

Professor Graham Cormode

Address: Department of Computer Science
University of Warwick
Coventry, UK CV4 7AL

Email: graham@cormode.org
Phone: +44 2476 523987
Web: <http://cormode.org>

Research Interests

- Algorithms for management and analysis of massive data, with applications to distributed and streaming data.
- Data anonymization and verifiable computation techniques for large data.
- Uncertain data representation, management and analysis.

Employment and Education

2013–present: Professor of Computer Science, University of Warwick, Coventry UK

Warwick University Liaison Director of the Alan Turing Institute.
NYU exchange professor with Center for Urban Science and Progress.

2006–2013: Principal Member of Technical Staff, AT&T Labs–Research, Florham Park, NJ USA.

Research scientist at AT&T Labs with focus on aspects of large-scale data management and analysis: streaming and distributed algorithms; privacy and anonymization; and representing and managing uncertain data.

2004–2006: Member of Technical Staff, Bell Labs, Lucent Technologies, Murray Hill, NJ USA.

Research scientist in Internet Management department at Bell Labs into next generation communications and data analysis. Worked on distributed communication and computation protocols, compressed sensing, large-scale data mining and network monitoring.

2002–2004: DIMACS Postdoctoral Fellow. Center for Discrete Mathematics and Computer Science (DIMACS), Piscataway, NJ USA.

Postdoctoral researcher on projects including massive data analysis, data stream algorithms, and data mining and epidemiology.

1998–2002: PhD. University of Warwick, UK. Thesis “Sequence Distance Embeddings”.

1999-2000 visiting scholar at Case Western Reserve University, Cleveland, Ohio.
2000 Summer Internship at AT&T Labs–Research.

1995–98: BA (Hons) Computer Science. King’s College, University of Cambridge, UK.

Awarded first class with honours. Won 1997 Addison-Wesley prize for best overall performance.

Honors and Awards

2014 Imre Simon Award, for most influential paper published in LATIN conference

2013 ACM Distinguished Scientist. Fewer than 300 of the 100,000 ACM members have been given this recognition.

2009 ICDE Best Paper Award for “Histograms and wavelets on probabilistic data”.

2008 VLDB Best Paper Award for “Finding frequent items in data streams”.

Grants and Funding Activities

- 2015–20 | PI, European Research Council Grant, Small Summaries for Big Data, €1.5M.
- 2015–20 | co-PI, EPSRC/Jaguar Land Rover, TASCC - The Cooperative Car, £2M.
- 2014–22 | co-PI, EPSRC Centre for Doctor Training in Urban Science and Progress, £4M.
- 2014–18 | PI, Microsoft EMEA Scholarship in Sketching Algorithms for Massive Graphs and Matrices, £79K
- 2013–14 | PI, European Commission Marie Curie Career Integration Grant, Doing Anonymization Practically, Privately, Effectively and Reusably, €100,000.
- 2013–15 | PI, Yahoo! Faculty Research Engagement Program. Smart Sketches for Modelling Massive Web Data, \$20K.
- 2011–15 | DIMACS Special Focus on Information Sharing and Dynamic Data Analysis. NSF award 1144502, total award \$0.7M.
- 2009–15 | AT&T PI, CCICADA — Command Control and Interoperability Center for Advanced Data Analysis (DHS Center of Excellence). Total award \$15M over 6 years.
- 2006–09 | AT&T co-PI, DyDAn — Center for Dynamic Data Analysis (DHS Center of Excellence). Total award \$5M over 3 years.
- 2006–07 | co-PI, KDD Research Grant for Blog Data Analysis. Total award \$200K.
- 2003–04 | NSF Joint Postdoc in Computational Epidemiology.

Books

- 2012 | G. Cormode, M. Garofalakis, P. Haas, and C. Jermaine. *Synopses for Massive Data: Samples, Histograms, Wavelets and Sketches*. now publishers, 2012
- 2010 | G. Cormode and M. Thottan, editors. *Algorithms for Next Generation Networks*. Springer, 2010
- 2006 | J. Abello and G. Cormode, editors. *Discrete Methods in Epidemiology*, volume 70 of DIMACS. AMS, 2006

Journal Publications

- 2016 | G. Luo, L. Wang, K. Yi, and G. Cormode. Quantiles over data streams: experimental comparisons, new analyses, and further improvements. *The VLDB Journal*, 25(4):449–472, 2016
- 2015 | K. Mirylenka, G. Cormode, T. Palpanas, and D. Srivastava. Conditional heavy hitters: detecting interesting correlations in data streams. *The VLDB Journal*, 24(3):395–414, 2015
S. Papadopoulos, G. Cormode, A. Deligiannakis, and M. N. Garofalakis. Lightweight query authentication on streams. *ACM Transactions on Database Systems*, 39(4):30:1–30:45, 2015
- 2014 | G. Cormode and H. Jowhari. A second look at counting triangles in graph streams. *Theoretical Computer Science*, 552:44–51, 2014
G. Cormode and D. Firmani. A unifying framework for l_0 -sampling algorithms. *Distributed and Parallel Databases*, 32(3):315–335, 2014. Special issue on Data Summarization on Big Data
A. Chakrabarti, G. Cormode, A. McGregor, and J. Thaler. Annotations in data streams. *ACM Transactions on Algorithms*, 11(1), 2014
- 2013 | P. K. Agarwal, G. Cormode, Z. Huang, J. M. Phillips, Z. Wei, and K. Yi. Mergeable summaries. *ACM Transactions on Database Systems*, 38(4):26, 2013
G. Cormode. What does an associate editor actually do? *SIGMOD Record*, 42(2):52–58, June 2013
G. Cormode. The continuous distributed monitoring model. *SIGMOD Record*, 42(1), Mar. 2013

