Arindam Banerjee

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

Education

Ph.D., University of Texas at Austin — August 2005
Dept of Electrical and Computer Engineering

M.Tech., Indian Institute of Technology, Kanpur — May 1999
Dept of Electrical Engineering

B.E., Jadavpur University — May 1997
Dept of Electronics and Telecommunication Engineering

Positions/Employment

University of Minnesota, Twin Cities	Associate Professor	2011–Present
University of Minnesota, Twin Cities	Assistant Professor	2005–2011

Current Membership in Professional Organizations

Association for Computing Machinery (ACM)
Institute of Electrical and Electronics Engineers (IEEE)
Society for Industrial and Applied Mathematics (SIAM)

Visiting Professorships or Visiting Scholar Positions

- Visiting Professor, University of Campinas (Unicamp), Campinas, Sao Paolo, Brazil, 2012.
- Long Term Visitor, Institute for Mathematics and its Applications (IMA), University of Minnesota, Twin Cities, 2011-12.
- Visiting Professor, Max Planck Institute (MPI) for Biological Cybernetics, Tubingen, Germany, 2008.
- Visitor, Institute of Pure and Applied Mathematics (IPAM), University of California, Los Angeles (UCLA), 2007.
- Visiting Fellow, Institute for Computational Engineering and Sciences (ICES), University of Texas at Austin, 2006.

HONORS and AWARDS

- Adobe Research Award, 2016.
- Best Student Paper Award, Honorable Mention, AI & Statistics (AISTATS) Conference, 2015.
- IBM Faculty Award, 2013.
- Yahoo Faculty Research and Engagement Program Award, 2013.
- Best Paper Award, Big Data Mining Workshop: ACM SIGKDD Conference on Knowledge Discovery and Data Mining, 2013.
- Best Application Paper Award, SIAM International Conference on Data Mining (SDM), 2013.
- Best Student Paper Award, SIAM International Conference on Data Mining (SDM), 2012.
- Invited participant, White House OSTP "Big Data" event, AAAS, March 2012.
- NSF CAREER Award, 2010-2015.
- Resident Fellow, Institute on the Environment (IonE), 2010 onwards.
- McKnight Land-Grant Professorship, University of Minnesota, 2009-11.
- Best of SDM Award, SIAM International Conference on Data Mining (SDM), 2007.
- Best Student Paper Runner-up, ACM International Workshop on Knowledge Discovery from Sensor Data (SensorKDD), 2007.
- J. T. Oden Faculty Research Fellowship, Visiting Fellow at the 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), 2004.
- IBM PhD fellowship for the academic years 2003-2004 and 2004-2005.
- 1^{st} 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 (INMO) and attended the International Mathematical Olympiads (IMO) camp for two successive years 1992, 1993.

RESEARCH

External Grants and Contracts

- III: Medium: Collaborative Research: Bayesian Modeling and Inference for Quantifying Terrestrial Ecosystem Functions
  National Science Foundation (NSF)
  PI, with Peter Reich (UMN), Sudipto Banerjee (UCLA)
  September 1, 2016 - August 31, 2019
  Total Amount: $1,085,987, UMN Amount: $724,000, UMN Direct Costs: $473,148.

- Fingerprints of AMOC Variations Derived From Machine Learning Methods
  National Oceanic and Atmospheric Administration (NOAA)
  Co-PI, with Timothy DelSole (GMU) and Barry Klinger (GMU)
  July 1, 2016 - June 30, 2019
  Total Amount: $308,980, UMN Amount: $21,300, UMN Direct Costs: $14,014.

- BIGDATA: F: DKA: Collaborative Research: High-Dimensional Statistical Machine Learning for Spatio-Temporal Climate Data
  National Science Foundation (NSF)
  PI, with Pradeep Ravikumar (CMU), Auroop Ganguly (NEU)
  September 1, 2014 - August 31, 2017
  Total Amount: $1,071,000, UMN Amount: $357,000, UMN Direct Costs: $254,148.

- RI: Small: Finding Patterns in Complex Data with Probablistic Graphical Models
  National Science Foundation (NSF)
  PI
  August 1, 2014 - July 31, 2017
  Total Amount: $449,991, UMN Amount: $449,991, UMN Direct Costs: $333,558.

- EAGER: Collaborative Research: Learning Relations between Extreme Weather Events and Planet-Wide Environmental Trends
  National Science Foundation (NSF)
  Co-PI, with Claire Monteleoni (PI, GWU), Tim Delsole (COLA/GMU)
  September 1, 2014 - August 31, 2016
  Total Amount: $300,000, UMN Amount: $100,000, UMN Direct Costs: $76,139.

