

## ARINDAM BANERJEE

Associate Professor  
 McKnight Land-Grant Professor (2009-11)  
 Department of Computer Science & Engineering  
 Resident Fellow, Institute on the Environment (2010-13)  
 University of Minnesota, Twin Cities  
<http://www-users.cs.umn.edu/~banerjee>

### Education

Degree	Institution	Date Degree Granted
B.E. Electronics & Telecomm. Engg.	Jadavpur University	1997
M. Tech. Electrical Engineering	Indian Institute of Technology (IIT) Kanpur	1999
Ph.D. Electrical & Computer Engineering Advisor: Joydeep Ghosh	University of Texas at Austin	2005

### Positions/Employment

University of Minnesota, Twin Cities	Associate Professor	2011–Present
University of Minnesota, Twin Cities	Assistant Professor	2005–2011
Graduate Appointments		
IBM T. J. Watson Research Center	Graduate Summer Intern	2003,2004
University of Texas at Austin	Graduate Research Assistant	2000-2003
Interwoven Inc.	Graduate Summer Intern	2000,2001
University of Texas at Austin	Teaching Assistant	1999-2000
Indian Institute of Technology (IIT), Kanpur	Graduate Research Assistant	1998-1999
Indian Institute of Technology (IIT), Kanpur	Teaching Assistant	1997-1998

### Current Membership in Professional Organizations

Association for Computing Machinery (ACM)

## HONORS, AWARDS, AND VISITING POSITIONS

### Honors and Awards

- NSF CAREER Award, 2010-2015.
- Resident Fellow, Institute on the Environment (IonE), 2010-2013.
- McKnight Land-Grant Professorship, 2009-11.
- Best of SDM Award, SIAM International Conference on Data Mining (SDM), April 2007.
- Best Student Paper runner-up, ACM International Workshop on Knowledge Discovery from Sensor Data (SensorKDD), August 2007.
- J. T. Oden Faculty Research Fellowship, Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Summer 2006.
- Nominated for Best Dissertation Award, University of Texas at Austin, 2006.
- Best Research Paper Award, University Cooperative Society Research Excellence Awards, University of Texas at Austin, March 2005.
- Best Paper Award, SIAM International Conference on Data Mining (SDM), April 2004.
- IBM PhD fellowship for the academic years 2003-2004 and 2004-2005.
- 1<sup>st</sup> in IBM internal data mining contest, Summer 2003.
- Various travel awards including KDD 2005, KDD 2004, KDD 2003, ISIT 2004 awards; GEC (UT Austin) travel grant; and NIPS 2004 complimentary registration.
- Bronze Medal as a special academic award from Jadavpur University, in Spring 1997.
- Selected for the Indian National Mathematical Olympiads and attended the International Mathematical Olympiads camp for two successive years 1992, 1993.

### Visiting Professorships or Visiting Scholar Positions

- Visiting Professor, Max Planck Institute (MPI) for Biological Cybernetics, Tübingen, Germany, August, 2008.
- Invited academic visitor, Institute of Pure and Applied Mathematics (IPAM), University of California, Los Angeles (UCLA), October, 2007.
- Visiting Fellow, Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, Summer 2006.

## RESEARCH

### External Grants and Contracts

- Collaborative Research: “Understanding Climate Change: A Data Driven Approach”  
National Science Foundation (NSF)  
Co-PI, with Vipin Kumar (PI), and others  
September 1, 2010 - August 31, 2015  
Total Amount: \$9,999,739.
- “CDI-Type II: Computational Tools for Behavioral Analysis, Diagnosis, and Intervention of At Risk Children”  
National Science Foundation (NSF)  
Co-PI, with Nikolas Papanikopoulos (PI), and others  
September 1, 2010 - August 31, 2014  
Total Amount: \$1,599,207.
- “NetSE: Small: Spatio-Temporal Network Traffic Dynamics and Interactions of Social-Technical Networks”  
National Science Foundation (NSF)  
Co-PI, with Zhi-Li Zhang (PI)  
September 1, 2010 - August 31, 2013  
Total Amount: \$500,000.
- “CAREER: Combinatorial Online Learning and its Applications”  
National Science Foundation (NSF)  
PI  
April 1, 2010 - March 31, 2015  
Total Amount: \$495,801.
- “RI: Small: Statistical Modeling of Dynamic Covariance Matrices”  
National Science Foundation (NSF)  
PI, with Daniel Boley (co-PI)  
September 1, 2009 - August 31, 2012  
Total Amount: \$499,999.
- “III-COR-Small: Multi-Relational Data Clustering with Probabilistic Mixture Models”  
National Science Foundation (NSF)  
PI  
September 1, 2008 - August 31, 2011  
Total Amount: \$399,609.

