

Michael Hay

Associate Professor
Department of Computer Science
Colgate University
13 Oak Drive, Hamilton, NY 13346
(315) 228-7650, mhay@colgate.edu

EDUCATION

University of Massachusetts Amherst
Ph.D., Computer Science, 2010
Dissertation: Enabling Accurate Analysis of Private Network Data
Committee: Gerome Miklau, David Jensen, Don Towsley, Andrew Papachristos

University of Massachusetts Amherst
M.S., Computer Science, 2007
GPA: 4.0 / 4.0

Dartmouth College
A.B. *cum laude*, Computer Science, 1998
GPA: 3.6 / 4.0

EMPLOYMENT

Colgate University, Hamilton, NY <i>Associate Professor of Computer Science</i>	2018 - present
Tumult Labs, Raleigh, NC <i>Co-founder & Principal Scientist</i>	2019 - present
U.S. Census Bureau, Washington, DC <i>Research Data Scientist, Center for Disclosure Avoidance Research</i>	2017 - 2019
Colgate University, Hamilton, NY <i>Assistant Professor of Computer Science</i>	2012 - 2018
University of Massachusetts Amherst, Amherst, MA <i>Visiting Research Professor of Computer Science</i>	2015 - 2016
Cornell University, Ithaca, NY <i>Computing Innovation Fellow</i>	2010 - 2012
University of Massachusetts Amherst, Amherst, MA <i>Research Assistant</i>	2002 - 2010
Adverplex, Cambridge, MA <i>Software Engineering Intern</i>	Summer 2008
Second Nature, Boston, MA <i>Program Manager</i>	2000 - 2002
Kenan Systems, Denver, CO <i>Software Engineer</i>	1998 - 2000

TEACHING

Colgate University
Undergraduate liberal arts institution with a 5 course teaching load

COSC 101: Intro to Computing I (every semester Fall 2012-Spring 2015)
COSC 290: Discrete Structures (Fall 2017, Spring 2018)
COSC 460: Database Management Systems (Fall 2014, Fall 2016, Fall 2018)
COSC 480: Topics: Database Management Systems (Fall 2012)
COSC 480: Topics: Artificial Intelligence (Spring 2014)
COSC 480: Topics: Data Science (Spring 2017)
Each COSC course taught includes a 2 hour laboratory in addition to lecture.
CORE 109S: In Data We Trust? (Spring 2017)
FSEM 125: In Data We Trust? (Fall 2016)

Cornell University

Courses taught during postdoctoral fellowship
First Year Writing seminar: CS 1330 Privacy in Bits: How Digital Technology is Reshaping Privacy (Spring 2012)
CS 2110 Object-Oriented Programming and Data Structures (Summer 2011)

PUBLISHED WORKS

Refereed Journal Publications

- [J-8] Yu-Hsuan Kuo, Cho-Chun Chiu, Daniel Kifer, Michael Hay, and Ashwin Machanavajjhala. “Differentially Private Hierarchical Count-of-counts Histograms”. In: *Proceedings of the VLDB Endowment* 11.11 (July 2018), pp. 1509–1521. ISSN: 2150-8097. DOI: 10.14778/3236187.3236202. URL: <https://doi.org/10.14778/3236187.3236202>.
- [J-7] Ryan McKenna, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. “Optimizing Error of High-dimensional Statistical Queries Under Differential Privacy”. In: *Proceedings of the VLDB Endowment* 11.10 (June 2018), pp. 1206–1219. ISSN: 2150-8097. DOI: 10.14778/3231751.3231769. URL: <https://doi.org/10.14778/3231751.3231769>.
- [J-6] Chao Li, Gerome Miklau, Michael Hay, Andrew McGregor, and Vibhor Rastogi. “The matrix mechanism: optimizing linear counting queries under differential privacy”. In: *International Journal on Very Large Data Bases (VLDB Journal)* 24.6 (2015), pp. 757–781.
- [J-5] Chao Li, Michael Hay, Gerome Miklau, and Yue Wang. “A data- and workload-aware query answering algorithm for range queries under differential privacy”. In: *Proceedings of the VLDB Endowment (PVLDB)* 7.5 (2014), pp. 341–352.
- [J-4] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Chao Li. “Resisting structural re-identification in anonymized social networks”. In: *International Journal on Very Large Data Bases (VLDB Journal)* 19.6 (2010), pp. 797–823.
- [J-3] Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu. “Boosting the accuracy of differentially private histograms through consistency”. In: *Proceedings of the VLDB Endowment (PVLDB)* 3.1-2 (2010), pp. 1021–1032.
- [J-2] Michael Hay, Gerome Miklau, David Jensen, Donald F. Towsley, and Philipp Weis. “Resisting structural re-identification in anonymized social networks”. In: *Proceedings of the VLDB Endowment (PVLDB)* 1.1 (2008), pp. 102–114.
- [J-1] Amy McGovern, Lisa Friedland, Michael Hay, Brian Gallagher, Andy Fast, Jen Neville, and David Jensen. “Exploiting relational structure to understand publication patterns in high-energy physics”. In: *ACM SIGKDD Explorations* 5.2 (2003), pp. 165–172.