- Global land model development: Time to shift from a plant functional type to a plant functional trait approach
  Department of Energy (DOE)
  Co-PI, with Peter Reich (PI)
  August 15, 2014 - August 14, 2017
  Total Amount: $1,347,518, UMN Amount: $1,347,518, UMN Direct Costs: $886,525.

- TWC: Medium: Collaborative Research: HIMALAYAS: Hierarchical Machine Learning Stack for Fine-Grained Analysis of Malware Domain Groups
  National Science Foundation (NSF)
  Co-PI, with Vinod Yegneswaran (PI, SRI), Shalini Ghosh (SRI), Guofei Gu (Texas A& M)
  September 1, 2013 - August 31, 2016
  Total Amount: $1,098,640, UMN Amount: $252,545, UMN Direct Costs: $182,840.

- MRI: Development of an Instrument that Monitors Behaviors Associated with Obsessive-Compulsive Behaviors and Schizophrenia
  National Science Foundation (NSF)
  Co-PI, with Nikolaos Papanikolopoulos (PI), and others

September 1, 2013 - August 31, 2017
Total Amount: $3,698,265, UMN Amount: $3,698,265, UMN Direct Costs: $2,896,200.

- Automated Detection of Precursors to Human-Automation Interaction Based Aviation Safety Incidents
  National Aeronautics and Space Agency (NASA)
  PI, with Vipin Kumar and Jaideep Srivastava (co-PIs)
  October 1, 2012 - September 30, 2015
  Total Amount: $447,090, UMN Amount: $447,090, UMN Direct Costs: $324,271.

- 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, UMN Amount: $6,400,000, UMN Direct Costs: $4,914,277.

- 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: $1,578,897, UMN Amount: $1,578,897.[1]

- 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, UMN Amount: $500,000, UMN Direct Costs: $376,267.

- CAREER: Combinatorial Online Learning and its Applications
  National Science Foundation (NSF)
  PI
  April 1, 2010 - March 31, 2015
  Total Amount: $495,801, UMN Amount: $495,801, UMN Direct Costs: $359,593.

- 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: $455,000, UMN Amount: $455,000, UMN Direct Costs: $332,018.

- 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, UMN Amount: $399,609, UMN Direct Costs: $290,203.

- 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.[1]

- 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, UMN Amount: $199,999, UMN Direct Costs: $189,573.

[1] Exact UMN (Direct Cost) Amount is not available.

- Discovering Effective Models for Home Visiting Practice
  Midwest Nursing Research Society
  Co-I, with Karen Monsen (PI)
  May 1, 2008 - April 30, 2009
  Total Amount: $9,824, UMN Amount: $9,824.[1]
- 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: $118,184, UMN Amount: $118,184, UMN Direct Costs: $87,125.

Internal Grants

- Computer Aided Cancer Diagnosis in Surgical Pathology
  Co-PI, with Nikolaos Papanikolopoulous (PI), and others
  Institute for Engineering in Medicine (IEM) Seed Grant
  January 1, 2012 - December 31, 2012
  Total Amount: $39,508.
- 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 Publications:

1. A. Gonçalves, F. J. Von Zuben, and A. Banerjee, "Multi-task Sparse Structure Learning with Gaussian Copula Models," Journal of Machine Learning Research, 17(33):1-30, 2016.
2. I. Melnyk, B. Matthews, H. Valizadegan, A. Banerjee, and Nikunj Oza, "Vector Autoregressive Model-based Anomaly Detection in Aviation Systems," Journal of Aerospace Information Systems, 13(4): 161-173, 2016.
3. F. Schrodt, J. Kattge, H. Shan, F. Fazayeli, A. Karpatne, A. Banerjee, M. Reichstein, M. Boenisch, S. Daz, J. Dickie, A. Gillison, V. Kumar, S. Lavorel, P.W. Leadley, C. Wirth, I. Wright, S.J. Wright, P.B. Reich, "HPMF a hierarchical Bayesian approach to gap-filling and trait prediction," Global Ecology and Biogeography, 24, 1510-1521, 2015.
4. A.R. Ganguly, E.A. Kodra, A. Agrawal, A. Banerjee, S. Boriah, Sn. Chatterjee, So. Chatterjee, A. Choudhary, D. Das, J. Faghmous, P. Ganguli, S. Ghosh, K. Hayhoe, C. Hays, W. Hendrix, Q. Fu, J. Kawale, D. Kumar, V. Kumar, W. Liao, S. Liess, R. Mawalagedara, V. Mithal, R. Oglesby, K. Salvi, P. K. Snyder, K. Steinhaeuser, D. Wang, and D. Wuebbles. "Toward enhanced understanding and projections of climate extremes using physics-guided data mining techniques," Nonlinear Processes in Geophysics, 21, 777-795, 2014.

