Graham Cormode

List of Publications by Year
in descending order

152 papers	8,368 citations	159358 30 h-index	91712 69 g-index
157 all docs	157 docs citations	157 times ranked	3704 citing authors

#	Article	IF	Citations
1	An improved data stream summary: the count-min sketch and its applications. Journal of Algorithms, 2005, 55, 58-75.	0.9	1,326
2	Key differences between Web 1.0 and Web 2.0. First Monday, 0, , .	0.6	341
3	Synopses for Massive Data: Samples, Histograms, Wavelets, Sketches. Foundations and Trends in Databases, 2011, 4, 1-294.	4.1	277
4	What's hot and what's not: tracking most frequent items dynamically. ACM Transactions on Database Systems, 2005, 30, 249-278.	1.5	269
5	Differentially Private Spatial Decompositions. , 2012, , .		268
6	PrivBayes. ACM Transactions on Database Systems, 2017, 42, 1-41.	1.5	197
7	Finding frequent items in data streams. Proceedings of the VLDB Endowment, 2008, 1, 1530-1541.	2.1	190
8	Finding the frequent items in streams of data. Communications of the ACM, 2009, 52, 97-105.	3.3	185
9	Holistic aggregates in a networked world. , 2005, , .		156
10	DPT. Proceedings of the VLDB Endowment, 2015, 8, 1154-1165.	2.1	147
11	Class-based graph anonymization for social network data. Proceedings of the VLDB Endowment, 2009, 2, 766-777.	2.1	140
12	PrivBayes. , 2014, , .		140
13	Communication-efficient distributed monitoring of thresholded counts. , 2006, , .		134
14	Methods for finding frequent items in data streams. VLDB Journal, 2010, 19, 3-20.	2.7	131
15	Semantics of Ranking Queries for Probabilistic Data and Expected Ranks. Proceedings - International Conference on Data Engineering, 2009, , .	0.0	130
16	Practical verified computation with streaming interactive proofs. , 2012, , .		129
17	What's new: finding significant differences in network data streams. IEEE/ACM Transactions on Networking, 2005, 13, 1219-1232.	2.6	123
18	What's hot and what's not. , 2003, , .		122

#	Article	IF	Citations
19	Anonymizing bipartite graph data using safe groupings. Proceedings of the VLDB Endowment, 2008, 1, 833-844.	2.1	122
20	Finding Hierarchical Heavy Hitters in Data Streams. , 2003, , 464-475.		121
21	Sketching probabilistic data streams. , 2007, , .		120
22	Privacy at Scale. , 2018, , .		119
23	Conservative or liberal? Personalized differential privacy. , 2015, , .		113
24	Mergeable summaries. ACM Transactions on Database Systems, 2013, 38, 1-28.	1.5	107
25	An Improved Data Stream Summary: The Count-Min Sketch and Its Applications. Lecture Notes in Computer Science, 2004, , 29-38.	1.0	107
26	Space efficient mining of multigraph streams. , 2005, , .		101
27	The string edit distance matching problem with moves. ACM Transactions on Algorithms, 2007, 3, 1-19.	0.9	97
28	Approximation algorithms for clustering uncertain data. , 2008, , .		92
29	Combinatorial Algorithms for Compressed Sensing. , 2006, , .		88
30	Holistic UDAFs at streaming speeds. , 2004, , .		87
31	Summarizing and Mining Skewed Data Streams. , 2005, , .		87
32	Diamond in the rough. , 2004, , .		86
33	Algorithms for distributed functional monitoring. ACM Transactions on Algorithms, 2011, 7, 1-20.	0.9	70
34	Comparing data streams using hamming norms (how to zero in). IEEE Transactions on Knowledge and Data Engineering, 2003, 15, 529-540.	4.0	68
35	Differentially private summaries for sparse data. , 2012, , .		68
36	Private Release of Graph Statistics using Ladder Functions. , 2015, , .		68

