

Graham Cormode

Address: AT&T Labs–Research, Rm E275,
180 Park Ave
Florham Park, NJ, 07932
United States of America

Telephone: (973) 360 8957

Date of Birth: 27 June 1977

Nationality: British

US Visa Status: H-1B

Email: graham @ research.att.com

Web page: <http://cormode.org>

Research Interests

- Algorithms for mining massive data sets, with applications to network and database streams.
- Embedding methods for dimensionality reduction and approximation algorithms.
- Distributed data management and continuous computations.

Employment and Education

2006 — Present: Technical Specialist, AT&T Labs–Research, Florham Park, NJ

2004 — 2006: Member of Technical Staff, Internet Management, Bell Labs, Lucent Technologies, Murray Hill, NJ

Researcher 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.

Postdoctoral researcher on projects including Massive Data Analysis, Data Stream Algorithms, and Data Mining and Epidemiology, under supervision of Professor S. Muthukrishnan.

1998 — 2002: PhD. University of Warwick, Coventry, United Kingdom with Mike Paterson

Including 1999-2000 at Case Western Reserve University, Cleveland, Ohio with Cenk Sahinalp and Summer 2000 Summer Manager at AT&T Research Shannon Labs with S. Muthukrishnan.

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

Awarded first class with honours. Placed in top five students in each year. Awarded the Addison-Wesley prize for best overall performance in the second year.

Journal Publications

1. “What’s New: Finding Significant Differences in Network Data Streams”, with S. Muthukrishnan, in *ACM Transactions on Networking* 13(6):1219–1232, 2006.
2. “What’s Hot and What’s Not: Tracking Most Frequent Items Dynamically”, with S. Muthukrishnan, invited submission in *ACM Transactions on Database Systems (TODS)* 30(1):249–278, 2005.
3. “An Improved Data Stream Summary: The Count-Min Sketch and its Applications”, with S. Muthukrishnan, in *Journal of Algorithms* 55(1):58–75.
4. “The String Edit Distance Matching Problem with Moves” with S. Muthukrishnan, invited submission to appear in *Journal of Algorithms*.

5. "Representations of the Research Student in Popular Culture", in *Annals of Improbable Research (AIR)* 10(1):26-27, 2004.
6. "Electronic Books in Digital Libraries", with G. Ozsoyoglu, H. Balkir and M. Ozsoyoglu, in *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 16(3):317-331, 2004.
7. "Comparing Data Streams Using Hamming Norms" with M. Datar, P. Indyk, S. Muthukrishnan, invited submission in *IEEE Transactions on Knowledge and Data Engineering (TKDE)* 15(3):529-541, 2003.

Conference Publications

1. "Combinatorial Algorithms for Compressed Sensing" with S. Muthukrishnan. In *Colloquium on Structural Information and Communication Complexity (SIROCCO)*, 2006.
2. "Space- and Time- Efficient Deterministic Algorithms for Biased Quantiles over Data Streams", with Flip Korn S. Muthukrishnan, and Divesh Srivastava. In *Proceedings of ACM Principles of Database Systems (PODS)*, 2006.
3. "Communication-efficient distributed monitoring of thresholded counts" with Ram Keralapura and Jai Ramimirtham. In *Proceedings of ACM SIGMOD*, 2006.
4. "Fast Approximate Wavelet Tracking on Streams" with Dimitris Sacharidis and Minos Garofalakis. In *Proceedings of Extending Database Technology (EDBT)*, 2006.
5. "What's Different: Distributed, Continuous Monitoring of Duplicate Resilient Aggregates on Data Streams", with S. Muthukrishnan and Wei Zhuang. In *Proceedings of the International Conference on Data Engineering*, 2006.
6. "Summarizing and mining inverse distributions on data streams via dynamic inverse sampling" with S. Muthukrishnan and Irina Rozenbaum. In *Proceedings of the International Conference on Very Large Data Bases*, 2005.
7. "Sketching streams through the net: Distributed approximate query tracking" with Minos Garofalakis. In *Proceedings of the International Conference on Very Large Data Bases*, 2005.
8. G. Cormode and S. Muthukrishnan. "Space efficient mining of multigraph streams" with S. Muthukrishnan. In *Proceedings of ACM Principles of Database Systems*, 2005.
9. "Holistic aggregates in a networked world: Distributed tracking of approximate quantiles" with Minos Garofalakis, S. Muthukrishnan, and Rajeev Rastogi. In *Proceedings of ACM SIGMOD International Conference on Management of Data*, pages 25-36, 2005.
10. "Summarizing and mining skewed data streams" with S. Muthukrishnan. In *SIAM Conference on Data Mining*, 2005.
11. "Effective computation of biased quantiles over data streams" with Flip Korn, S. Muthukrishnan, and Divesh Srivastava. In *21st International Conference on Data Engineering (ICDE)*, pages 20-31, 2005.
12. "Substring compression problems" with S. Muthukrishnan. In *ACM-SIAM Symposium on Discrete Algorithms*, pages 321-330, 2005.
13. "What's New: Finding Significant Differences in Network Data Streams", with S. Muthukrishnan in *IEEE INFOCOM*, pages 1534-1545, 2004.