- A. Chakrabarti, G. Cormode, R. Kondapally, and A. McGregor. Information cost tradeoffs for augmented index and streaming language recognition. *SIAM Journal on Computing (SICOMP)*, 42(1):61–83, 2013
- G. Cormode and D. Firmani. A unifying framework for l_0 -sampling algorithms. *Distributed and Parallel Databases*, 32(3):315–335, 2014. Special issue on Data Summarization on Big Data
- G. Cormode, Q. Ma, S. Muthukrishnan, and B. Thompson. Socializing the h-index. *Journal of Informetrics*, 7(3):718 – 721, 2013
- G. Cormode, M. Mitzenmacher, and J. Thaler. Streaming graph computations with a helpful advisor. *Algorithmica*, 65(2):409–442, 2013
- 2012 G. Cormode, S. Muthukrishnan, K. Yi, and Q. Zhang. Continuous sampling from distributed streams. *Journal of the ACM (JACM)*, 59(2), Apr. 2012 (Invited as one of the best papers from PODS 2010)
- G. Cormode, S. Muthukrishnan, and J. Yan. Studying the source code of scientific research. *SIGKDD Explorations*, 14(2):59–62, Dec. 2012
- G. Cormode and S. Muthukrishnan. Approximating data with the count-min data structure. *IEEE Software*, 2012
- 2011 G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed functional monitoring. *ACM Transactions on Algorithms*, 7(2):1–21, 2011
- G. Cormode, J. Jests, F. Li, and K. Yi. Semantics of ranking queries for probabilistic data. *IEEE Transactions on Knowledge and Data Engineering*, 23(12):1903–1917, 2011 (Invited as one of the best papers from ICDE 2010)
- 2010 G. Cormode, B. Krishnamurthy, and W. Willinger. A manifesto for modeling and measurement in social media. *First Monday*, 15(9), Sept. 2010
- G. Cormode and M. Garofalakis. Histograms and wavelets on probabilistic data. *IEEE Transactions on Knowledge and Data Engineering*, 22(8):1142–1157, Aug. 2010 (Invited as one of the best papers from ICDE 2009)
- A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. *ACM Transactions on Algorithms*, 6(3), 2010
- R. Berinde, G. Cormode, P. Indyk, and M. Strauss. Space-optimal heavy hitters with strong error bounds. *ACM Transactions on Database Systems*, 35(4), 2010 (Invited as one of the best papers from PODS 2009)
- G. Cormode, D. Srivastava, T. Yu, and Q. Zhang. Anonymizing bipartite graph data using safe groupings. *The VLDB Journal*, 19(1):115–139, 2010 (Invited as one of the best papers from VLDB 2009)
- G. Cormode and M. Hadjieleftheriou. Methods for finding frequent items in data streams. *The VLDB Journal*, 19(1):3–20, 2010 (Invited as one of the best papers from VLDB 2008)
- 2009 K. Yi, F. Li, G. Cormode, M. Hadjieleftheriou, G. Kollios, and D. Srivastava. Small synopses for group-by query verification on outsourced data streams. *ACM Transactions on Database Systems*, 34(3), 2009
- G. Cormode and M. Hadjieleftheriou. Finding the frequent items in streams of data. *Communications of the ACM (CACM)*, 52(10):97–105, 2009 (Invited from VLDB 2008)
- G. Cormode, S. Tirthapura, and B. Xu. Time-decaying sketches for robust aggregation of sensor data. *SIAM Journal on Computing (SICOMP)*, 39(4):1309–1339, 2009
- G. Cormode, S. Tirthapura, and B. Xu. Time-decayed correlated aggregates over data streams. *Statistical Analysis and Data Mining*, 2(5-6):294–310, 2009 (Invited as one of the best papers from SDM 2009)
- 2008 G. Cormode and M. Garofalakis. Approximate continuous querying over distributed streams. *ACM Transactions on Database Systems*, 33(2), June 2008
- G. Cormode and B. Krishnamurthy. Key differences between web 1.0 and web 2.0. *First Monday*, 13(6), June 2008
- G. Cormode. How not to review a paper: The tools and techniques of the adversarial reviewer. *SIGMOD Record*, 37(4):100–104, Dec. 2008
- G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava. Finding hierarchical heavy hitters in streaming data. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(4), Jan. 2008