#	Article	IF	Citations
37	Conquering the Divide: Continuous Clustering of Distributed Data Streams. , 2007, , .		65
38	Mergeable summaries. , 2012, , .		63
39	Finding hierarchical heavy hitters in streaming data. ACM Transactions on Knowledge Discovery From Data, 2008, 1, 1-48.	2.5	62
40	Forward Decay: A Practical Time Decay Model for Streaming Systems. Proceedings - International Conference on Data Engineering, 2009, , .	0.0	62
41	Marginal Release Under Local Differential Privacy. , 2018, , .		62
42	Data sketching. Communications of the ACM, 2017, 60, 48-55.	3.3	61
43	Personal privacy vs population privacy. , 2011, , .		59
44	The continuous distributed monitoring model. SIGMOD Record, 2013, 42, 5-14.	0.7	58
45	Set cover algorithms for very large datasets. , 2010, , .		56
46	Space- and time-efficient deterministic algorithms for biased quantiles over data streams. , 2006, , .		53
47	Approximate continuous querying over distributed streams. ACM Transactions on Database Systems, 2008, 33, 1-39.	1.5	53
48	Answering range queries under local differential privacy. Proceedings of the VLDB Endowment, 2019, 12, 1126-1138.	2.1	53
49	Optimal sampling from distributed streams. , 2010, , .		50
50	Anonymizing bipartite graph data using safe groupings. VLDB Journal, 2010, 19, 115-139.	2.7	46
51	Privacy in dynamic social networks. , 2010, , .		46
52	Continuous sampling from distributed streams. Journal of the ACM, 2012, 59, 1-25.	1.8	45
53	Exponentially Decayed Aggregates on Data Streams. , 2008, , .		39
54	Anonymized data. , 2009, , .		39

#	ARTICLE	IF	CITATIONS
55	Time-decaying aggregates in out-of-order streams. , 2008, , .		37
56	Robust lower bounds for communication and stream computation. , 2008, , .		36
57	Space-optimal heavy hitters with strong error bounds. ACM Transactions on Database Systems, 2010, 35, 1-28.	1.5	36
58	Comparing Data Streams Using Hamming Norms (How to Zero In). , 2002, , 335-345.		36
59	Quantiles over data streams. , 2013, , .		35
60	Approximating Data with the Count-Min Sketch. IEEE Software, 2012, 29, 64-69.	2.1	33
61	Quantiles over data streams: experimental comparisons, new analyses, and further improvements. VLDB Journal, 2016, 25, 449-472.	2.7	31
62	Fast Approximate Wavelet Tracking on Streams. Lecture Notes in Computer Science, 2006, , 4-22.	1.0	31
63	Estimating the confidence of conditional functional dependencies. , 2009, , .		30
64	Minimizing minimality and maximizing utility. Proceedings of the VLDB Endowment, 2010, 3, 1045-1056.	2.1	30
65	Sampling for big data. , 2014, , .		30
66	Time-decaying sketches for sensor data aggregation. , 2007, , .		29
67	Verifying computations with streaming interactive proofs. Proceedings of the VLDB Endowment, 2011, 5, 25-36.	2.1	29
68	Histograms and Wavelets on Probabilistic Data. Proceedings - International Conference on Data Engineering, 2009, , .	0.0	26
69	Semantics of Ranking Queries for Probabilistic Data. IEEE Transactions on Knowledge and Data Engineering, 2011, 23, 1903-1917.	4.0	26
70	Small synopses for group-by query verification on outsourced data streams. ACM Transactions on Database Systems, 2009, 34, 1-42.	1.5	25
71	Space-optimal heavy hitters with strong error bounds. , 2009, , .		24
72	Accurate and efficient private release of datacubes and contingency tables. , 2013, , .		23