5. A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos, "Jensen-Bregman LogDet Divergence with Application to Efficient Similarity Search for Covariance Matrices," IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(9): 2161-2174, 2013.
6. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection for Discrete Sequences: A Survey," IEEE Transactions on Knowledge and Data Engineering, 24(5): 823-839, 2012.
7. H. Wang, H. Shan, and A. Banerjee, "Bayesian Cluster Ensembles," Statistical Analysis and Data Mining, 4(1), 54-70, 2011.
8. H. Shan and A. Banerjee, "Mixed-Membership Naive Bayes Models," Data Mining and Knowledge Discovery, 23(1), 1-62, 2011.
9. 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.
10. 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.
11. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Survey," ACM Computing Surveys, 41(3), Article 15, 2009.
12. 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.
13. 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.
14. 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.
15. A. Banerjee and J. Ghosh, "Scalable Clustering Algorithms with Balancing Constraints," Data Mining and Knowledge Discovery, 13(3), 365-395, November, 2006.
16. 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.
17. 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.
18. A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," Journal of Machine Learning Research, 6 (Oct), 1705-1749, 2005.
19. 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.
20. 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 Publications:

1. Q. Gu, A. Banerjee, "High Dimensional Structured Superposition Models," Advances in Neural Information Processing Systems (NIPS), 2016. (Acceptance rate: 22.72%)
2. S. Chen and A. Banerjee, "Structured Matrix Recovery via the Generalized Dantzig Selector," Advances in Neural Information Processing Systems (NIPS), 2016. (Acceptance rate: 22.72%)

3. I. Melnyk, A. Banerjee, B. Matthews, and N. Oza, "Semi-Markov Switching Vector Autoregressive Model-based Anomaly Detection in Aviation Systems," International Conference on Knowledge Discovery and Data Mining (KDD), 2016. (Acceptance Rate: 18.11%)
4. F. Fazayeli and A. Banerjee, "Generalized Direct Change Estimation in Ising Model Structure," International Conference on Machine Learning (ICML), 2016. (Acceptance Rate: 24.27%)
5. I. Melnyk and A. Banerjee, "Estimating Structured Vector Autoregressive Model," International Conference on Machine Learning (ICML), 2016. (Acceptance Rate: 24.27%)
6. F. Fazayeli and A. Banerjee, "The Matrix Generalized Inverse Gaussian Distribution: Properties and Applications," European Conference on Machine Learning and Principles and Practice of Knowledge Discovery(ECML-PKDD), 2016. (Acceptance Rate: 28%)
7. A. Taheri, S. Chatterjee, and A. Banerjee, "High Dimensional Structured Estimation with Noisy Designs," SIAM International Conference on Data Mining (SDM), 2016. (Acceptance Rate: 25.88%)
8. S. Chatterjee, S. Liess, A. Banerjee, and V. Kumar, "Understanding Dominant Factors for Precipitation over the Great Lakes Region," AAAI Conference on Artificial Intelligence (AAAI), 2016. (Acceptance Rate: 25.75%)
9. S. Chen and A. Banerjee, "Structured Estimation with Atomic Norms: General Bounds and Applications," Advances in Neural Information Processing Systems (NIPS), 2015. (Acceptance Rate: 21.93%)
10. V. Sivakumar, A. Banerjee, and P. Ravikumar, "Beyond Sub-Gaussian Measurements: High-Dimensional Structured Estimation with Sub-Exponential Designs," Advances in Neural Information Processing Systems (NIPS), 2015. (Acceptance Rate: 21.93%)
11. S. Gunasekar, A. Banerjee, and J. Ghosh, "Unified View of Matrix Completion under General Structural Constraints," Advances in Neural Information Processing Systems (NIPS), 2015. (Acceptance Rate: 21.93%)
12. M. Kadkhodaie, K. Christakopoulou, M. Sanjabi, and A. Banerjee, "Accelerated Alternating Direction Method of Multipliers," International Conference on Knowledge Discovery and Data Mining (KDD), 2015. (Acceptance Rate: 19.41%)
13. N. Johnson and A. Banerjee, "Structured Hedging for Resource Allocations with Leverage," International Conference on Knowledge Discovery and Data Mining (KDD), 2015. (Acceptance Rate: 19.41%)
14. G. Golnari, A. Taheri, A. Banerjee, and Z.-L. Zhang, "Revisiting Non-Progressive Influence Models: Scalable Influence Maximization in Social Networks," Conference on Uncertainty in Artificial Intelligence (UAI), 2015. (Acceptance Rate: 34.25%)
15. A. Gonçalves, F. Von Zuben, and A. Banerjee, "Muti-label structure learning with Ising model selection," International Joint Conference on Artificial Intelligence (IJCAI), 2015. (Acceptance Rate: 28.81%)
16. K. Christakopoulou and A. Banerjee, "Collaborative Ranking with a Push at the Top," International World Wide Web Conference (WWW), 2015. (Acceptance Rate: 14.1%)
17. I. Melnyk and A. Banerjee, "A Spectral Algorithm for Inference in Hidden Semi-Markov Models," International Conference on Artificial Intelligence and Statistics (AISTATS), 2015. (Acceptance Rate: 28.73%)
18. S. Chen and A. Banerjee, "One-bit Compressed Sensing with the k-Support Norm," International Conference on Artificial Intelligence and Statistics (AISTATS), 2015. (Acceptance Rate: 28.73%)
19. N. Johnson and A. Banerjee, "Online Resource Allocation with Structured Diversification," SIAM International Conference on Data Mining (SDM), 2015. (Acceptance Rate: 14.7%)
20. K. Subbian, A. Banerjee, and S. Basu, "Link Prediction Using Multiple Sources," SIAM International Conference on Data Mining (SDM), 2015. (Acceptance Rate: 14.7%)
21. C. Jin, Q. Fu, W. Hendrix, Z. Chen, A. Agrawal, W. Liao, A. Choudhary, H. Wang, and A. Banerjee, "Running MAP Inference on Million Node Graphical Models: A High Performance Computing Perspective," IEEE/ACM International Symposium on Cluster, Cloud and Grid Computing (CCGrid), 2015. (Acceptance Rate: 25.7%)