- 2007 G. Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. *ACM Transactions on Algorithms*, 3(1), 2007 (Invited as one of the best papers of SODA 2002)
- 2005 G. Cormode and S. Muthukrishnan. What's new: Finding significant differences in network data streams. *Transactions on Networking*, 13(6):1219–1232, December 2005 (Invited as one of the best papers of Infocom 2004)
- G. Cormode and S. Muthukrishnan. An improved data stream summary: The count-min sketch and its applications. *Journal of Algorithms*, 55(1):58–75, April 2005
- G. Cormode and S. Muthukrishnan. What's hot and what's not: Tracking most frequent items dynamically. *ACM Transactions on Database Systems*, 30(1):249–278, March 2005 (Invited as one of the best papers of PODS 2003)
- 2004 G. Ozsoyoglu, N. H. Balkir, G. Cormode, and Z. M. Ozsoyoglu. Electronic books in digital libraries. *IEEE Transactions on Knowledge and Data Engineering*, 16(3):317–331, 2004
- G. Cormode. Representations of the research student in popular culture. *Annals of Improbable Research*, 10(1):26–27, 2004
- 2003 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 (Invited as one of the best papers of VLDB 2002)

Peer-reviewed Conference Publications

- 2016 Z. Jorgensen, T. Yu, and G. Cormode. Publishing attributed social graphs with formal privacy guarantees. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 107–122, 2016
- R. Chitnis, G. Cormode, H. Esfandiari, M. Hajiaghayi, A. McGregor, M. Monemizadeh, and S. Vorotnikova. Kernelization via sampling with applications to dynamic graph streams. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2016
- 2015 K. J. Ahn, G. Cormode, S. Guha, A. McGregor, and A. Wirth. Correlation clustering in data streams. In *International Conference on Machine Learning, ICML*, pages 2237–2246, 2015
- X. He, G. Cormode, A. Machanavajjhala, C. M. Procopiuc, and D. Srivastava. DPT: differentially private trajectory synthesis using hierarchical reference systems. In *Proceedings of the VLDB Endowment*, volume 8, pages 1154–1165, 2015
- R. Chitnis, G. Cormode, H. Esfandiari, M. Hajiaghayi, and M. Monemizadeh. New streaming algorithms for parameterized maximal matching & beyond. In *Symposium on Parallelism in Algorithms*, pages 56–58, 2015
- J. Zhang, G. Cormode, M. Procopiuc, D. Srivastava, and X. Xiao. Private release of graph statistics using ladder functions. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2015
- A. Chakrabarti, G. Cormode, A. McGregor, J. Thaler, and S. Venkatasubramanian. Verifiable stream computation and arthur-merlin communication. In *Computational Complexity Conference*, 2015
- Z. Jorgensen, T. Yu, and G. Cormode. Conservative or liberal? personalized differential privacy. In *International Conference on Data Engineering (ICDE)*, 2015
- R. Chitnis, G. Cormode, M. Hajiaghayi, and M. Monemizadeh. Parameterized streaming: Maximal matching and vertex cover. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2015
- 2014 G. Cormode, M. Procopiuc, D. Srivastava, X. Xiao, and J. Zhang. Privbayes: Private data release via bayesian networks. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2014
- A. Chakrabarti, G. Cormode, N. Goyal, and J. Thaler. Annotations for sparse data streams. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2014
- Q. Ma, S. Muthukrishnan, B. Thompson, and G. Cormode. Modeling collaboration in academia: A game theoretic approach. In *WWW Workshop on Big Scholarly Data*, 2014
- G. Cormode, S. Muthukrishnan, and J. Yan. People like us: Mining scholarly data for comparable researchers. In *WWW Workshop on Big Scholarly Data*, 2014