Refereed Conference Publications¹

¹Computer Science is atypical from most sciences in that *conference proceedings constitute a primary publication venue*, on par with or in some cases preferable to journals. Premier conferences are highly selective and conference proceedings are archival. More information can be found in “Evaluating Computer Scientists and Engineers For Promotion and Tenure,” a Computing Research Association Best Practice Memo.

- [C-19] Sameera Ghayyur, Yan Chen, Roberto Yus, Ashwin Machanavajjhala, Michael Hay, Gerome Miklau, and Sharad Mehrotra. “IoT-Detective: Analyzing IoT Data Under Differential Privacy”. In: *Proceedings of the 2018 International Conference on Management of Data*. SIGMOD ’18. Houston, TX, USA: ACM, 2018, pp. 1725–1728. ISBN: 978-1-4503-4703-7. DOI: 10.1145/3183713.3193571. URL: <http://doi.acm.org/10.1145/3183713.3193571>.
- [C-18] Dan Zhang, Ryan McKenna, Ios Kotsogiannis, Michael Hay, Ashwin Machanavajjhala, and Gerome Miklau. “EKTELO: A Framework for Defining Differentially-Private Computations”. In: *Proceedings of the 2018 International Conference on Management of Data*. SIGMOD ’18. Houston, TX, USA: ACM, 2018, pp. 115–130. ISBN: 978-1-4503-4703-7. DOI: 10.1145/3183713.3196921. URL: <http://doi.acm.org/10.1145/3183713.3196921>.
- [C-17] Garret Bernstein, Ryan McKenna, Tao Sun, Michael Hay, Gerome Miklau, and Dan Sheldon. “Differentially private learning of graphical models using CGMs”. In: *International Conference on Machine Learning (ICML)*. 2017.
- [C-16] Yan Chen, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau. “PeGaSus: Data-Adaptive Differentially Private Stream Processing”. In: *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security*. CCS ’17. Dallas, Texas, USA: ACM, 2017, pp. 1375–1388. ISBN: 978-1-4503-4946-8. DOI: 10.1145/3133956.3134102. URL: <http://doi.acm.org/10.1145/3133956.3134102>.
- [C-15] Michael Hay, Gerome Miklau, and Liudmila Elagina. “Differentially private rank aggregation”. In: *SIAM International Conference on Data Mining (SDM)*. 2017.
- [C-14] Ios Kotsogiannis, Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, and Margaret Orr. “DIAS: differentially private interactive algorithm selection using Pythia”. In: *ACM International Conference on the Management of Data (SIGMOD) Demo Track*. 2017.
- [C-13] Ios Kotsogiannis, Ashwin Machanavajjhala, Michael Hay, and Gerome Miklau. “Pythia: data dependent differentially private algorithm selection”. In: *ACM International Conference on the Management of Data (SIGMOD)*. 2017.
- [C-12] Ashwin Machanavajjhala, Xi He, and Michael Hay. “Differential privacy in the wild: a tutorial on current practices & open challenges”. In: *ACM International Conference on the Management of Data (SIGMOD)*. 2017.
- [C-11] Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, and Dan Zhang. “Principled evaluation of differentially private algorithms using DPBench”. In: *ACM International Conference on the Management of Data (SIGMOD)*. 2016, pp. 139–154.
- [C-10] Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, Dan Zhang, and George Bissias. “Exploring privacy-accuracy tradeoffs using DPComp”. In: *ACM International Conference on the Management of Data (SIGMOD) Demo Track*. 2016, pp. 2101–2104.
- [C-9] Johannes Gehrke, Michael Hay, Edward Lui, and Rafael Pass. “Crowd-blending privacy”. In: *Conference on Advances in Cryptology (CRYPTO)*. 2012, pp. 479–496.
- [C-8] Michael Hay, Kun Liu, Gerome Miklau, Jian Pei, and Evimaria Terzi. “Privacy-aware data management in information networks (tutorial)”. In: *ACM International Conference on the Management of Data (SIGMOD)*. 2011, pp. 1201–1204.
- [C-7] Xiaokui Xiao, Gabriel Bender, Michael Hay, and Johannes Gehrke. “iReduct: differential privacy with reduced relative errors”. In: *ACM International Conference on the Management of Data (SIGMOD)*. 2011, pp. 229–240.
- [C-6] Chao Li, Michael Hay, Vibhor Rastogi, Gerome Miklau, and Andrew McGregor. “Optimizing linear counting queries under differential privacy”. In: *ACM Symposium on Principles of Database Systems (PODS)*. 2010, pp. 123–134.