14. "An Improved Data Stream Summary: The Count-Min Sketch and its Applications", with S. Muthukrishnan, in *Latin American Theoretical Informatics (LATIN)*, pages 29–38, 2004.
15. "Diamond in the Rough: Finding Hierarchical Heavy Hitters in Multi-Dimensional Data", with Flip Korn, S. Muthukrishnan, Divesh Srivastava in *ACM Special Interest Group on Management of Data (SIGMOD)*, pages 155-166, 2004.
16. "Holistic UDAFs at streaming speeds" with Flip Korn, Ted Johnson, S. Muthukrishnan, Oliver Spatscheck, Divesh Srivastava in *ACM Special Interest Group on Management of Data (SIGMOD)*, pages 35–46, 2004.
17. "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.
18. "How to increase the acceptance ratios of top conferences" with A. Czumaj, S. Muthukrishnan in *Proceedings of Third International Conference on Fun with Algorithms*, pages 262–272, 2004.
19. "What's Hot and What's Not: Tracking Most Frequent Items Dynamically", with S. Muthukrishnan, in *ACM Principles of Database Systems (PODS)*, pages 296–306, 2003
20. "Finding Hierarchical Heavy Hitters in Data Streams", with F. Korn, S. Muthukrishnan, D. Srivastava, in *International Conference on Very Large Data Bases (VLDB)*, pages 464–475, 2003.
21. "Stable distributions for stream computations: it's as easy as 0,1,2", in *Workshop on Management and Processing of Data Streams (MPDS)*, pages 37–38, 2003.
22. "Estimating Dominance Norms of Multiple Data Streams" with S. Muthukrishnan, in *European Symposium on Algorithms (ESA)*, pages 148–160, 2003.
23. "Comparing Data Streams Using Hamming Norms" with M. Datar, P. Indyk, S. Muthukrishnan, in *International Conference on Very Large Data Bases (VLDB)*, pages 335–545, 2002.
24. "The String Edit Distance Matching Problem with Moves" with S. Muthukrishnan, in *Symposium on Discrete Algorithms (SODA)*, pages 667–676, 2002.
25. "Fast Mining of Tabular Data via Approximate Distance Computations" with P. Indyk, N. Koudas, S. Muthukrishnan, in *International Conference on Data Engineering (ICDE)*, pages 605–616, 2002
26. "Permutation Editing and Matching via Embeddings" with S. Muthukrishnan, C. Sahinalp, in *International Colloquium on Automata, Languages and Programming (ICALP)*, pages 481–492, 2001
27. "Communication Complexity of Document Exchange", with M. Paterson, C. Sahinalp and U. Vishkin, in *Symposium on Discrete Algorithms (SODA)*, pages 197–206, 2000
28. "Electronic Books in Digital Libraries", with G. Ozsoyoglu, H. Balkir and M. Ozsoyoglu, in *IEEE Advances in Digital Libraries (ADL)*, pages 5–14, 2000.

Teaching and Mentoring

Spring 2006 Blog Data Analysis – Graduate Seminar Class at Rutgers University.

Summer 2003 DIMACS Research Experience for Undergraduates Mentor.

Summer 2004 Worked closely with six undergraduate students in research projects.