- 2013 G. Cormode, X. Gong, C. M. Procopiuc, E. Shen, D. Srivastava, and T. Yu. UMicS: From anonymized data to usable microdata. In *ACM Conference on Information and Knowledge Management (CIKM)*, 2013
- S. Papadopoulos, G. Cormode, A. Deligiannakis, and M. Garofalakis. Lightweight authentication of linear algebraic queries on data streams. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2013
- 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 (SIGMOD)*, 2013
- G. Cormode, S. Muthukrishnan, and J. Yan. First author advantage: Citation labeling in research. In *Proceedings of the Computational Scientometrics: Theory and Applications Workshop at CIKM*, 2013
- G. Cormode, C. M. Procopiuc, E. Shen, D. Srivastava, and T. Yu. Empirical privacy and empirical utility of anonymized data. In *Privacy-Preserving Data Publication and Analysis (PrivDB)*, 2013
- G. Cormode, K. Mirylenka, T. Palpanas, and D. Srivastava. Finding interesting correlations with conditional heavy hitters. In *International Conference on Data Engineering (ICDE)*, 2013
- G. Cormode, C. M. Procopiuc, D. Srivastava, and G. Yaroslavtsev. Accurate and efficient private release of datacubes and contingency tables. In *International Conference on Data Engineering (ICDE)*, 2013
- G. Cormode and D. Firmani. On unifying the space of l_0 -sampling algorithms. In *SIAM Meeting on Algorithm Engineering and Experiments*, 2013
- 2012 A. Goyal, H. Daumé III, and G. Cormode. Sketch algorithms for estimating point queries in NLP. In *EMNLP-CoNLL*, pages 1093–1103, 2012
- G. Cormode, S. Muthukrishnan, and J. Yan. Scienceography: the study of how science is written. In *Proceedings of the International Conference on Fun with Algorithms (FUN)*, 2012
- P. Agarwal, G. Cormode, Z. Huang, J. Phillips, Z. Wei, and K. Yi. Mergeable summaries. In *ACM Principles of Database Systems (PODS)*, 2012
- E. Cohen, G. Cormode, and N. Duffield. Don't let the negatives bring you down: Sampling from streams of signed updates. In *ACM Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2012
- G. Cormode and K. Yi. Tracking distributed aggregates over time-based sliding windows. In *Scientific and Statistical Database Management (SSDBM)*, 2012
- G. Cormode, M. Procopiuc, D. Srivastava, and T. Tran. Differentially private publication of sparse data. In *International Conference on Database Theory (ICDT)*, 2012
- G. Cormode, M. Procopiuc, E. Shen, D. Srivastava, and T. Yu. Differentially private spatial decompositions. In *International Conference on Data Engineering (ICDE)*, 2012
- M. Lu, S. Bangalore, G. Cormode, M. Hadjieleftheriou, and D. Srivastava. A dataset search engine for the research document corpus. In *International Conference on Data Engineering (ICDE)*, 2012
- G. Cormode, E. Shen, D. Srivastava, and T. Yu. Aggregate query answering on possibilistic data with cardinality constraints. In *International Conference on Data Engineering (ICDE)*, 2012
- G. Cormode, M. Mitzenmacher, and J. Thaler. Practical verified computation with streaming interactive proofs. In *Innovations in Theoretical Computer Science (ITCS)*, 2012
- 2011 G. Cormode. Personal privacy vs population privacy: Learning to attack anonymization. In *ACM SIGKDD Conference*, 2011
- P. Agarwal, G. Cormode, Z. Huang, J. Phillips, Z. Wei, and K. Yi. Mergeable coresets. In *Third Workshop on Massive Data Algorithmics (MASSIVE)*, 2011
- G. Cormode and K. Yi. Tracking distributed aggregates over time-based sliding windows (brief announcement). In *ACM Principles of Distributed Computing (PODC)*, 2011
- E. Cohen, G. Cormode, and N. Duffield. Structure-aware sampling on data streams. In *ACM Conference on Measurement and Modeling of Computer Systems (SIGMETRICS)*, 2011
- E. Cohen, G. Cormode, and N. Duffield. Structure-aware sampling: Flexible and accurate summarization. In *International Conference on Very Large Data Bases (VLDB)*, 2011
- 2010 G. Cormode, H. Karloff, and T. Wirth. Set cover algorithms for very large datasets. In *ACM Conference on Information and Knowledge Management (CIKM)*, 2010

- A. Chakrabarti, G. Cormode, R. Kondapally, and A. McGregor. Information cost tradeoffs for augmented index and streaming language recognition. In *IEEE Foundations of Computer Science (FOCS)*, 2010
- G. Cormode, M. Mitzenmacher, and J. Thaler. Streaming graph computations with a helpful advisor. In *European Symposium on Algorithms*, 2010
- G. Cormode, S. Muthukrishnan, K. Yi, and Q. Zhang. Optimal sampling from distributed streams. In *ACM Principles of Database Systems (PODS)*, 2010
- S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava. Privacy in dynamic social networks. In *World Wide Web Conference (WWW)*, 2010
- G. Cormode, N. Li, T. Li, and D. Srivastava. Minimizing minimality and maximizing utility: Analyzing method-based attacks on anonymized data. In *International Conference on Very Large Data Bases (VLDB)*, 2010
- S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava. Prediction promotes privacy in dynamic social networks. In *Workshop on Online Social Networks (WOSN)*, 2010
- 2009 S. Bhagat, G. Cormode, B. Krishnamurthy, and D. Srivastava. Class-based graph anonymization for social network data. In *International Conference on Very Large Data Bases (VLDB)*, 2009
- G. Cormode, A. Deligiannakis, M. Garofalakis, and A. McGregor. Probabilistic histograms for probabilistic data. In *International Conference on Very Large Data Bases (VLDB)*, 2009
- A. Chakrabarti, G. Cormode, and A. McGregor. Annotations in data streams. In *International Colloquium on Automata, Languages and Programming (ICALP)*, 2009
- G. Cormode, L. Golab, F. Korn, A. McGregor, D. Srivastava, and X. Zhang. Estimating the confidence of conditional functional dependencies. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2009
- R. Berinde, G. Cormode, P. Indyk, and M. Strauss. Space-optimal heavy hitters with strong error bounds. In *ACM Principles of Database Systems (PODS)*, 2009
- G. Cormode, S. Tirthapura, and B. Xu. Time-decayed correlated aggregates over data streams. In *SIAM Conference on Data Mining (SDM)*, 2009
- G. Cormode and M. Garofalakis. Histograms and wavelets on probabilistic data. In *International Conference on Data Engineering (ICDE)*, 2009. Best paper award
- G. Cormode, F. Li, and K. Yi. Semantics of ranking queries for probabilistic data and expected ranks. In *International Conference on Data Engineering (ICDE)*, 2009
- G. Cormode, V. Shkapenyuk, D. Srivastava, and B. Xu. Forward decay: A practical time decay model for streaming systems. In *International Conference on Data Engineering (ICDE)*, 2009
- 2008 G. Cormode and M. Hadjieleftheriou. Finding frequent items in data streams. In *International Conference on Very Large Data Bases (VLDB)*, 2008. Best paper award
- G. Cormode, D. Srivastava, T. Yu, and Q. Zhang. Anonymizing bipartite graph data using safe groupings. In *International Conference on Very Large Data Bases (VLDB)*, 2008
- G. Cormode, F. Korn, and S. Tirthapura. Time-decaying aggregates in out-of-order streams. In *ACM Principles of Database Systems (PODS)*, 2008
- G. Cormode and A. McGregor. Approximation algorithms for clustering uncertain data. In *ACM Principles of Database Systems (PODS)*, 2008
- G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava. Summarizing two-dimensional data with skyline-based statistical descriptors. In *Scientific and Statistical Database Management (SSDBM)*, 2008
- A. Chakrabarti, G. Cormode, and A. McGregor. Robust lower bounds for communication and stream computation. In *ACM Symposium on Theory of Computing (STOC)*, 2008
- G. Cormode, F. Korn, S. Muthukrishnan, and Y. Wu. On signatures for communication graphs. In *International Conference on Data Engineering (ICDE)*, 2008
- G. Cormode, F. Korn, and S. Tirthapura. Exponentially decayed aggregates on data streams. In *International Conference on Data Engineering (ICDE)*, 2008
- G. Cormode, S. Muthukrishnan, and K. Yi. Algorithms for distributed, functional monitoring. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2008
- 2007 S. Bhagat, G. Cormode, and I. Rozenbaum. Applying link-based classification to label blogs. In *Joint WEBKDD and SNA-KDD Workshop*, 2007