- [C-5] Michael Hay, Chao Li, Gerome Miklau, and David Jensen. “Accurate estimation of the degree distribution of private networks”. In: *IEEE International Conference on Data Mining (ICDM)*. 2009, pp. 169–178.
- [C-4] Vibhor Rastogi, Michael Hay, Gerome Miklau, and Dan Suciu. “Relationship privacy: output perturbation for queries with joins”. In: *ACM Symposium on Principles of Database Systems (PODS)*. 2009, pp. 107–116.
- [C-3] Ben Wellner, Andrew McCallum, Fuchun Peng, and Michael Hay. “An integrated, conditional model of information extraction and coreference for citation matching”. In: *Conference on Uncertainty in Artificial Intelligence (UAI)*. 2004, pp. 593–601.
- [C-2] David Jensen, Jennifer Neville, and Michael Hay. “Avoiding bias when aggregating relational data with degree disparity”. In: *International Conference on Machine Learning (ICML)*. 2003, pp. 274–281.
- [C-1] Jennifer Neville, David Jensen, Lisa Friedland, and Michael Hay. “Learning relational probability trees”. In: *ACM International Conference on Knowledge Discovery and Data Mining (KDD)*. 2003, pp. 625–630.

Refereed Workshop Publications

- [W-5] Ryan McKenna, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. “Optimizing Error of High-dimensional Statistical Queries Under Differential Privacy”. In: *To appear at Theory and Practice of Differential Privacy Workshop at CCS*. 2018.
- [W-4] Garret Bernstein, Ryan McKenna, Tao Sun, Michael Hay, Gerome Miklau, and Dan Sheldon. “Differentially private learning of graphical models using CGMs”. In: *Private and Secure Machine Learning at ICML*. 2017.
- [W-3] Dan Zhang, Ryan McKenna, Ios Kotsogiannis, Gerome Miklau, Michael Hay, and Ashwin Machanavajjhala. “EKTELO: A framework for defining differentially-private computations”. In: *Theory and Practice of Differential Privacy Workshop at the ACM Conference on Computer and Communications Security*. 2017.
- [W-2] Michael Hay, Ashwin Machanavajjhala, Gerome Miklau, Yan Chen, and Dan Zhang. “Principled evaluation of differentially private algorithms”. In: *Theory and Practice of Differential Privacy Workshop at ICML*. 2016.
- [W-1] Dan Zhang, Michael Hay, Gerome Miklau, and Brendan O’Connor. “Challenges of visualizing differentially private data”. In: *Theory and Practice of Differential Privacy Workshop at ICML*. 2016.