22. A. Banerjee, S. Chen, F. Fazayeli, and V. Sivakumar, "Estimation with Norm Regularization," Advances in Neural Information Processing Systems (NIPS), 2014. (Acceptance Rate: 24.67%)

23. S. Chatterjee, S. Chen, and A. Banerjee, "Generalized Dantzig Selector: Application to the k-support norm," Advances in Neural Information Processing Systems (NIPS), 2014. (Acceptance Rate: 24.67%)

24. H. Wang, A. Banerjee, and Z.-Q. Luo, "Parallel Direction Method of Multipliers," Advances in Neural Information Processing Systems (NIPS), 2014. (Acceptance Rate: 24.67%)

25. H. Wang and A. Banerjee, "Bregman Alternating Direction Method of Multipliers," Advances in Neural Information Processing Systems (NIPS), 2014. (Acceptance Rate: 24.67%)

26. A. R. Gonçalves, P. Das, S. Chatterjee, V. Sivakumar, F. J. Von Zuben, A. Banerjee, "Multi-task Sparse Structure Learning," International Conference on Information and Knowledge Management (CIKM), 2014. (Acceptance Rate: 20.79%)

27. P. Das, N. Johnson, and A. Banerjee, "Online Portfolio Selection with Group Sparsity," AAAI Conference on Artificial Intelligence (AAAI), 2014. (Acceptance Rate: 28.31%)

28. H. Wang, F. Fazayeli, S. Chatterjee, and A. Banerjee, "Gaussian Copula Precision Estimation with Missing Values," International Conference on Artificial Intelligence and Statistics (AISTATS), 2014. (Acceptance Rate: 35.8%)

29. H. Wang, A. Banerjee, C. Hsieh, P. Ravikumar, and I. Dhillon, "Large Scale Distributed Sparse Precision Estimation," Advances in Neural Information Processing Systems (NIPS), 2013. (Acceptance Rate: 25.4%)

30. Q. Fu, H. Wang, and A. Banerjee, "Bethe-ADMM for Tree Decomposition based Parallel MAP inference," Conference on Uncertainty in Artificial Intelligence (UAI), 2013. (Acceptance Rate: 31%)

31. P. Das, N. Johnson, and A. Banerjee, "Online Lazy Updates for Portfolio Selection with Transaction Costs," AAAI Conference on Artificial Intelligence (AAAI), 2013. (Acceptance Rate: 29%)

32. S. Acharyya, A. Banerjee, and D. Boley, "Bregman Divergences and Triangle Inequality," SIAM International Conference on Data Mining (SDM), 2013. (Acceptance Rate: 25.5%)

33. K. Subbian and A. Banerjee, "Climate Multi-model Regression Using Spatial Smoothing," SIAM International Conference on Data Mining (SDM), 2013. (Acceptance Rate: 25.5%)