- “Detecting Anomalies from Numeric and Textual Data using Data Mining”  
National Aeronautics and Space Agency (NASA)  
Co-PI, with Jaideep Srivastava (PI) and others  
January 1, 2008 - December 31, 2010  
Total Amount: \$994,859.
- “CRI: Research Infrastructure for Emerging Network Systems and Applications”  
National Science Foundation (NSF)  
Co-PI, with Zhi-Li Zhang (PI) and others  
September 1, 2007 - August 31, 2009  
Total Amount: \$199,999.
- “Discovering Effective Models for Home Visiting Practice”  
Midwest Nursing Research Society  
Co-I, with Karen Monsen (PI)  
May 01, 2008 - April 30, 2009  
Total Amount: \$9,824.
- “Dynamic Graphical Models for Knowledge Discovery and Predictive Modeling of Social Networks”  
Oak Ridge National Labs (ORNL)  
PI  
January 16, 2006 - September 30, 2007  
Total Amount: \$92,310.

## Internal Grants

- “Automation of Problem List Generation for a Personal Health Summary”  
Co-PI, with Lee Pyles (PI), and others  
University of Minnesota Interdisciplinary Informatics (UMII)  
July 1, 2011 - June 30, 2012  
Total Amount: \$74,866.
- “Data Mining based Intelligent Clinical Decision Information Systems”  
Co-PI, with George Karypis (PI)  
Digital Technology Center (DTC)  
September 1, 2009 - August 31, 2010  
Total Amount: \$80,200.
- “Towards Social Search: Topic Modeling and Query Routing in the Social Web”  
Grant-in-Aid, University of Minnesota Graduate School  
PI  
January 1, 2007 - June 31, 2008  
Total Amount: \$24,810.

## Publications

Refereed Journal Articles: (\* = PhD Student Advisee)

- J15. H. Wang, H. Shan\*, and A. Banerjee, "Bayesian Cluster Ensembles," *Statistical Analysis and Data Mining*, 4(1), 54-70, 2011.
- J14. H. Shan\* and A. Banerjee, "Mixed-Membership Naive Bayes Models," *Data Mining and Knowledge Discovery*, 23(1), 1-62, 2011.
- J13. V. Chandola\*, A. Banerjee, and V. Kumar, "Anomaly Detection for Discrete Sequences: A Survey," *IEEE Transactions on Knowledge and Data Engineering*, <http://doi.ieeeecomputersociety.org/10.1109/TKDE.2010.235>, 2010.
- J12. K. Monsen, A. Banerjee, and P. Das\*, "Discovering Client and Intervention Patterns in Home Visiting Data," *Western Journal of Nursing Research*, 32(8), 1031-1054, 2010.
- J11. Q. He, K. Chang, E.-P. Lim, and A. Banerjee, "Keep it Simple with Time: A re-examination of Probabilistic Topic Detection Models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(10), 1795-1808, 2010.
- J10. V. Chandola\*, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," *ACM Computing Surveys*, 41(3), Article 15, 2009.
- J9. A. Agovic\*, A. Banerjee, A. Ganguly, and V. Protopopescu, "Anomaly Detection in Transportation Corridors using Manifold Embedding," *Intelligent Data Analysis*, 13(3), 435-455, 2009.
- J8. J. Wan, S. Kang, C. Tang, J. Yan, Y. Ren, J. Liu, X. Gao, A. Banerjee, L. Ellis, and T. Li, "Meta-prediction of Phosphorylation Sites with Weighted Voting and Restricted Grid Search Parameter Selection," *Nucleic Acids Research*, 36(4), e22, doi: 10.1093/nar/gkm848, 2008.
- J7. A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation," *Journal of Machine Learning Research*, 8 (Aug), 1919-1986, 2007.
- J6. A. Banerjee and J. Ghosh, "Scalable Clustering Algorithms with Balancing Constraints," *Data Mining and Knowledge Discovery*, 13(3), 365-395, November, 2006.
- J5. V. Monga, A. Banerjee, and B. Evans, "A Clustering Based Approach to Perceptual Image Hashing," *IEEE Transactions on Information Forensics and Security*, 1(1), 68-79, March 2006.
- J4. A. Banerjee, X. Guo, and H. Wang, "On the Optimality of Conditional Expectation as a Bregman Predictor," *IEEE Transactions of Information Theory*, 51(7), 2664-2669, 2005.
- J3. A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," *Journal of Machine Learning Research*, 6 (Oct), 1705-1749, 2005.
- J2. A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Clustering on the Unit Hypersphere using von Mises-Fisher Distributions," *Journal of Machine Learning Research*, 6 (Sep), 1345-1382, 2005.
- J1. A. Banerjee and J. Ghosh, "Frequency Sensitive Competitive Learning for Scalable Balanced Clustering on High Dimensional Hyperspheres," *IEEE Transactions on Neural Networks*, 15(3), 702-719, May 2004.