Tutorials²

- [T-1] Ashwin Machanavajjhala, Xi He, and Michael Hay. “Differential privacy in the wild: a tutorial on current practices & open challenges”. In: *Proceedings of the VLDB Endowment (PVLDB)* 9.13 (2016), pp. 1611–1614.

Book Chapters

- [B-1] Michael Hay, Gerome Miklau, and David Jensen. “Enabling accurate analysis of private network data”. In: *Privacy-Aware Knowledge Discovery: Novel Applications and New Techniques*. CRC Press, 2010.

²Tutorials are scholarly works created to present an in-depth survey of a particular topic, providing non-experts with the essential knowledge to commence research in the topic. They are in some ways akin to a review paper but presented in the format of a 1.5-3 hour conference presentation. Tutorial selection is a competitive process based on peer-review of proposals.

INVITED TALKS

L'Ecole Nationale de la Statistique et de l'Analyse de l'Information (ENSAI) European Course in Advanced Statistics: Statistical Disclosure Control for Official Statistics. “*Differential Privacy: Tutorial and Applications.*” Bruz, France, February 2018.

IEEE Symposium on Privacy-Aware Computing, HotPrivacy Workshop. “*ektelo: A Framework for Defining Differentially Private Computations.*” Washington DC, August 2017.

Schloss Dagstuhl. “*Fairness Through Awareness & Learning Fair Representations.*” Wadern, Germany, July 2016.

Mount Holyoke College. “*Privacy in the Era of Big Data.*” South Hadley, MA, March 2016.

Boston University, Charles River Workshop on Private Analysis of Social Networks. “*Private Analysis of Network Data.*” Boston, MA, May 2014.

DIMACS, Rutgers University, Workshop on Recent Work on Differential Privacy across Computer Science. “*iReduct: Differential Privacy with Reduced Relative Errors.*” Piscataway, NJ, October 2012.

ACM International Conference on Knowledge Discovery and Data Mining (KDD). “*Enabling Accurate Analysis of Private Network Data.*” August 2011, Doctoral Dissertation Award Talk. San Diego, CA

IBM Thomas J. Watson Research Center. “*Enabling Accurate Analysis of Private Network Data.*” Yorktown Heights, NY, October 2009.

PRESENTATIONS

Computing Community Consortium (CCC) Early Career Researcher Symposium. “*Towards Systems for Data Science with Formal Privacy Guarantees.*” Poster presentation, Washington, DC, August 2018.

ACM International Conference on the Management of Data (SIGMOD). “*IoT-Detective: Analyzing IoT Data Under Differential Privacy.*” Demo presentation, Houston, TX, June 2018.

DARPA Brandeis Program PI Meeting. “*System P: A Data Analytics Engine With Customizable Privacy and Optimized Utility.*” Talk, Berkeley, CA, April 2018.

ACM International Conference on the Management of Data (SIGMOD). “*Differential Privacy In The Wild: A Tutorial On Current Practices & Open Challenges.*” Tutorial presentation, Chicago, IL, May 2017.

ACM International Conference on the Management of Data (SIGMOD). “*DIAS: Differentially Private Interactive Algorithm Selection using Pythia.*” Demo presentation, Chicago, IL, May 2017.

SIAM International Conference on Data Mining (SDM). “*Differentially Private Rank Aggregation.*” Paper presentation, Houston, TX, April 2017.

Colgate University. “*Data Analysis with Privacy Protection: Seeing the Forest But Not the Trees.*” NASC Colloquium, Hamilton, NY, April 2017.

National Science Foundation SaTC PI Meeting. “*Pythia: Algorithm Selection for Differential Privacy.*” Poster presentation, Arlington, VA, January 2017.