34. S. Kasiviswanathan, H. Wang, A. Banerjee, P. Melville, "Online L1-Dictionary Learning with Application to Novel Document Detection," Advances in Neural Information Processing Systems (NIPS), 2012. (Acceptance Rate: 25.2%)

35. C. Hsieh, I. Dhillon, P. Ravikumar, A. Banerjee, "A Divide-and-Conquer Method for Sparse Inverse Covariance Estimation," Advances in Neural Information Processing Systems (NIPS), 2012. (Acceptance Rate: 25.2%)

36. A. Taheri, M. Tepper, A. Banerjee, and G. Sapiro, "If You are Happy and You Know It ... Tweet," ACM Conference on Information and Knowledge Management (CIKM), 2012. (Acceptance Rate: 27.8%)

37. H. Wang and A. Banerjee, "Online Alternating Direction Method," International Conference on Machine Learning (ICML), 2012. (Acceptance Rate: 27.2%)

38. H. Shan, J. Kattge, P. B. Reich, A. Banerjee, F. Schrodt, and M. Reichstein, "Gap Filling in the Plant Kingdom—Trait Prediction Using Hierarchical Probabilistic Matrix Factorization," International Conference on Machine Learning (ICML), 2012. (Acceptance Rate: 27.2%)

39. T. Zhou, H. Shan, A. Banerjee, and G. Sapiro, "Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information," SIAM International Conference on Data Mining (SDM), 2012. (Acceptance Rate: 15%)

40. S. Chatterjee, K. Steinhaeuser, A. Banerjee, S. Chatterjee, and A. Ganguly, "Sparse Group Lasso: Consistency and Climate Applications," SIAM International Conference on Data Mining (SDM), 2012. (Acceptance Rate: 15%)

41. Q. Fu, A. Banerjee, S. Liess, and P. Snyder, "Drought Detection for the Last Century: A MRF-based Approach," SIAM International Conference on Data Mining (SDM), 2012. (Acceptance Rate: 15%)

42. S. Kasiviswanathan, P. Melville, A. Banerjee, and V. Sindhwani, "Emerging Topic Detection using Dictionary Learning," ACM Conference on Information and Knowledge Management (CIKM), 2011. (Acceptance Rate: 35%)

43. 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. (Acceptance Rate: 23.7%)

44. 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. (Acceptance Rate: 17.6%)

45. P. Das and A. Banerjee, "Meta Optimization and its Application to Portfolio Selection," ACM SIGKKD International Conference on Knowledge Discovery and Data Mining (KDD), 2011. (Acceptance Rate: 17.6%)

46. A. Agovic, A. Banerjee, and S. Chatterjee, "Probabilistic Matrix Addition," International Conference on Machine Learning (ICML), 2011. (Acceptance Rate: 25.8%)

47. H. Shan and A. Banerjee, "Generalized Probabilistic Matrix Factorizations for Collaborative Filtering," IEEE International Conference on Data Mining (ICDM), 2010. (Acceptance Rate: 19.45%)

48. N. Pathak, A. Banerjee, and J. Srivastava, "A Generalized Linear threshold Model for Multiple Cascades," IEEE International Conference on Data Mining (ICDM), 2010. (Acceptance Rate: 19.45%)

49. A. Agovic and A. Banerjee, "Gaussian Process Topic Models," Conference on Uncertainty in Artificial Intelligence (UAI), 2010. (Acceptance Rate: 33.8%)

50. H. Shan and A. Banerjee, "Residual Bayesian Co-clustering for Matrix Approximation," SIAM International Conference on Data Mining (SDM), 2010. (Acceptance Rate: 23.36%)

51. 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. (Acceptance Rate: 48.8%)

52. H. Shan, A. Banerjee, and N. Oza, "Discriminative Mixed-membership Models," IEEE International Conference on Data Mining (ICDM), 2009. (Acceptance Rate: 17.68%)

53. Q. Fu and A. Banerjee, "Bayesian Overlapping Subspace Clustering," IEEE International Conference on Data Mining (ICDM), 2009. (Acceptance Rate: 17.68%)

54. S. Jegelka, S. Sra, and A. Banerjee, "Approximation Algorithms for Tensor Clustering," The 20th International Conference on Algorithmic Learning Theory (ALT), 2009. (Acceptance Rate: 43.33%)

55. H. Wang, H. Shan, and A. Banerjee, "Bayesian Cluster Ensembles," SIAM International Conference on Data Mining (SDM), 2009. (Acceptance Rate: not known, usually 12-18%)

56. 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. (Acceptance Rate: 43%)

57. H. Shan and A. Banerjee, "Bayesian Co-clustering," IEEE International Conference on Data Mining (ICDM), 2008. (Acceptance Rate: 19.9%)

