. Metwally, D. Agrawal, and A. E. Abbadi. Efficient computation of frequent and top-k elements in data streams. In *International Conference on Database Theory*, 2005.

[176] J. Misra and D. Gries. Finding repeated elements. *Science of Computer Programming*, 2:143–152, 1982.

[177] M. Mitzenmacher. *Bloom Filters*, pages 252–255. Springer, 2009.

[178] M. Mitzenmacher and E. Upfal. *Probability and Computing: Randomized Algorithms and Probabilistic Analysis*. Cambridge University Press, 2005.

[179] M. Mitzenmacher and G. Varghese. Biff (bloom filter) codes: Fast error correction for large data sets. In *IEEE International Symposium on Information Theory (ISIT)*, pages 483–487, 2012.

[180] M. Molloy. Cores in random hypergraphs and boolean formulas. *Random Structures and Algorithms*, 27(1):124–135, 2005.

[181] M. Monemizadeh and D. P. Woodruff. 1-pass relative-error l_p-sampling with applications. In *ACM-SIAM Symposium on Discrete Algorithms*, 2010.

[182] R. Morris. Counting large numbers of events in small registers. *Communications of the ACM*, 21(10):840–842, 1977.

[183] S. Moser and P. N. Chen. *A Student's Guide to Coding and Information Theory*. Cambridge University Press, 2012.

[184] R. Motwani and P. Raghavan. *Randomized Algorithms*. Cambridge University Press, 1995.

[185] D. Mount and S. Arya. ANN: Library for approximate nearest neighbor searching. Technical report, University of Maryland, 2010.

[186] J. I. Munro and M. S. Paterson. Selection and sorting with limited storage. *Theoretical Computer Science*, 12:315–323, 1980.

[187] J. Nelson and H. L. Nguyen. OSNAP: faster numerical linear algebra algorithms via sparser subspace embeddings. In *IEEE Conference on Foundations of Computer Science*, pages 117–126, 2013.

[188] J. Nelson and H. L. Nguyen. Sparsity lower bounds for dimensionality reducing maps. In *ACM Symposium on Theory of Computing*, pages 101–110, 2013.

[189] J. Nelson and D. Woodruff. Fast manhattan sketches in data streams. In *ACM Principles of Database Systems*, 2010.

[190] R. O'Donnell, Y. Wu, and Y. Zhou. Optimal lower bounds for locality-sensitive hashing (except when q is tiny). *ACM Transactions on Computation Theory*, 6(1):5, 2014.

[191] J. Pach and G. Tardos. Tight lower bounds for the size of epsilon-nets. *J. Amer. Math. Soc.*, 26:645–658, 2013.

[192] R. Pagh. Compressed matrix multiplication. In *ITCS*, pages 442–451, 2012.

[193] A. Pavan, K. Tangwongsan, S. Tirthapura, and K. Wu. Counting and sampling triangles from a graph stream. *PVLDB*, 6(14):1870–1881, 2013.

[194] A. Pavan and S. Tirthapura. Range-efficient counting of distinct elements in a massive data stream. *SIAM Journal on Computing*, 37(2):359–379, 2007.

[195] N. Pham and R. Pagh. Fast and scalable polynomial kernels via explicit feature maps. In *ACM SIGKDD*, pages 239–247, 2013.

[196] R. Pike, S. Dorward, R. Griesemer, and S. Quinlan. Interpreting the data: Parallel analysis with sawzall. *Dynamic Grids and Worldwide Computing*, 13(4):277–298, 2005.

[197] A. A. Razborov. On the distributional complexity of disjointness. *Theoretical Computer Science*, 106(2):385–390, 1992.

[198] T. Sarlós. Improved approximation algorithms for large matrices via random projections. In *IEEE Conference on Foundations of Computer Science*, pages 143–152, 2006.

[199] C.-E. Särndal, B. Swensson, and J. Wretman. *Model Assisted Survey Sampling*. Springer, 1992.

[200] S. E. Schechter, C. Herley, and M. Mitzenmacher. Popularity is everything: A new approach to protecting passwords from statistical-guessing attacks. In *5th USENIX Workshop on Hot Topics in Security*, 2010.

[201] J. P. Schmidt, A. Siegel, and A. Srinivasan. Chernoff-Hoeffding bounds for applications with limited independence. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 331–340, 1993.

[202] R. Schweller, Z. Li, Y. Chen, Y. Gao, A. Gupta, Y. Zhang, P. A. Dinda, M.-Y. Kao, and G. Memik. Reversible sketches: enabling monitoring and analysis over high-speed data streams. *IEEE Transactions on Networks*, 15(5):1059–1072, 2007.

[203] Q. Shi, J. Petterson, G. Dror, J. Langford, A. J. Smola, and S. V. N. Vishwanathan. Hash kernels for structured data. *Journal of Machine Learning Research*, 10:2615–2637, 2009.

[204] N. Shrivastava, C. Buragohain, D. Agrawal, and S. Suri. Medians and beyond: New aggregation techniques for sensor networks. In *ACM SenSys*, 2004.

[205] O. Simpson, C. Seshadhri, and A. McGregor. Catching the head, tail,

and everything in between: A streaming algorithm for the degree distribution. In *IEEE International Conference on Data Mining*, pages 979–984, 2015.