- S. Ganguly and G. Cormode. On estimating frequency moments of data streams. In *Proceedings of RANDOM*, 2007
- G. Cormode, S. Tirthapura, and B. Xu. Time-decaying sketches for sensor data aggregation. In *ACM Principles of Distributed Computing (PODC)*, 2007
- G. Cormode and M. Garofalakis. Sketching probabilistic data streams. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2007
- S. Bhagat, G. Cormode, S. Muthukrishnan, I. Rozenbaum, and H. Xue. No blog is an island analyzing connections across information networks. In *International Conference on Weblogs and Social Media*, 2007
- G. Cormode, S. Muthukrishnan, and W. Zhuang. Conquering the divide: Continuous clustering of distributed data streams. In *International Conference on Data Engineering (ICDE)*, 2007
- A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for computing the entropy of a stream. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, 2007
- 2006 G. Cormode and S. Muthukrishnan. Combinatorial algorithms for compressed sensing. In *SIROCCO*, 2006
- 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 (PODS)*, 2006
- G. Cormode, R. Keralapura, and J. Ramimirtham. Communication-efficient distributed monitoring of thresholded counts. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, 2006
- G. Cormode, M. Garofalakis, and D. Sacharidis. Fast approximate wavelet tracking on streams. In *Extending Database Technology*, pages 4–22, 2006
- G. Cormode, S. Muthukrishnan, and W. Zhuang. What’s different: Distributed, continuous monitoring of duplicate-resilient aggregates on data streams. In *International Conference on Data Engineering (ICDE)*, pages 20–31, 2006
- 2005 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 (VLDB)*, pages 25–36, 2005
- G. Cormode and M. Garofalakis. Sketching streams through the net: Distributed approximate query tracking. In *International Conference on Very Large Data Bases (VLDB)*, pages 13–24, 2005
- G. Cormode and S. Muthukrishnan. Space efficient mining of multigraph streams. In *ACM Principles of Database Systems (PODS)*, pages 271–282, 2005
- G. Cormode, M. Garofalakis, S. Muthukrishnan, and R. Rastogi. Holistic aggregates in a networked world: Distributed tracking of approximate quantiles. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 25–36, 2005
- G. Cormode and S. Muthukrishnan. Summarizing and mining skewed data streams. In *SIAM Conference on Data Mining (SDM)*, 2005
- G. Cormode and S. Muthukrishnan. Substring compression problems. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 321–330, 2005
- 2004 G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava. Diamond in the rough: Finding hierarchical heavy hitters in multi-dimensional data. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 155–166, 2004
- G. Cormode, F. Korn, S. Muthukrishnan, T. Johnson, O. Spatscheck, and D. Srivastava. Holistic UDAFs at streaming speeds. In *ACM SIGMOD International Conference on Management of Data (SIGMOD)*, pages 35–46, 2004
- G. Cormode. The hardness of the lemmings game, or Oh no, more NP-completeness proofs. In *Proceedings of Third International Conference on Fun with Algorithms*, pages 65–76, 2004
- G. Cormode, A. Czumaj, and S. Muthukrishnan. How to increase the acceptance ratios of top conferences. In *Proceedings of Third International Conference on Fun with Algorithms*, pages 262–273, 2004
- G. Cormode and S. Muthukrishnan. What’s new: Finding significant differences in network data streams. In *Proceedings of IEEE Infocom*, pages 1534–1545, 2004