58. Q. Fu and A. Banerjee, "Multiplicative Mixture Models for Overlapping Clustering," IEEE International Conference on Data Mining (ICDM), 2008. (Acceptance Rate: 19.9%)

59. K. Hsu, A. Banerjee, and J. Srivastava, "I/O Scalable Bregman Co-clustering," Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD), 2008. (Acceptance Rate: 12%)

60. A. Banerjee and H. Shan, "Latent Dirichlet Conditional Naive Bayes Models," IEEE International Conference on Data Mining (ICDM), 2007. (Acceptance Rate: 19.1%)

61. A. Banerjee, S. Basu, S. Merugu, "Multi-way Clustering on Relation Graphs," SIAM International Conference on Data Mining (SDM), 2007. 2007.(Acceptance Rate: 12%)

62. A. Banerjee, "An Analysis of Logistic Models: Exponential Family Connections and Online Performance," SIAM International Conference on Data Mining (SDM), 2007. 2007.(Acceptance Rate: 12%)

63. 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. 2007.(Acceptance Rate: 12%)

64. A. Banerjee, "On Bayesian Bounds," International Conference on Machine Learning (ICML), 2006. (Acceptance Rate: 20%)

65. A. Banerjee, C. Krumpelman, S. Basu, R. Mooney, and J. Ghosh, "Model-based Overlapping Clustering," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005. (Acceptance Rate: 21.28%)

66. A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2004. (Acceptance Rate: not known, 20-24%)

67. A. Banerjee and J. Langford, "An Objective Evaluation Criterion for Clustering," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2004. (Acceptance Rate: not known, 20-24%)

68. A. Banerjee, I. Dhillon, J. Ghosh, and S. Merugu, "An Information Theoretic Analysis of Maximum Likelihood Mixture Estimation for Exponential Families," International Conference on Machine Learning (ICML), 2004. (Acceptance Rate: 32%)

69. A. Banerjee, X. Guo, and H. Wang, "Optimal Bregman Prediction and Jensens Equality," International Symposium on Information Theory (ISIT), 2004. (Acceptance Rate: not known, $\approx 45\%$)

70. A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," SIAM International Conference on Data Mining (SDM), 2004. (Acceptance Rate: 14.3%)

71. S. Basu, A. Banerjee, and R. Mooney, "Active Semi-supervision for Pairwise Constrained Clustering," SIAM International Conference on Data Mining (SDM), 333-344, April 2004. (Acceptance Rate: 14.3%)

72. A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Generative Model-based Clustering of Directional Data," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2003. (Acceptance Rate: 13.2%)

73. A. Banerjee, and J. Ghosh, "Competitive Learning Mechanisms for Scalable, Incremental and Balanced Clustering of Streaming Texts," International Joint Conference on Neural Networks (IJCNN), 2003. (Acceptance Rate: 33.6%)

74. S. Basu, A. Banerjee, and R. Mooney, "Semi-supervised Clustering by Seeding," International Conference on Machine Learning (ICML), 2002. (Acceptance Rate: 33%)

75. A. Banerjee and J. Ghosh, "Frequency Sensitive Competitive Learning for Clustering on High-dimensional Hyperspheres," International Joint Conference on Neural Networks (IJCNN), 2002. (Acceptance Rate: not known, $\approx$ 35%)

76. A. Banerjee and J. Ghosh, "On Scaling Up Balanced Clustering Algorithms," SIAM International Conference on Data Mining (SDM), 2002. (Acceptance Rate: 13.6%)

Book Chapters:

1. A. R. Gonçalves, S. Chatterjee, V. Sivakumar, and A. Banerjee, "Structured Estimation in High Dimensions: Applications in Climate," in Large-Scale Machine Learning in the Earth Sciences, A. Srivastava and K. Steinhaeuser, editors, CRC Press, accepted for publication, 2016.

2. V. Chandola, A. Banerjee, and V. Kumar, "Anomaly Detection: A Modern Perspective," in Encyclopedia of Machine Learning and Data Mining, C. Sammut and G. I. Webb, editors, Springer, accepted for publication, 2016.