Refereed Conference Proceedings: (\* = PhD Student Advisee, Name = Conference Presenter)

- C36. S. Kasiviswanathan, P. Melville, A. Banerjee, and V. Sindhvani, "Emerging Topic Detection using Dictionary Learning," *ACM Conference on Information and Knowledge Management (CIKM)*, 2011.
- C35. A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos, "Efficient Similarity Search for Covariance Matrices via the Jensen-Bregman LogDet Divergence," *International Conference on Computer Vision (ICCV)*, 2011.
- C34. H. Wang\*, A. Banerjee, and D. Boley, "Common Component Analysis for Multiple Covariance Matrices," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.
- C33. P. Das\* and A. Banerjee, "Meta Optimization and its Application to Portfolio Selection," *ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD)*, 2011.
- C32. A. Agovic\*, A. Banerjee, and S. Chatterjee, "Probabilistic Matrix Addition," *International Conference on Machine Learning (ICML)*, 2011.
- C31. H. Shan\* and A. Banerjee, "Generalized Probabilistic Matrix Factorizations for Collaborative Filtering," *IEEE International Conference on Data Mining (ICDM)*, 2010.
- C30. N. Pathak\*, A. Banerjee, and J. Srivastava, "A Generalized Linear threshold Model for Multiple Cascades," *IEEE International Conference on Data Mining (ICDM)*, 2010.
- C29. A. Agovic\*, H. Shan\*, and A. Banerjee, "Analyzing aviation safety reports: From topic modeling to scalable multi-label classification," *Conference on Intelligent Data Understanding (CIDU)*, 2010.
- C28. A. Agovic\* and A. Banerjee, "Gaussian Process Topic Models," *Conference on Uncertainty in Artificial Intelligence (UAI)*, 2010.
- C27. H. Shan\* and A. Banerjee, "Residual Bayesian Co-clustering for Matrix Approximation," *SIAM International Conference on Data Mining (SDM)*, 2010.
- C26. H. Zhu, G. Mateos, G. Giannakis, N. Sidiropoulos, and A. Banerjee, "Sparsity-Cognizant Overlapping Co-clustering for Behavior Inference in Social Networks," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, 2010.
- C25. H. Shan\*, A. Banerjee, and N. Oza, "Discriminative Mixed-membership Models," *IEEE International Conference on Data Mining (ICDM)*, 2009.
- C24. Q. Fu\* and A. Banerjee, "Bayesian Overlapping Subspace Clustering," *IEEE International Conference on Data Mining (ICDM)*, 2009.
- C23. S. Jegelka, S. Sra, and A. Banerjee, "Approximation Algorithms for Tensor Clustering," *The 20th International Conference on Algorithmic Learning Theory (ALT)*, 2009.
- C22. H. Wang, H. Shan\*, and A. Banerjee, "Bayesian Cluster Ensembles," *SIAM International Conference on Data Mining (SDM)*, 2009.
- C21. A. Agovic\*, M. Gini, and A. Banerjee, "Semi-Supervised Learning of User-Preferred Travel Schedules," *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2009.
- C20. H. Shan\* and A. Banerjee, "Bayesian Co-clustering," *IEEE International Conference on Data Mining (ICDM)*, 2008.
- C19. Q. Fu\* and A. Banerjee, "Multiplicative Mixture Models for Overlapping Clustering," *IEEE International Conference on Data Mining (ICDM)*, 2008.