DARPA Brandeis Program PI Meeting. “*System P: A Data Analytics Engine With Customizable Privacy and Optimized Utility.*” Talk, Woburn, MA, October 2016.

ACM International Conference on the Management of Data (SIGMOD). “*Principled Evaluation of Differentially Private Algorithms Using DPBench.*” Paper presentation, San Francisco, CA, June 2016.

ACM International Conference on the Management of Data (SIGMOD). “*Exploring Privacy-Accuracy Tradeoffs Using DPComp.*” Demo presentation, San Francisco, CA, June 2016.

Theory and Practice of Differential Privacy Workshop at ICML. “*Principled Evaluation of Differentially Private Algorithms.*” Paper presentation, New York, NY, June 2016.

UMass Amherst. “*Principled Evaluation of Differentially Private Algorithms Using DPBench.*” Presentation to Database Group, Amherst, MA, December 2015.

Cornell University. “*Analyzing Private Network Data.*” Presentation to Database group, Ithaca, NY, November 2011.

ACM International Conference on the Management of Data (SIGMOD). “*iReduct: Differential Privacy with Reduced Relative Errors.*” Paper presentation, Athens, Greece, June 2011.

ACM International Conference on the Management of Data (SIGMOD). “*Privacy-aware Data Management In Information Networks.*” Tutorial presentation, Athens, Greece, June 2011.

International Conference on Very Large Data Bases (VLDB). “*Boosting the Accuracy of Differentially-Private Histograms Through Consistency.*” Paper presentation, Singapore, September 2010.

Institute for Pure & Applied Mathematics, UCLA. “*Accurate Estimation of the Degree Distribution of Private Networks.*” Poster presentation, Los Angeles, CA, February 2010.

IEEE International Conference on Data Mining (ICDM). “*Accurate Estimation of the Degree Distribution of Private Networks.*” Best student paper award talk, Miami, FL, December 2009.

International Conference on Very Large Data Bases (VLDB). “*Resisting Structural Re-identification in Anonymized Social Networks.*” Paper presentation, Auckland, New Zealand, August 2008.

GRANTS & FELLOWSHIPS (EXTERNAL)

Defense Advanced Research Projects Agency: *System P: A Data Analytics Engine With Customizable Privacy and Optimized Utility*
\$2,781,810 total, \$433,920 to Colgate. Role: Co-Investigator (PI: Gerome Miklau, UMass Amherst), 2015-2020.

National Science Foundation: *TWC: Medium: Collaborative: Re[DP]: Realistic Data Mining Under Differential Privacy*
\$273,424. Role: Principal Investigator, 2014-2018.

National Science Foundation/Computing Research Association: *Computing Innovation Fellowship*
\$150,000, Role: Postdoctoral Fellow, 2010-2012.

GRANTS & FELLOWSHIPS (COLGATE)

Teaching with Technology Micro Grant, Spring 2017
Exploring Cloud-based Data Management Systems in Core Scientific Perspectives

Colgate Picker Fellowship, 2015-2016
Analyzing Network Mobility Traces Under Differential Privacy

Teaching with Technology Micro Grant, Spring 2015
Peer Instruction in Introduction to Computing I

Faculty Development Grant, Fall 2012
Exploring How to Broaden Participation in Computer Science Through Innovation in the Introductory Curriculum

PROFESSIONAL SERVICE

Program Committees, Review Boards & Panels

PC Member, ACM International Conference on the Management of Data (SIGMOD), 2019, 2018, 2017, 2016, 2015, 2012
→ **Distinguished PC Member**, SIGMOD 2018
Review Board, Proceedings of the VLDB Endowment (PVLDB), 2019, 2017, 2013, 2012, 2010
PC Member, Theory and Practice of Differential Privacy Workshop (TPDP), 2018
Panel Member, National Science Foundation, 2016, 2011
PC Member, IEEE International Conference on Data Engineering (ICDE), 2014, 2011
PC Member, ACM SIGMOD-SIGACT-SIGART Symposium on Principles Of Database Systems (PODS), 2014, 2013
PC Member, ACM SIGMOD Workshop on Databases and Social Networks (DBSocial), 2013
PC Member, IEEE ICDE International Workshop on Privacy-Preserving Data Publication and Analysis (PrivDB), 2013
PC Member, Privacy Enhancing Technologies Symposium (PETS), 2013
PC Member, World Wide Web Conference (WWW), 2012
PC Member, International Conference on Machine Learning (ICML), 2011
PC Member, Privacy Aspects of Data Mining Workshop at the IEEE International Conference on Data Mining (ICDM), 2011