3. C. Monteleoni, G. A. Schmidt, F. Alexander, A. Niculescu-Mizil, K. Steinhaeuser, M. Tippett, A. Banerjee, M. B. Blumenthal, A. R. Ganguly, J. E. Smerdon, M. Tedesco, "Climate Informatics," in Computational Intelligent Data Analysis for Sustainable Development, T. Yu, N. Chawla and S. Simoff, editors, Chapman & Hall/CRC Press, 2013.
4. A. R. Ganguly, E. Kodra, S. Chatterjee, A. Banerjee, and H. N. Najm, "Computational data sciences for actionable insights on climate extremes and uncertainty," in Computational Intelligent Data Analysis for Sustainable Development, T. Yu, N. Chawla and S. Simoff, editors, CRC Press, 2013.
5. A. Agovic and A. Banerjee, "Semi-supervised Clustering," in Data Clustering: Algorithms and Applications, C. C. Aggarwal and C. K. Reddy, editors, Chapman & Hall/CRC press, 2013.
6. H. Shan and A. Banerjee, "Discriminative Mixed-membership Models," in Handbook of Mixed Membership Models, E. Airoldi, D. Blei, E. Erosheva, S. E. Fienberg, editors, Chapman & Hall/CRC Press, 2012.
7. H. Shan, A. Agovic, and A. Banerjee, "Discriminative topic models," in Machine Learning and Knowledge Discovery for Engineering Systems Health Management, A. Srivastava and J. Han, editors, Chapman & Hall/CRC Press, 2011.
8. A. Banerjee and H. Shan, "Model based Clustering," in Encyclopedia of Machine Learning, C. Sammut and G. Webb, editors, Springer, 2010.
9. 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, A. Srivastava and M. Sahami, editors, Chapman & Hall/CRC Press, 2009.
10. A. Agovic, A. Banerjee, A. Ganguly, and V. Protopopescu, "Anomaly Detection in Transportation Corridors using Manifold Embedding," in Knowledge Discovery from Sensor Data, O. Omitaomu and A. Ganguly, editors, CRC Press, 2009.
11. 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.
12. S. Basu, M. Bilenko, A. Banerjee, and R. Mooney, "Probabilistic Semi-supervised Clustering with Constraints," in Semi-supervised Learning, O. Chapelle, B. Scholkopf, and A. Zien, editors, MIT Press, 2006.

Workshops, Magazine Articles & Other Publications (lightly reviewed):

1. X. Liu, P. Cao, D. Zhao, and A. Banerjee, "Multi-task Spare Group Lasso for Characterizing Alzheimer's Disease," Workshop on Data Mining for Medicine and Healthcare (DMMH), in conjunction with SIAM International Conference on Data Mining, 2016.
2. A. Gonçalves, F. J. Von Zuben, and A. Banerjee, "A Multitask Learning View on the Earth System Model Ensemble," Computing in Science & Engineering, 2015.
3. A. Banerjee, S. Chen, and V. Sivakumar, "Open Problem: Restricted Eigenvalue Condition for Heavy Tailed Designs," Proceedings of the 28th Conference on Learning Theory, 2015.
4. F. Fazayeli , A Banerjee, J Kattge, F Schrodt, P Reich, "Uncertainty Quantified Matrix Completion using Bayesian Hierarchical Matrix Factorization," International Conference on Machine Learning Applications (ICMLA), 2014.
5. J. H. Faghmous, A. Banerjee, S. Shekhar, M. Steinbach, V. Kumar, A. R. Ganguly, and N. F. Samatova, "Theory-Guided Data Science for Climate Change," IEEE Computer, 47(11): 74-78, 2014.
6. C. Jin, Q. Fu, H. Wang, A. Agrawal, W. Hendrix, W.-K. Liao, M. A. Patwary, A. Banerjee, and A. Choudhary, "Solving Combinatorial Optimization Problems using Relaxed Linear Programming: A High Performance Computing Perspective," BigMine workshop (ACM SIGKDD), 2013.
7. A. Kaplan, X. Li, R. Sivalingam, G. Somasundaram, A. Banerjee, V. Morellas, N. Papanikolopoulos, and A. Truskinovsky. "Computer Vision Methods in Surgical Pathology: Diagnosing Carcinoma of the Breast," Annual Meeting of the United States & Canadian Academy of Pathology (USCAP), 2013.