- C18. K. Hsu, A. Banerjee, and J. Srivastava, "I/O Scalable Bregman Co-clustering," *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*, 2008.
- C17. A. Banerjee and H. Shan\*, "Latent Dirichlet Conditional Naive Bayes Models," *IEEE International Conference on Data Mining (ICDM)*, 2007.
- C16. A. Banerjee, S. Basu, S. Merugu, "Multi-way Clustering on Relation Graphs," *SIAM International Conference on Data Mining (SDM)*, 2007.
- C15. A. Banerjee, "An Analysis of Logistic Models: Exponential Family Connections and Online Performance," *SIAM International Conference on Data Mining (SDM)*, 2007.
- C14. A. Banerjee and S. Basu, "Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning," *SIAM International Conference on Data Mining (SDM)*, 2007.
- C13. A. Banerjee, "On Bayesian Bounds," *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, 81-88, 2006.
- C12. A. Banerjee, C. Krumpelman, S. Basu, R. Mooney, and J. Ghosh, "Model-based Overlapping Clustering," *Proceedings of the 11th International Conference on Knowledge Discovery and Data Mining (KDD)*, 532-537, August 2005.
- C11. A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation," *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, 509-514, August 2004.
- C10. A. Banerjee and J. Langford, "An Objective Evaluation Criterion for Clustering," *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, 515-520, August 2004.
- C9. A. Banerjee, I. Dhillon, J. Ghosh, and S. Merugu, "An Information Theoretic Analysis of Maximum Likelihood Mixture Estimation for Exponential Families," *Proceedings of the 21st International Conference on Machine Learning (ICML)*, 57-64, July 2004.
- C8. A. Banerjee, X. Guo, and H. Wang, "Optimal Bregman Prediction and Jensen's Equality," *Proceedings of the International Symposium on Information Theory (ISIT)*, 169, June 2004.
- C7. A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, 234-245, April 2004.
- C6. S. Basu, A. Banerjee, and R. Mooney, "Active Semi-supervision for Pairwise Constrained Clustering," *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, 333-344, April 2004.
- C5. A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Generative Model-based Clustering of Directional Data," *Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining (KDD)*, 19-28, August 2003.
- C4. A. Banerjee, and J. Ghosh, "Competitive Learning Mechanisms for Scalable, Incremental and Balanced Clustering of Streaming Texts," *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, July 2003.
- C3. S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised Clustering by Seeding," *Proceedings of the 19th International Conference on Machine Learning (ICML)*, 19-26, July 2002.
- C2. A. Banerjee and J. Ghosh, "Frequency Sensitive Competitive Learning for Clustering on High-dimensional Hyperspheres," *Proceedings of the International Joint Conference on Neural Networks (IJCNN)*, 1590-1595, May 2002.
- C1. A. Banerjee and J. Ghosh, "On Scaling Up Balanced Clustering Algorithms," *Proceedings of the 2nd SIAM International Conference on Data Mining (SDM)*, 333-349, April 2002.

Refereed Book Chapters: (\* = PhD Student Advisee)

- B6. H. Shan\*, A. Agovic\*, and A. Banerjee, “Discriminative topic models,” in *Data Mining in Systems Health Management*, CRC Press, A. Srivastava and J. Han, editors, 2011, invited chapter, to appear.
- B5. A. Banerjee and H. Shan\*, “Model based Clustering,” in *Encyclopedia of Machine Learning*, Springer, C. Sammut and G. Webb, editors, 2010.
- B4. A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, “Text Clustering with Mixtures of von Mises-Fisher Distributions,” in *Text Mining: Theory, Applications, and Visualization*, Chapman & Hall/CRC Press, A. Srivastava and M. Sahami, editors, 2009.
- B3. A. Agovic\*, A. Banerjee, A. Ganguly, and V. Protopopescu, “Anomaly Detection in Transportation Corridors using Manifold Embedding,” in *Knowledge Discovery from Sensor Data*, CRC Press, O. Omitaomu and A. Ganguly, editors, 2009.
- B2. A. Banerjee and J. Ghosh, “Clustering with Balancing Constraints,” in *Constrained Clustering: Advances in Algorithms, Theory, and Applications*, S. Basu, I. Davidson, and K. L. Wagstaff, editors, CRC Press, 2008.
- B1. S. Basu, M. Bilenko, A. Banerjee, and R. Mooney, “Probabilistic Semi-supervised Clustering with Constraints,” in *Semi-supervised Learning*, O. Chapelle, B. Schölkopf, and A. Zien, editors, MIT Press, 2006.