Summer 2006

Spring 2003 Taught Processing Massive Data Sets 198:671, at Rutgers University.
Gave four 2 hour lectures on fundamental techniques for massive data analysis.
Guided and moderated student presentations.

2000 — 2002 Seminar leader for Discrete Mathematics CS127, University of Warwick.
Recitations with 30-40 Undergraduate Students, totalling over 50 hours classroom time.

Spring 2000 Teaching Assistant for Graduate Algorithms EECS454, Case Western Reserve University.
Grading and weekly office hours for 40 students.

Patents (Granted and Applied for)

- 20050131946 “A method for finding hierarchical heavy hitters in data streams” with F. Korn, D. Srivastava and S. Muthukrishnan. Application filed April 2003.
- Lucent/Cormode 1-23-19-65 “Method And Apparatus For Globally Approximating Quantiles In A Distributed Monitoring Environment” with M. Garofalakis, S. Muthukrishnan and R. Rastogi.
- Lucent/Cormode 2-27 “A Method For Distributed Tracking of Approximate Join Size and Related Summaries” with M. Garofalakis.
- Lucent/Cormode 3-28 “Fast Approximate Wavelet Tracking On Streams” with M. Garofalakis.
- Provisional Patent 60/560666 “System and Method for Identifying Hierarchical Heavy Hitters in a Multidimensional Environment” with F. Korn, S. Muthukrishnan and D. Srivastava.
- Provisional Patent 11/293655 “Method and Apparatus for Finding Biased Quantiles in Data Streams” with F. Korn, S. Muthukrishnan and D. Srivastava.
- AT&T/2005-0370 “System and Method for Generating Statistical Descriptors of a Data Stream”, with F. Korn, S. Muthukrishnan and D. Srivastava.
- AT&T/2005-0492 and 2005-0492a “Method for Compressed Sensor Measurements” with S. Muthukrishnan.
- Lucent/Cormode 4-1-1 “Communication-Efficient Distributed Monitoring Of Thresholded Counts”, with J. Ramimirtham and R. Keralapura.

Recent Invited Talks and Tutorials

- “Streaming in a Connected World: Querying and Tracking Distributed Data Stream”. Tutorial presented at VLDB 2006.
- “Clustering and Data Stream Analysis”, DIMACS Tutorial on Data Mining and Epidemiology, March 2006.
- “Towards an algorithmic theory of compressed sensing”, Schloss Dagstuhl, July 2005, CISS Conference (Princeton) March 2006.
- “Algorithms for processing massive data at network line speed”, Iowa, Minnesota, Dartmouth, Google, AT&T, CWRU, Poly, 2004.
- “What’s Hot, What’s Not, What’s New and What’s Next” at DIMACS Mixer, AT&T Shannon Labs, September 2003.

- “Zeroing in on the L_0 Metric” at DIMACS Workshop on Discrete Metric Spaces and their Algorithmic Applications, Princeton, August 2003.
- “Tracking Frequent Items Dynamically” at the Institute of Advanced Studies, DIMACS, Stonybrook and U. Penn, Spring 2003.
- “Algorithmic Embeddings for Comparing Large Text Streams” at CCR/DIMACS Workshop on Mining Massive Data Sets and Streams: Mathematical Methods and Algorithms for Homeland Defense, June 2002.
- “Embeddings of Metrics on Strings and Permutations” at Workshop on Discrete Metric Spaces and their Algorithmic Applications, Haifa, Israel, March 2002.

Professional Activities

- Co-editor, “AMS-DIMACS Volume on Discrete Methods in Epidemiology” with James Abello. To appear 2006.
- Program Committee Member for MineNet 2006, DAWAK 2006, DMSN 2006, VLDB 2005, ICDE 2004, KDD 2004, ICDM 2004, 2005.
- Co-organizer of DIMACS Seminar Series on Quantitative Biology and the DIMACS Working Group on Data Mining and Epidemiology, 2003-2004.
- Conference referee for STOC, FOCS, SODA, ICDE, PODS, FSTTCS, ICALP, ALENEX, CPM, SDM, SWAT, and ESA.
- Journal referee for VLDB Journal, ACM Transactions on Algorithms, IEEE TKDE, Algorithmica, SICOMP Journal, and others.