#	Article	IF	Citations
73	A unifying framework for â„" 0-sampling algorithms. Distributed and Parallel Databases, 2014, 32, 315-335.	1.0	23
74	Conditional heavy hitters: detecting interesting correlations in data streams. VLDB Journal, 2015, 24, 395-414.	2.7	23
75	Annotations in Data Streams. Lecture Notes in Computer Science, 2009, , 222-234.	1.0	22
76	Empirical privacy and empirical utility of anonymized data. , 2013, , .		22
77	Streaming in a connected world. , 2007, , .		21
78	A near-optimal algorithm for estimating the entropy of a stream. ACM Transactions on Algorithms, 2010, 6, 1-21.	0.9	21
79	Continuous distributed monitoring. , 2011, , .		21
80	Lightweight authentication of linear algebraic queries on data streams. , 2013, , .		20
81	Frequency estimation under local differential privacy. Proceedings of the VLDB Endowment, 2021, 14, 2046-2058.	2.1	20
82	Probabilistic histograms for probabilistic data. Proceedings of the VLDB Endowment, 2009, 2, 526-537.	2.1	19
83	Histograms and Wavelets on Probabilistic Data. IEEE Transactions on Knowledge and Data Engineering, 2010, 22, 1142-1157.	4.0	19
84	Parameterized Streaming: Maximal Matching and Vertex Cover. , 2015, , .		19
85	On Estimating Frequency Moments of Data Streams. Lecture Notes in Computer Science, 2007, , 479-493.	1.0	19
86	A manifesto for modeling and measurement in social media. First Monday, 0, , .	0.6	19
87	Count-Min Sketch. , 2009, , 511-516.		18
88	Time-decaying Sketches for Robust Aggregation of Sensor Data. SIAM Journal on Computing, 2010, 39, 1309-1339.	0.8	16
89	Estimating Dominance Norms of Multiple Data Streams. Lecture Notes in Computer Science, 2003, , 148-160.	1.0	16
90	Tracking Distributed Aggregates over Time-Based Sliding Windows. Lecture Notes in Computer Science, 2012, , 416-430.	1.0	16

#	Article	IF	Citations
91	Permutation Editing and Matching via Embeddings. Lecture Notes in Computer Science, 2001, , 481-492.	1.0	14
92	Streaming Graph Computations with a Helpful Advisor. Algorithmica, 2013, 65, 409-442.	1.0	13
93	A second look at counting triangles in graph streams. Theoretical Computer Science, 2014, 552, 44-51.	0.5	11
94	A second look at counting triangles in graph streams (corrected). Theoretical Computer Science, 2017, 683, 22-30.	0.5	11
95	Information Cost Tradeoffs for Augmented Index and Streaming Language Recognition. , 2010, , .		10
96	Relative Error Streaming Quantiles. , 2021, , .		10
97	A Tight Lower Bound for Comparison-Based Quantile Summaries. , 2020, , .		10
98	Aggregate Query Answering on Possibilistic Data with Cardinality Constraints. , 2012, , .		9
99	A Dataset Search Engine for the Research Document Corpus. , 2012, , .		9
100	Socializing the h-index. Journal of Informetrics, 2013, 7, 718-721.	1.4	9
101	Annotations in Data Streams. ACM Transactions on Algorithms, 2014, 11, 1-30.	0.9	9
102	Don't let the negatives bring you down. Performance Evaluation Review, 2012, 40, 343-354.	0.4	8
103	Correlation Clustering in Data Streams. Algorithmica, 2021, 83, 1980-2017.	1.0	8
104	Privacy-Preserving Synthetic Location Data in the Real World. , 2021, , .		8
105	On Signatures for Communication Graphs. , 2008, , .		7
106	How NOT to review a paper. SIGMOD Record, 2009, 37, 100-104.	0.7	7
107	Streaming in a connected world. , 2008, , .		6
108	Time-decayed correlated aggregates over data streams. Statistical Analysis and Data Mining, 2009, 2, 294-310.	1.4	6