Technical Reports: (\* = PhD Student Advisee)

- T15. P. Das\* and A. Banerjee, “Meta Algorithms for Portfolio Selection,” *Technical Report TR-10-022*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2010.
- T14. H. Wang\*, A. Banerjee, and D. Boley, “Modeling Time Varying Covariance Matrices in Low Dimensions,” *Technical Report TR-10-017*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2010.
- T13. N. Pathak\*, A. Banerjee, and J. Srivastava, “StochColor: Stochastic Coloring based Graph Partitioning,” *Technical Report TR-10-011*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2010.
- T12. V. Chandola\*, A. Banerjee, and V. Kumar, “Anomaly Detection for Discrete Sequences: A Survey,” *Technical Report TR-09-015*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2009.
- T11. A. Agovic\* and A. Banerjee, “A Unified View of Graph-based Semi-Supervised Learning: Label Propagation, Graph-Cuts, and Embeddings,” *Technical Report TR-09-012*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2009.
- T10. H. Shan\* and A. Banerjee, “Mixed-Membership Naive Bayes Models,” *Technical Report TR-09-002*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2009.
- T9. S. Sra, S. Jegelka, and A. Banerjee, “Approximation Algorithms for Bregman Clustering Co-clustering and Tensor Clustering,” *MPI Technical Report #177*, Max Planck Institute of Biological Cybernetics, 2008.
- T8. A. Banerjee and N. Srivastava, “Conditionally Positive Definite Kernels and Infinitely Divisible Distributions,” *Technical Report TR-08-034*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.
- T7. H. Wang, H. Shan\*, and A. Banerjee, “Bayesian Cluster Ensembles,” *Technical Report TR-08-028*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.

- T6. H. Shan\* and A. Banerjee, “Bayesian Co-clustering,” *Technical Report TR-08-022*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.
- T5. A. Banerjee and S. Basu, “A Social Query Model for Decentralized Search,” *Technical Report TR-08-017*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.
- T4. N. Pathak\*, C. Delong, K. Erickson, and A. Banerjee, “Social Topic Models for Community Extraction,” *Technical Report TR-08-005*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.
- T3. V. Chandola\*, A. Banerjee, and V. Kumar, “Anomaly Detection: A Survey,” *Technical Report TR 07-017*, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2007.
- T2. A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, “Clustering with Bregman Divergences,” *Technical Report TR-03-19*, Department of Computer Sciences, University of Texas at Austin, 2003.
- T1. A. Banerjee, I. Dhillon, J. Ghosh and S. Sra, “Expectation Maximization for Clustering on Hyperspheres,” *Technical Report TR-03-07*, Department of Computer Sciences, University of Texas at Austin, 2003.

Workshop & Other Publications: (\* = PhD Student Advisee, Name = Presenter)