[206] A. Srinivasan. Improving the discrepancy bound for sparse matrices: Better approximations for sparse lattice approximation problems. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 692–701, 1997.

[207] S. Suri, C. D. Tóth, and Y. Zhou. Range counting over multidimensional data streams. *Discrete and Computational Geometry*, 26(4):633–655, 2006.

[208] M. Szegedy. The dlt priority sampling is essentially optimal. In *ACM Symposium on Theory of Computing*, 2006.

[209] M. Szegedy and M. Thorup. On the variance of subset sum estimation. In *European Symposium on Algorithms (ESA)*, 2007.

[210] M. Talagrand. Sharper bounds for gaussian and empirical processes. *The Annals of Probability*, 22(1):28–76, 1994.

[211] D. P. Team. Learning with privacy at scale. *Apple Machine Learning Journal*, 1(8), Dec. 2017.

[212] M. Thorup. Even strongly universal hashing is pretty fast. In *ACM-SIAM Symposium on Discrete Algorithms*, 2000.

[213] M. Thorup. Equivalence between priority queues and sorting. *Journal of the ACM*, 54(6), 2007.

[214] M. Thorup and Y. Zhang. Tabulation based 4-universal hashing with applications to second moment estimation. In *ACM-SIAM Symposium on Discrete Algorithms*, 2004.

[215] D. Ting. Count-min: Optimal estimation and tight error bounds using empirical error distributions. In *ACM SIGKDD*, pages 2319–2328, 2018.

[216] S. Tirthapura and D. P. Woodruff. Rectangle-efficient aggregation in spatial data streams. In *ACM Principles of Database Systems*, pages 283–294, 2012.

[217] S. Tirthapura and D. P. Woodruff. A general method for estimating correlated aggregates over a data stream. *Algorithmica*, 73(2):235–260, 2015.

[218] A. Tridgell and P. Mackerras. The rsync algorithm. Technical Report TR-CS-96-05, Department of Computer Science, The Australian National University, 1996.

[219] I. W. Tsang, J. T. Kwok, and P.-M. Cheung. Core vector machines: Fast SVM training on very large data sets. *Journal of Machine Learning Research*, 6:363–392, 2005.

[220] V. N. Vapnik and A. Y. Chervonenkis. On the uniform convergence of relative frequencies of events to their probabilities. *Theory of Probability and its Applications*, 16:264–280, 1971.

[221] S. Venkataraman, D. X. Song, P. B. Gibbons, and A. Blum. New streaming algorithms for fast detection of superspreaders. In *Proceedings of the Network and Distributed System Security Symposium, NDSS 2005, San Diego, California, USA*, 2005.

[222] J. S. Vitter. Random sampling with a reservoir. *ACM Transactions on Mathematical Software*, 11(1):37–57, Mar. 1985.

[223] J. Wang, W. Liu, S. Kumar, and S.-F. Chang. Learning to hash for indexing big data: a survey. *Proceedings of the IEEE*, 104(1):34–57, 2016.

[224] L. Wang, G. Luo, K. Yi, and G. Cormode. Quantiles over data streams: An experimental study. In *ACM SIGMOD International Conference on Management of Data*, 2013.

[225] K. Y. Whang, B. T. Vander-Zanden, and H. M. Taylor. A linear-time probabilistic counting algorithm for database applications. *ACM Transactions on Database Systems*, 15(2):208, 1990.

[226] D. Woodruff. Optimal space lower bounds for all frequency moments. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 167–175, 2004.

[227] D. P. Woodruff. Low rank approximation lower bounds in row-update streams. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 1781–1789, 2014.

[228] D. P. Woodruff. Sketching as a tool for numerical linear algebra. *Found. Trends Theor. Comput. Sci.*, 10(1–2):1–157, Oct. 2014.

[229] D. P. Woodruff and Q. Zhang. Tight bounds for distributed functional monitoring. In *ACM Symposium on Theory of Computing*, 2012.

[230] D. P. Woodruff and Q. Zhang. Subspace embeddings and ℓ_p-regression using exponential random variables. In *Conference on Learning Theory*, pages 546–567, 2013.

[231] H. Yu, P. K. Agarwal, R. Poreddy, and K. R. Varadarajan. Practical methods for shape fitting and kinetic data structures using coresets. *Algorithmica*, 52(3):378–402, 2008.

[232] H. Zarrabi-Zadeh. An almost space-optimal streaming algorithm for coresets in fixed dimensions. *Algorithmica*, 60(1):46–59, 2011.

[233] Q. Zhang, J. Pell, R. Canino-Koning, A. C. Howe, and C. T. Brown. These are not the k-mers you are looking for: Efficient online k-mer counting using a probabilistic data structure. *PLOS ONE*, 9(7):1–13, 07 2014.

[234] Y. Zhang, S. Singh, S. Sen, N. Duffield, and C. Lund. Online identification of hierarchical heavy hitters: Algorithms, evaluation and applications. In *Internet Measurement Conference (IMC)*, 2004.

[235] Q. Zhao, M. Ogihara, H. Wang, and J. Xu. Finding global icebergs over distributed data sets. In *ACM Principles of Database Systems*, 2006.

[236] V. M. Zolotarev. *One Dimensional Stable Distributions*, volume 65 of *Translations of Mathematical Monographs*. American Mathematical Society, 1983.

Index