#	Article	IF	Citations
109	Tracking distributed aggregates over time-based sliding windows. , 2011, , .		6
110	Don't let the negatives bring you down. , 2012, , .		6
111	Information Cost Tradeoffs for Augmented Index and Streaming Language Recognition. SIAM Journal on Computing, 2013, 42, 61-83.	0.8	6
112	Data Sketching. Queue, 2017, 15, 49-67.	0.8	6
113	L_p Samplers and Their Applications. ACM Computing Surveys, 2020, 52, 1-31.	16.1	6
114	Streaming Graph Computations with a Helpful Advisor. Lecture Notes in Computer Science, 2010, , 231-242.	1.0	6
115	Structure-aware sampling. Proceedings of the VLDB Endowment, 2011, 4, 819-830.	2.1	6
116	Finding interesting correlations with conditional heavy hitters. , 2013, , .		5
117	Annotations for Sparse Data Streams. , 2014, , .		5
118	People like us. , 2014, , .		5
119	On the Tradeoff between Stability and Fit. ACM Transactions on Algorithms, 2016, 13, 1-24.	0.9	5
120	Anonymized Data: Generation, models, usage. , 2010, , .		4
121	What does an associate editor actually do?. SIGMOD Record, 2013, 42, 52-58.	0.7	4
122	Summarizing Two-Dimensional Data with Skyline-Based Statistical Descriptors. Lecture Notes in Computer Science, 2008, , 42-60.	1.0	4
123	Current Trends in Data Summaries. SIGMOD Record, 2022, 50, 6-15.	0.7	4
124	Time-Decayed Correlated Aggregates over Data Streams. , 2009, , .		3
125	Scienceography: The Study of How Science Is Written. Lecture Notes in Computer Science, 2012, , 379-391.	1.0	3
126	Summary Data Structures for Massive Data. Lecture Notes in Computer Science, 2013, , 78-86.	1.0	3

#	Article	IF	Citations
127	An evaluation of multi-probe locality sensitive hashing for computing similarities over web-scale query logs. PLoS ONE, 2018, 13, e0191175.	1.1	3
128	Semantics of Ranking Queries for Probabilistic Data. IEEE Transactions on Knowledge and Data Engineering, 2010, , .	4.0	2
129	Join Sizes, Frequency Moments, and Applications. Data-centric Systems and Applications, 2016, , 87-102.	0.2	2
130	Streaming Algorithms for Bin Packing and Vector Scheduling. Theory of Computing Systems, 2021, 65, 916.	0.7	2
131	Theory meets Practice at the Median: A Worst Case Comparison of Relative Error Quantile Algorithms. , 2021, , .		2
132	Count-Min Sketch. , 2014, , 1-6.		2
133	Count-Min Sketch. , 2018, , 653-659.		2
134	UMicS. , 2013, , .		1
135	First author advantage. , 2013, , .		1
136	Modeling collaboration in academia. , 2014, , .		1
137	Lightweight Query Authentication on Streams. ACM Transactions on Database Systems, 2014, 39, 1-45.	1.5	1
138	Compact Summaries over Large Datasets. , 2015, , .		1
139	Data-Independent Space Partitionings for Summaries. , 2021, , .		1
140	Misra-Gries Summaries. , 2016, , 1334-1337.		1
141	Streaming Algorithms for Bin Packing and Vector Scheduling. Lecture Notes in Computer Science, 2020, , 72-88.	1.0	1
142	Count-Min Sketch. , 2016, , 464-468.		1
143	Stable Distributions in Streaming Computations. Data-centric Systems and Applications, 2016, , 283-300.	0.2	1
144	Studying the source code of scientific research. SIGKDD Explorations: Newsletter of the Special Interest Group (SIG) on Knowledge Discovery & Data Mining, 2013, 14, 59-62.	3.2	0

#	Article	IF	Citations
145	On Unifying the Space of â„"₀-Sampling Algorithms. , 2013, , 163-172.		0
146	Corrigendum to â€œA second look at counting triangles in graph streamsâ€•[Theoret. Comput. Sci. 552 (2014) 44â€"51]. Theoretical Computer Science, 2017, 683, 31-32.	0.5	0
147	Verifiable Stream Computation and Arthur--Merlin Communication. SIAM Journal on Computing, 2019, 48, 1265-1299.	0.8	0
148	Selected data mining concepts. DIMACS Series in Discrete Mathematics and Theoretical Computer Science, 2007, , 1-40.	0.0	0
149	Misra-Gries Summaries. , 2014, , 1-5.		0
150	AMS Sketch. , 2016, , 76-78.		0
151	Count-Min Sketch. , 2017, , 1-6.		0
152	Relative Error Streaming Quantiles. SIGMOD Record, 2022, 51, 69-76.	0.7	0