- W15. R. Sivalingam, G. Somasundaram, A. Ragipindi, A. Banerjee, V. Morellas, N. Papanikolopoulos, and A. Truskinovsky, “Diagnosing Endometrial Carcinoma via Computer-Assisted Image Analysis,” *Annual Meeting of the United States & Canadian Academy of Pathology (USCAP)*, 2011.
- W14. A. Banerjee, D. Boley, and S. Acharyya, “Symmetrized Bregman Divergences and Metrics,” *Snowbird Learning Workshop*, 2009.
- W13. K. A. Monsen, M. J. Kerr, K. Abe, K. S. Martin, and A. Banerjee, “Use of computerized datasets and data mining methods to predict public health nurse home visiting client outcomes,” *World Academy of Nursing Science*, 2009.
- W12. K. A. Monsen, A. Banerjee, V. K. Ramadoss, P. Das\*, and K. Savik, “Discovering Effective Models for Home Visiting Practice,” *Midwest Nursing Research Society Annual Meeting*, 2009.
- W11. A. Banerjee and S. Basu, “A Social Query Model for Decentralized Search,” *2nd ACM International Workshop on Social Network Mining and Analysis (SNAKDD)*, August, 2008.
- W10. N. Pathak\*, C. Delong, K. Erickson, and A. Banerjee, “Social Topic Models for Community Extraction,” *2nd ACM International Workshop on Social Network Mining and Analysis (SNAKDD)*, August, 2008.
- W9. K. A. Monsen, A. Banerjee, B. Westra, and M. J. Kerr, “Using data mining methods with standardized terminology data sets for home visiting intervention effectiveness research,” *American Public Health Association Annual Meeting*, 2008.
- W8. A. Agovic\*, A. Banerjee, A. Ganguly, and V. Protopopescu, “Anomaly Detection in Transportation Corridors using Manifold Embedding,” *1st ACM International Workshop on Knowledge Discovery from Sensor Data (Sensor-KDD)*, August, 2007.
- W7. V. Monga, A. Banerjee, and B. Evans, “Clustering Algorithms for Perceptual Image Hashing,” *Proceedings of IEEE Digital Signal Processing Workshop*, August, 2004.
- W6. A. Banerjee, I. Dhillon, J. Ghosh, and S. Merugu, “Rate Distortion, Bregman Divergences and Maximum Likelihood Mixture Estimation,” *The Learning Workshop at Snowbird*, April, 2004.
- W5. A. Banerjee and J. Ghosh, “Mean Model Clustering,” *The Learning Workshop at Snowbird*, April, 2003.
- W4. A. Banerjee and J. Ghosh, “Characterizing Visitors to a Website Across Multiple Sessions,” *Proceedings of the National Science Foundation(NSF) Workshop on Next Generation Data Mining*, pp. 218-227, Nov 2002.

- W3. A. Banerjee and J. Ghosh, "Clickstream Clustering using Weighted Longest Common Subsequence," *Proceedings of the 1st SIAM International Conference on Data Mining (SDM): Workshop on Web Mining*, pp. 33-40, April 2001.
- W2. A. Banerjee and J. Ghosh, "Concept-based Clustering of Clickstream Data," In *Proceedings of the 3rd International Conference on Information Technology*, pp. 145-160, Dec 2000.
- W1. A. Banerjee, "Computerized Tumor Boundary Detection Using Genetic Algorithm," *Proceedings of the National Conference on Applications of Signal Processing*, Sept 1998.

## Presentations

### Invited Presentations:

- University of Notre Dame, 2011.
- University of Oregon, 2011.
- Columbia University, 2010.
- Macalester College, 2010.
- University of California at San Diego, 2010.
- MITRE, Workshop on Aviation Safety, 2009.
- Max Planck Institute (MPI) for Biological Cybernetics, Tübingen, Germany, 2008.
- Thomson-Reuters, 2008.
- Stanford University, Workshop on Algorithms for Modern Massive Datasets (MMDS), 2008.
- SRI International, 2006.
- Oak Ridge National Labs, 2006.
- University of Florida at Gainesville, 2005.
- StonyBrook University, 2005.
- IBM T. J. Watson Research Center, 2003 and 2004.
- Toyota Technological Institute, Chicago, 2003.

### Conference Presentations:

- H. Shan\* and A. Banerjee, “Generalized Probabilistic Matrix Factorizations for Collaborative Filtering,” *IEEE International Conference on Data Mining (ICDM)*, 2010.
- N. Pathak\*, A. Banerjee, and J. Srivastava, “A Generalized Linear threshold Model for Multiple Cascades,” *IEEE International Conference on Data Mining (ICDM)*, 2010.
- A. Banerjee, S. Basu, S. Merugu, “Multi-way Clustering on Relation Graphs,” *SIAM International Conference on Data Mining (SDM)*, Minneapolis, April, 2007.
- A. Banerjee, “An Analysis of Logistic Models: Exponential Family Connections and Online Performance,” *SIAM International Conference on Data Mining (SDM)*, Minneapolis, April, 2007.
- A. Banerjee and S. Basu, “Topic Models over Text Streams: A Study of Batch and Online Unsupervised Learning,” *SIAM International Conference on Data Mining (SDM)*, Minneapolis, April, 2007.
- A. Banerjee, “On Bayesian Bounds,” *Proceedings of the 23rd International Conference on Machine Learning (ICML)*, Pittsburgh, June, 2006.
- A. Banerjee, C. Krumpelman, S. Basu, R. Mooney, and J. Ghosh, “Model-based Overlapping Clustering,” *Proceedings of the 11th International Conference on Knowledge Discovery and Data Mining (KDD)*, Chicago, August, 2005.
- A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, “A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation,” *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, Seattle, August, 2004.