- G. Cormode and S. Muthukrishnan. An improved data stream summary: The count-min sketch and its applications. In *Proceedings of Latin American Theoretical Informatics (LATIN)*, pages 29–38, 2004
- 2003 G. Cormode. Stable distributions for stream computations: it's as easy as 0,1,2. In *Workshop on Management and Processing of Massive Data Streams at FCRC*, 2003
- G. Cormode and S. Muthukrishnan. What's hot and what's not: Tracking most frequent items dynamically. In *ACM Principles of Database Systems (PODS)*, pages 296–306, 2003
- G. Cormode, F. Korn, S. Muthukrishnan, and D. Srivastava. Finding hierarchical heavy hitters in data streams. In *International Conference on Very Large Data Bases (VLDB)*, pages 464–475, 2003
- G. Cormode and S. Muthukrishnan. Estimating dominance norms of multiple data streams. In *European Symposium on Algorithms*, volume 2838 of *LNCS*, 2003
- 2002 G. Cormode, M. Datar, P. Indyk, and S. Muthukrishnan. Comparing data streams using Hamming norms. In *International Conference on Very Large Data Bases (VLDB)*, pages 335–345, 2002
- G. Cormode, P. Indyk, N. Koudas, and S. Muthukrishnan. Fast mining of tabular data via approximate distance computations. In *International Conference on Data Engineering (ICDE)*, pages 605–616, 2002
- G. Cormode and S. Muthukrishnan. The string edit distance matching problem with moves. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 667–676, 2002
- 2001 G. Cormode, S. Muthukrishnan, and S. C. Sahinalp. Permutation editing and matching via embeddings. In *International Colloquium on Automata, Languages and Programming (ICALP)*, volume 2076, pages 481–492, 2001
- 2000 G. Cormode, M. Paterson, S. C. Sahinalp, and U. Vishkin. Communication complexity of document exchange. In *ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 197–206, 2000
- G. Ozsoyoglu, N. H. Balkir, G. Cormode, and Z. M. Ozsoyoglu. Electronic books in digital libraries. In *Proceedings of IEEE Advances in Digital Libraries (ADL)*, pages 5–14, 2000

Teaching and Mentoring

- 2015 Advanced Databases, BSc course at Warwick University.
- 2014 Summer school course on Sketches, streaming and big data at University of Copenhagen.
- 2014 Mathematics for Computer Scientists, BSc course at Warwick University.
- 2013-6 Foundations of Data Analytics, MSc course at Warwick University.
- 2012 Summer School Course on Algorithms for Distributed Streams, MADALGO Center, Aarhus University, Denmark.
- 2011 Map Reduce Algorithms Graduate Class, Rutgers University.
Guest lecturer on Data Anonymization, Columbia University.
- 2009, 10 Tutorial on Anonymized Data at SIGMOD 2009, ICDE 2010
- 2008 Summer school on Data Stream Algorithms at Bristol University, UK.
- 2006, 07, 08 Tutorial on querying and tracking distributed streams. Presented at VLDB 2006, SIGMOD 2007, EDBT 2008.
- 2006 Blog Data Analysis — Graduate Seminar Class at Rutgers University.
- 2003–12 DIMACS Research Experience for Undergraduates Mentor.
- 2003 Processing Massive Data Sets — Graduate Class, Rutgers University.
- 2000–02 Seminar leader for Discrete Mathematics, University of Warwick.
- 2000 Teaching Assistant for Graduate Algorithms EECS454, Case Western Reserve University.

Student Supervision

PhD Students

Jacques Dark (2015-2018)
Tejas Kulkarni (2015-2018)
Charlie Dickens (2016-2019)
Chris Hickey (2016-2019)

Research Experience for Undergraduates (REU)

Daniel Stubbs (2012)
Roy Luo (2010)
Justin Thaler (2007)
Yinmeng Zhang (2004)
Shiri Azenkot (2004)
Adam Kirsch (2003)

Graduate Student Interns

Jun Zhang (2013)
Vishesh Karwa (2012)
Donatella Firmani (2012)
Katya Mirylenka (2011)
Grigory Yaroslavtsev (2011)
Thanh Tran (2010)
Meiyu Lu (2010)
Entong Shen (2010)
Smriti Bhagat (2009)
Zhifeng Bao (2009)
Tiancheng Li (2009)
Xi Zhang (2008)
Bojian Xu (2008)
Qing Zhang (2007)
Alessio Cavallini (2007)

