Computer Science Department
University of Massachusetts Amherst
140 Governors Drive
Amherst, MA 01003

MICHAEL HAY

413-835-5618
mhay@cs.umass.edu
www.cs.umass.edu/~mhay

RESEARCH INTERESTS

Knowledge discovery, data mining, database management, social network analysis, data privacy.

EDUCATION

University of Massachusetts Amherst — expected 2010
Ph.D., Computer Science

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

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

RESEARCH AND PROFESSIONAL EXPERIENCE

University of Massachusetts Amherst — 2006 - present
Graduate Research Assistant
Advisors: Gerome Miklau and David Jensen

Thesis work focuses on developing tools that enable accurate analysis of social network data while protecting privacy. Results thus far: Demonstrated that simple techniques for privacy protection are vulnerable to attack. Designed algorithm that anonymizes the network via clustering. Developed techniques for accurately estimating network statistics (clustering coefficient, degree distribution) while ensuring strong privacy guarantees.

University of Massachusetts Amherst — 2002 - 2006
Graduate Research Assistant
Advisor: David Jensen

Developed machine learning algorithms for information extraction, duplication detection, and classification in relational data. Used our techniques to win the 2003 ACM KDD Cup.

Adverplex, Cambridge, MA — Summer 2008
Intern

Developed machine learning tools to identify relevant keywords for search engine marketing campaign.

Second Nature, Boston, MA — 2000 - 2002
Program Manager/IT Director

Responsible for maintaining IT infrastructure. Developed web portal for environmental resources.

Kenan Systems Corporation (acquired by Lucent), Denver, CO — 1998 - 2000
Software Engineer/Consultant

Helped develop award-winning customer care software. Developed tools for importing data from legacy systems. Consulted clients on software installation, configuration, and live production.

PUBLICATIONS

Working papers

Boosting the accuracy of differentially private queries through consistency
Michael Hay, Vibhor Rastogi, Gerome Miklau, and Dan Suciu
submitted to Proceedings of the VLDB Endowment (VLDB)

Resisting structural re-identification in anonymized social networks
Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Chao Li
submitted to VLDB Journal

Refereed Publications

Accurate estimation of the degree distribution of private networks
Michael Hay, Chao Li, Gerome Miklau, and David Jensen
IEEE International Conference on Data Mining series (ICDM) 2009

Relationship privacy: output perturbation for queries with joins
Vibhor Rastogi, Michael Hay, Gerome Miklau, and Dan Suciu
ACM Symposium on Principles of Database Systems (PODS) 2009

Resisting structural re-identification in anonymized social networks
Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis
International Conference on Very Large Data Bases (VLDB) 2008

An integrated, conditional model of information extraction and coreference for citation matching
Ben Wellner, Andrew McCallum, Fuchun Peng, and Michael Hay
Conference on Uncertainty in Artificial Intelligence (UAI) 2004

Exploiting relational structure to understand publication patterns in high-energy physics
Amy McGovern, Lisa Friedland, Michael Hay, Brian Gallagher, Andy Fast, Jen Neville, David Jensen
ACM SIGKDD Explorations 2003

Learning relational probability trees
Jennifer Neville, David Jensen, Lisa Friedland, and Michael Hay
ACM International Conference on Knowledge Discovery and Data Mining (KDD) 2003

Avoiding bias when aggregating relational data with degree disparity
David Jensen, Jennifer Neville, and Michael Hay
International Conference on Machine Learning (ICML) 2003

Technical Reports

Anonymizing social networks
Michael Hay, Gerome Miklau, David Jensen, Philipp Weis, and Siddharth Srivastava
University of Massachusetts Amherst Technical Report UM-CS-2007-021 2007

Understanding the effects of search constraints on structure learning
Michael Hay, Andrew Fast, and David Jensen
University of Massachusetts Amherst Technical Report UM-CS-2007-019 2007

SOFTWARE DEVELOPMENT EXPERIENCE

- Contributed to PROXIMITY, open-source software for relational knowledge discovery. PROXIMITY has installations at national labs, industrial corporations, and other academic institutions; it has been downloaded over 10,000 times. My contributions include: the relational probability tree algorithm; a graphical user-interface for data browsing; a query optimization engine for the QGRAPH language.
- Programming languages: Java, Python, C++, R, SQL.

TEACHING EXPERIENCE

University of Massachusetts Amherst

Instructor, CS 121 Introduction to Problem Solving with Computers, Summer 2006
Average rating of 4.7/5.0 on students' course evaluations.

Teaching Assistant, CS 121 Introduction to Problem Solving with Computers, Spring 2006

Guest Lecturer, CS 121 (Fall 2006), CS 383 (Fall 2006), CS 591Y (Spring 2005), CS745 (Fall 2008)

Organizer, Graduate-level reading group on "Inference in graphical models," Summer 2004

AWARDS

Nominee for Graduate Student Fellowship, University of Massachusetts Amherst, 2004

Winner of KDD Cup at ACM International Conference on Knowledge Discovery and Data Mining, 2003

Citation for Meritorious Scholarship, Dartmouth College, 1996

SERVICE

Professional reviewing

External reviewer: ICDE (2006), ICDM (2005, 2007), SIGMOD (2009), VLDB (2007)

Reviewer: TKDE (2008), Chapman & Hall/CRC Press (2009)

REFERENCES

Gerome Miklau (co-advisor)

Assistant Professor
Computer Science Department
University of Massachusetts Amherst
miklau@cs.umass.edu
(413) 545-4032

David Jensen (co-advisor)

Associate Professor
Computer Science Department
University of Massachusetts Amherst
jensen@cs.umass.edu
(413) 545-9677