- A. Banerjee and J. Langford, “An Objective Evaluation Criterion for Clustering,” *Proceedings of the 10th International Conference on Knowledge Discovery and Data Mining (KDD)*, Seattle, August 2004.
- A. Banerjee, X. Guo, and H. Wang, “Optimal Bregman Prediction and Jensen’s Equality,” *Proceedings of the International Symposium on Information Theory (ISIT)*, Chicago, June, 2004.
- A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, “Clustering with Bregman Divergences,” *Proceedings of the 4th SIAM International Conference on Data Mining (SDM)*, Orlando, April, 2004.
- A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, “Generative Model-based Clustering of Directional Data,” *Proceedings of the 9th International Conference on Knowledge Discovery and Data Mining (KDD)*, Washington DC, August 2003.

#### Tutorials:

- A. Banerjee, “Introduction to Graphical Models for Data Mining,” ACM Conference on Knowledge Discovery and Data Mining (KDD), July 2010, Washington DC.
- A. Banerjee, V. Chandola, A. Lazarevic, V. Kumar, and J. Srivastava, “Data Mining for Anomaly Detection,” European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML PKDD), September 2008, Antwerp, Belgium.
- A. Banerjee, V. Chandola, A. Lazarevic, V. Kumar, and J. Srivastava, “Anomaly Detection: A Tutorial,” SIAM Data Mining Conference (SDM), April 2008, Atlanta, GA.

#### **Other Key Activities and Accomplishments**

- Project with NASA, developed and delivered text classification and topic modeling package for the analysis of aviation safety reports, 2010.
- Project with Oak Ridge National Labs, developed and delivered data mining package for anomaly detection in truck weigh-station data, 2006.
- Member of startup Neonyoyo Inc., developed recommendation systems for wireless and internet applications. The company was formed and later acquired by Interwoven Inc. in 2000.

## TEACHING AND CURRICULUM DEVELOPMENT

### University of Minnesota

<i>Semester</i>	<i>Year</i>	<i>Course</i>
Spring	2010	CSci 5512: Artificial Intelligence II
Fall	2009	CSci 4041: Algorithms and Data Structures
Spring	2009	CSci 4041: Algorithms and Data Structures
Fall	2008	CSci 5525: Machine Learning
Spring	2008	CSci 5512W: Artificial Intelligence II
Fall	2007	CSci 8980: Advanced Topics in Graphical Models
Spring	2007	CSci 5512W: Artificial Intelligence II
Fall	2006	CSci 5525: Machine Learning
Spring	2006	CSci 8980: Topics in Machine Learning

### Curriculum Development

- CSci 5525: Machine Learning, first offered Fall 2006  
Machine Learning is one of the most active areas in Computer Science. Almost all research universities have one or more graduate courses on the topic. The course was designed to introduce graduate students to the major developments in the field over the past few decades. The course is offered every alternate year in Fall.
- CSci 8980: Advanced Topics in Graphical Models, offered Fall 2007  
Graphical models constitute one of the most active areas of research in Machine Learning. The course was designed for advanced graduate students doing research in Machine Learning, and applying such ideas to related domains such as robotics and statistical natural language processing.
- CSci 8980: Topics in Machine Learning, offered Spring 2006  
The course was designed for graduate students with some background in machine learning, statistics and data analysis. The course covered important recent advances in machine learning in the context of online learning, and their connections to advances in game theory.

### Faculty Development Activities regarding teaching

- Participated in the Early Career Faculty Learning Community to Develop and Enhance Teaching skills.

## ADVISING AND MENTORING

### Undergraduate Student Activities

- Garrison Kubis, UROP, Fall, 2011.
- Tinghui Zhou, UROP, Fall, 2010.
- Kim Kawatra, UROP, Spring, 2010.
- Charles Curtsinger, UROP, Fall, 2007.