Reviewer

ACM Transactions on Database Systems (TODS)
ACM Transactions on Information and System Security (TISSEC)
ACM Transactions on Knowledge Discovery from Data (TKDD)
Computers & Security
Data Mining and Knowledge Discovery (DMKD)
European Symposium on Algorithms (ESA)
IEEE Transactions on Dependable and Secure Computing (TDSC)
IEEE Transactions on Information Forensics and Security
IEEE Transactions on Knowledge and Data Engineering (TKDE)
IEEE Transactions on Network Science and Engineering (TNSE)
IEEE Transactions on Parallel and Distributed Systems (TPDS)
Information Processing Letters
International Journal on Very Large Data Bases (VLDBJ)
Journal of Biomedical Informatics (JBI)
Journal of the American Medical Informatics Association (JAMIA)
Machine Learning Journal
National Science Foundation (External Reviewer)
Nature Human Behavior
Swiss National Science Foundation (External Reviewer)
Transactions on Data Privacy (TDP)

CONSULTANT SERVICES

U.S. Census Bureau, Washington, DC (2017 - present)
Research Data Scientist, Center for Disclosure Avoidance Research

AWARDS

Distinguished Program Committee Member, ACM Conference on Management of Data (SIGMOD), 2018.
Dissertation Award, ACM Special Interest Group on Knowledge Discovery & Data Mining, 2011
Dissertation Award, U. of Massachusetts Amherst, CS Department, sponsored by Yahoo, 2011

Best Student Paper Award, IEEE International Conference on Data Mining, 2009
KDD Cup Winner, ACM International Conference on Knowledge Discovery & Data Mining, 2003
Citation for Meritorious Scholarship, Dartmouth College, 1996

UNIVERSITY SERVICE

University Committees

Elected Member, Student Conduct Board (AY 2017-present, 3 year term)
Appointed Member, Student Affairs Board (AY 2017)
Appointed Member, Picker ISI Committee (AY 2016, 2017, 2018)
Appointed Member, *ad hoc* Research Computing Committee (AY 2016, 2017, 2018)
Chair, Committee on Information Technology (Fall 2016)
Member (*ex officio*), Committee on Information Technology (Fall 2013-Spring 2015)
Member, Search Committee for Research Computing Consultant in ITS (Summer 2014)
Appointed Member, Campus Communication Council (Spring 2014)

Departmental Service

Adviser to 20-25 students
Mentor for undergraduate research
Asad Jamil, Summer 2017
Margaret Orr, Spring 2017
Ahsan Mahmood, Fall 2016
Cindy Han, Summer 2015
Abeneazer Chafamo, Summer 2015
Soo Bin Kwon, Summer-Fall 2014
Dong Mai, Summer 2014
Member of search committees: tenure stream assistant professor (2014, 2015, 2016, 2018),
visiting assistant professor (2015, 2016, 2017, 2018), lab instructor (2013)
Departmental adviser to the Computer Science national honor society, Upsilon Pi Epsilon
(2013-2015, 2016-2017)
Liaison to Colgate Admissions office (2014-2015, 2016-present)
Weekly department tea organizer (Spring 2015) and presenter (various dates)

Additional Service Activities

Panelist for NASC Colloquium, Fall 2016, "*What I wish I had known about grants...*"
Organized CLTR Teaching Table, Spring 2015 "*Making Thinking Visible*"
Taught Colgate High School Seminar on "*Amazing Algorithms: the Incredible Ideas that Drive the Digital Age*", Fall 2013