Granted US Patents

9,251,216	Efficient publication of sparse data
9,116,958	Methods and apparatus to sample data connections
8,825,640	Methods and apparatus for ranking uncertain data in a probabilistic database
8,799,754	Verification of data stream computations using third-party-supplied annotations
8,732,295	Methods and apparatus for providing real friends count
8,631,500	Generating minimality-attack-resistant data
8,627,488	Methods and apparatus to anonymize a dataset of spatial data
8,612,649	Validation of priority queue processing
8,595,194	Forward decay temporal data analysis
8,590,049	Method and apparatus for providing anonymization of data
8,589,443	Method and apparatus for providing anonymization of data
8,538,938	Interactive proof to validate outsourced data stream processing
8,484,269	Computing time-decayed aggregates under smooth decay functions
8,458,326	Sampling from distributed streams of data
8,438,650	Anonymization of data over multiple temporal releases
8,391,164	Computing time-decayed aggregates in data streams
8,386,534	Automatic gleaning of semantic information in social networks
8,386,412	Methods and apparatus to construct histogram and wavelet synopses for probabilistic data
8,160,837	Methods and apparatus to determine statistical dominance point descriptors for multidimensional data

8,145,669	Methods and apparatus for representing probabilistic data using a probabilistic histogram
8,112,802	Verification of outsourced data streams
8,078,710	Method and apparatus for monitoring functions of distributed data
7,885,911	Fast approximate wavelet tracking on streams
7,783,647	Method and apparatus for globally approximating quantiles in a distributed monitoring environment
7,756,805	Method for distributed tracking of approximate join size and related summaries
7,742,424	Communication-efficient distributed monitoring of thresholded counts
7,657,503	System and method for generating statistical descriptors for a data stream
7,590,657	System and method for identifying hierarchical heavy hitters in a multidimensional environment
7,584,396	System and method for decoding a signal using compressed sensor measurements
7,450,032	System and method for encoding a signal using compressed sensor measurements

Selected Invited Talks and Tutorials

2016	Sub-quadratic recovery of correlated pairs, June 2016. Talk at Google Research, Facebook, Simons Institute
2015	Compact summaries over large datasets, May 2015. Invited tutorial in PODS 2015 and BICOD 2015 Trusting the cloud with practical interactive proofs, Apr. 2015. Talk at Google NYC, Bristol, Oxford Algorithms Day, Durham The confounding problem of private data release, Mar. 2015. Invited talk at EDBT/ICDT 2015
2014	Sampling for big data, Aug. 2014. Tutorial at SIGKDD 2014 conference Sketches, streaming and big data, July 2014. Summer school on Hashing at University of Copenhagen
2013	Summary data structures for massive data, July 2013. Invited talk in Session on Data Streams and Compression, Computability in Europe 2013 Computing + statistics = data science, June 2013. An introduction to data science for teenagers, IGGY DUX awards, Experience Warwick Streaming verification of outsourced computation, May 2013. Talk at Big Data Analytics Workshop, Microsoft Research Cambridge, and University of Warwick Privacy and big data: Challenges and promise, Mar. 2013. Invited panel at NYU Abu Dhabi conference on Big Data Systems, Applications, and Privacy Building blocks of privacy: Differentially private mechanisms, Apr. 2013. Invited tutorial talk at Privacy Preserving Data Publication and Analysis (PrivDB) workshop Current industry trends in computer science research, Mar. 2013. Invited Talk/Panel at NSF Research Experience for Undergraduates PI Meeting
2012	Data-driven concerns in private data release, Sept. 2012. Talk at Stevens Institute of Technology; AT&T Labs; UMass Amherst; Rutgers University-Newark; Bell Labs; NYU-Abu Dhabi Sketches: Past, present and future, 2012. Invited Panel on Sketching and Streaming at SAMSI Workshop, 2012 Small summaries for Big Data, 2012. Talk at Duke ARO workshop on Big Data at Large; MSR Cambridge; Princeton
2011	Some sketchy results, May 2011. Talk at DIMACS Workshop on Algorithms in the Field (8F) Mergeable summaries, Apr. 2011. Talk at Harvard University; DIMACS; Johns Hopkins; University of Pennsylvania; AT&T Labs; Warwick University Continuous distributed monitoring: A short survey, Sept. 2011. Invited keynote at Algorithms and Models for Distributed Event Processing (AlMoDEP) Distributed summaries, 2011. Talk at DIMACS workshop on Parallelism: a 2020 vision
2010	G. Cormode and D. Srivastava. Anonymized data: Generation, models, usage, Mar. 2010. Tutorial at ICDE 2010