### Graduate Student Activities

#### Master's Thesis Directed

- Subrahmanya Bhat, MS, 2010.  
Title: Probabilistic Graph Partitioning for Topic Modeling in Text Streams
- Roman Briskine, MS, 2008.  
Title: Clustering based Meta-prediction of Phosphorylation Sites

#### Doctoral Dissertations Directed

- Amrudin Agovic, co-advised with Maria Gini, PhD, 2011.  
Title: Predictive Modeling using Dimensionality Reduction and Dependency Structures  
First Appointment: CEO, Reliancy
- Varun Chandola, co-advised with Vipin Kumar, PhD, 2009.  
Title: Anomaly Detection for Symbolic Sequences and Times Series Data  
First Appointment: Oak Ridge National Labs (ORNL)

#### Doctoral Students Advised (Current)

- Hanhuai Shan  
Status: Passed Oral Prelims Spring'08
- Qiang Fu  
Status: Passed Oral Prelims Spring'09
- Nishith Pathak, co-advised with Jaideep Srivastava  
Status: Passed Oral Prelims Fall'10
- Puja Das  
Status: Passed WPE Spring'10
- Huahua Wang  
Status: Third Year PhD Student
- Soumyadeep Chatterjee  
Status: Third Year PhD Student
- Amir Asiaee Taheri  
Status: Second Year PhD Student

## **SERVICE AND PUBLIC OUTREACH**

### **Service To The Discipline/Profession/Interdisciplinary Areas**

#### *Journal Reviewer Experience*

- Journal of Machine Learning Research
- Data Mining and Knowledge Discovery
- IEEE Transactions on Information Theory
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Transactions on Neural Networks
- IEEE Transactions on Knowledge and Data Engineering
- IEEE Transactions on System, Man and Cybernetics
- IEEE Transactions on Robotics
- ACM Transactions on Knowledge Discovery from Data
- Machine Learning Journal
- Statistical Analysis and Data Mining
- Geoinformatica
- Communications of the ACM
- Neurocomputing
- Applied Intelligence

#### *Review panels for external funding agencies, foundations, etc.*

- National Science Foundation (NSF) Panels, 2009, 2010, 2011.
- Israeli Science Foundation (ISF) Reviewer, 2010.
- NASA Intelligent Systems Reviewer, 2004.

#### *Organization of conferences, workshops, panels, symposia*

##### Chair/Co-Chair:

- Workshop Co-chair, IMA Workshop, Machine Learning: Theory and Computation, 2012.
- Workshop Co-chair, ICML Workshop on “Machine Learning for Global Challenges,” 2011.
- Workshop Co-chair, SIAM Conference on Data Mining (SDM), 2011.
- Technical Chair, NASA Conference on Intelligent Data Understanding (CIDU), 2008.

Senior Program Committee:

- SIAM Conference on Data Mining (SDM), 2012.
- IEEE International Conference on Data Mining (ICDM), 2011.
- ACM International Conference on Knowledge Discovery and Data Mining (KDD), 2011.
- SIAM Conference on Data Mining (SDM), 2010.

Program Committee:

- International Conference on Machine Learning (ICML'07, ICML'08, ICML'09, ICML'10, ICML'11).
- IEEE International Conference on Data Mining (ICDM'05, ICDM'07, ICDM'09, ICDM'10).
- ACM International Conference on Knowledge Discovery and Data Mining (KDD'08, KDD'09).
- Advances in Neural Information Processing Systems (NIPS'06, NIPS'07, NIPS'08, NIPS'09, NIPS'10).
- SIAM Conference on Data Mining (SDM'06, SDM'07, SDM'10).
- National Conference on Artificial Intelligence (AAAI'06, AAAI'07).
- International Conference on Artificial Intelligence and Statistics (AISTATS'09).

Other Activities:

- Book Reviewing: “Text Mining: Predictive Methods for Analyzing Unstructured Information” by Sholom Weiss, Nitin Indurkha, Tong Zhang, Fred Damerau, published October 2004 by Springer-Verlag.
- Invited panelist on “Text mining: The discipline that never was” organized by Prabhakar Raghavan, head of Yahoo! Research, at the 11th ACM International Conference on Knowledge Discovery and Data Mining (KDD'05).

**Service To The University/College/Department***Department/Unit Service:*

2009-10 Graduate Admissions Committee  
 2008-09 Graduate Admissions Committee  
 2007-08 Graduate Admissions Committee  
 2006-07 Strategic Planning Committee  
 2005-06 Graduate Admissions Committee