- Sipping from the firehose: Streaming interactive proofs for verifying computations, February 2010. Invited talk at Bristol Algorithms Days 2010; University of Maryland
 Progress in data anonymization: from k-anonymity to the minimality attack, February 2010. Talk in Bristol
- 2009 Finding frequent items in data streams, March 2009. Talk at DIMACS Working group on Streaming, Coding and Compressive Sensing; AT&T Labs; UMass Amherst; Dartmouth College
 Anonymization and uncertainty in social network data, Oct. 2009. Invited talk at DBIR Day 2009 at NYU Poly
 G. Cormode and D. Srivastava. Anonymized data: Generation, models, usage, July 2009. Tutorial at SIGMOD 2009
 Processing graph streams: Upper and lower bounds, June 2009. Talk at Workshop on Algorithms and Models for Complex Networks, Bristol UK
- 2008 On 'selection and sorting with limited storage', Sept. 2008. Talk at Mike66 Workshop celebrating Mike Paterson
 Algorithms for distributed functional monitoring, Aug. 2008. Talk at Dagstuhl Seminar on Sublinear Algorithms
 Data stream algorithms, July 2008. Tutorial at Bristol Summer School on Probabilistic Techniques in Computer Science
 G. Cormode and M. Garofalakis. Streaming in a connected world: Querying and tracking distributed data streams, March 2008. Tutorial at VLDB 2006, SIGMOD 2007, EDBT 2008
- 2007 Analyzing web 2.0, blogs and social networks, Dec. 2007. Talk at AT&T Labs
 Computational fundamentals of analyzing and mining data streams, March 2007. Tutorial at Workshop on Data Stream Analysis, Caserta, Italy
- 2006 Computing the entropy of a stream, December 2006. AT&T Labs; Bell Labs; DyDAn Center
 A compact survey of compressed sensing, December 2006. Workshop on Algorithms for Data Streams, IIT Kanpur, India
 Biased quantiles, June 2006. Bertinoro
 Cluster and data stream analysis, March 2006. Tutorial at DIMACS Workshop on Data Mining and Epidemiology
- 2005 Tracking inverse distributions of massive data streams, July 2005. Network Sampling Workshop in Paris, Bell Labs Research Seminar
 Towards an algorithmic theory of compressed sensing, July 2005. Schloss Dagstuhl
 Summarizing and mining skewed data streams, May 2005. NJIT
- 2004 Algorithms for processing massive data at network line speed, March 2004. Talk at U. Iowa; U. Minnesota; Dartmouth; Google; AT&T; CWRU; Poly
 How hard are computer games?, February 2004. Talk at DIMACS
- 2003 What's hot, what's not, what's new and what's next, October 2003. Bell Labs; DIMACS Mixer at AT&T Labs
 Zeroing in on the l0 metric, August 2003. DIMACS Workshop on Discrete Metric Spaces and their Algorithmic Applications at Princeton
 Tracking frequent items dynamically, 2003. Institute of Advanced Studies; DIMACS; Stonybrook; U. Pennsylvania
- 2002 Algorithmic embeddings for comparing large text streams, June 2002. CCR/DIMACS Workshop/Tutorial on Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense
 Embeddings of metrics on strings and permutations, March 2002. Workshop on Discrete Metric Spaces and their Algorithmic Applications in Haifa, Israel; BCTCS
- 2000 Short string signatures, September 2000. DIMACS Workshop on Sublinear Algorithms in Princeton, NJ

Other Professional Activities

- University Liaison Director (Warwick), Alan Turing Institute (2016–present).
- Associate Editor, ACM Transactions on Database Systems (TODS), 2012–present.
- Associate Editor, Journal of Discrete Algorithms, 2014–present.
- Associate Editor, IEEE Transactions on Knowledge and Data Engineering (TKDE), 2009–2014.
- Finance Chair, ACM SIGMOD Conference 2013.
- Special Focus Leader, DIMACS Special Focus on Information Sharing and Dynamic Data Analysis, 2011–2015.
- AT&T PI, CCICADA: Command, Control and Interoperability Center for Advanced Data Analysis, 2009–15.
- Expert Evaluator, European Commission FP7, 2012
- National Science Foundation Panel Member, 2007, 2010.
- Member DIMACS Council (2008–present), DIMACS Graduate student selection cttee (2005–2010)
- Workshop and Meeting Organizer:
 - Cryptocurrency Day, Alan Turing Institute, 2016
 - Alan Turing Institute Scoping Workshops: Anomaly and Change Detection in Streaming Big Data; Statistical and computational challenges in large-scale data analysis; Computing Systems Research For Big Data; Information-Theoretic Foundations and Algorithms for Large-Scale Data Inference, 2015.
 - Big Data and Computational Scalability, Warwick, 2015.
 - First International Workshop on Big Dynamic Distributed Data, VLDB, 2013
 - Big Data in the Mathematical Sciences, Warwick, 2013
 - Shonan Meeting on Large-scale Distributed Computation, 2012
 - Second International Workshop on Management of Uncertain Data (MOUND), 2010
 - North-east DBIR Day, 2010
 - DIMACS Mixer, 2009
 - DIMACS Tutorial on Algorithms for Next Generation Networks, 2007
 - DIMACS Tutorial on Data Mining, 2006.
 - DIMACS Working Group on Data Mining and Epidemiology, 2004.
 - DIMACS Seminar Series on Quantitative Biology, 2003–04
- Conference Program Committee Member for STOC 2013; PODS 2016, 2012, 2010, 2008; VLDB 2014 (senior PC), 2012, 2011, 2010, 2009, 2008, 2007, 2005; SIGMOD 2017, 2013, 2010 (group leader); ICALP 2016; ICDE 2015, 2013, 2012, 2009, 2004; ICDT 2016, 2014, 2011; CIKM track chair 2014, 2008; EDBT 2015; PETS 2016, 2015, 2014; KDD 2016 (senior PC), 2011, 2004; ICDM 2006, 2005, 2004.