8. S. Kasiviswanathan, H. Wang, A. Banerjee and P. Melville. "Novel Document Detection using Online L1-Dictionary Learning," ACM SIGKDD Workshop on Social Media Analytics (SOMA), 2012.
9. D. Das, A. R. Ganguly, A. Banerjee, and Z. Obradovic, "Towards understanding dominant processes in complex dynamical systems: Case of precipitation extremes," ACM SIGKDD Workshop on Knowledge Discovery from Sensor Data (SensorKDD), 2012.
10. R. Sivalingam, G. Somasundaram, X. Li, A. Banerjee, V. Morellas, N. Papanikolopoulos, and A. Truskinovsky, "Diagnosing Adenocarcinoma of the Prostate by Computer Vision Methods," Annual Meeting of the United States & Canadian Academy of Pathology (USCAP), 2012.
11. 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.
12. 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.
13. A. Banerjee, D. Boley, and S. Acharyya, "Symmetrized Bregman Divergences and Metrics," Snowbird Learning Workshop, 2009.
14. 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.
15. 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.
16. A. Banerjee and S. Basu, "A Social Query Model for Decentralized Search," 2nd ACM International Workshop on Social Network Mining and Analysis (SNAKDD), 2008.
17. 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), 2008.
18. 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.
19. 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), 2007.
20. V. Monga, A. Banerjee, and B. Evans, "Clustering Algorithms for Perceptual Image Hashing," Proceedings of IEEE Digital Signal Processing Workshop, 2004.
21. A. Banerjee, I. Dhillon, J. Ghosh, and S. Merugu, "Rate Distortion, Bregman Divergences and Maximum Likelihood Mixture Estimation," The Learning Workshop at Snowbird, 2004.
22. A. Banerjee and J. Ghosh, "Mean Model Clustering," The Learning Workshop at Snowbird, 2003.
23. 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, 2002.
24. 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, 2001.
25. A. Banerjee and J. Ghosh, "Concept-based Clustering of Clickstream Data," In Proceedings of the 3rd International Conference on Information Technology, 2000.
26. A. Banerjee, "Computerized Tumor Boundary Detection Using Genetic Algorithm," Proceedings of the National Conference on Applications of Signal Processing, 1998.

Technical Reports (Unreviewed):

1. A. A. Taheri, M. Tepper, A. Banerjee, and G. Sapiro, "If You are Happy and Know It ... Tweet," Technical Report TR-12-017, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2012.
2. A. Cherian, S. Sra, A. Banerjee, and N. Papanikolopoulos, "Jensen-Bregman LogDet Divergence for Efficient Similarity Computations on Positive Definite Tensors," Technical Report TR-12-013, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2012.
3. P. Das and A. Banerjee, "Online Quadratically Constrained Convex Optimization with Applications to Risk Adjusted Portfolio Selection," Technical Report TR-12-008, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2012.
4. Q. Fu, H. Wang, A. Banerjee, S. Liess, and P. K. Snyder, "MAP Inference on Million Node Graphical Models: KL-divergence based Alternating Directions Method," Technical Report TR-12-007, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2012.
5. H. Shan, A. Banerjee and R. Natarajan, "Probabilistic Tensor Factorization for Tensor Completion," Technical Report TR-11-0026, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2011.
6. 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.
7. 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.
8. 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.
9. 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.
10. 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.
11. 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.
12. 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.
13. 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.
14. 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.
15. H. Shan and A. Banerjee, "Bayesian Co-clustering," Technical Report TR-08-022, Department of Computer Science & Engineering, University of Minnesota, Twin Cities, 2008.
16. 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.
17. 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.

18. 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.
19. 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.
20. 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.

Patents:

1. "Inferring Emerging and Evolving Topics in Streaming Text," jointly with IBM Research, US Patent number: 8909643, December 9, 2014.
2. "Multiple Imputation of Missing Data in Multidimensional Retail Data Sets Via Tensor Factorization," jointly with IBM Research, US Patent number: 8818919, August 26, 2014.

Presentations

Invited Presentations:

- Information Theory and Applications (ITA) workshop, 2016.
- American Geophysical Union (AGU), 2015.
- DOE Plant Trait Workshop, 2015.
- IMA Data Science Seminar, 2015.
- UMN Industrial Advisory Committee (IAC) workshop, 2015.
- NSF IIS-GEO Workshop, 2015.
- Information Theory and Applications (ITA) workshop, 2015.
- Center for Ocean, Land, and Atmospheric Studies (COLA), George Mason University, 2015.
- Wisconsin Institute for Discovery, University of Wisconsin, Madison, 2014.
- NEC, Princeton, 2014.
- Information Theory and Applications (ITA) workshop, 2014.
- ICML Workshop on Covariance Selection and Graphical Model Structure Learning (ICML), 2014.
- SDM Mini-Symposium on Multiple Clusterings, Multi-view Data, and Multi-Source Knowledge-driven Clustering (SDM), 2014.
- Information Theory and Applications (ITA) workshop, 2013.
- 3M, 2012.
- University of Texas at Austin, 2012.
- Climate Knowledge Discovery Workshop, SuperComputing (SC), 2011.
- New York Academy of Sciences, Workshop on Climate Informatics (CI), 2011.
- 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, Tubingen, 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:

- A. Taheri, S. Chatterjee, and A. Banerjee, "High Dimensional Structured Estimation with Noisy Designs," SIAM International Conference on Data Mining (SDM), 2016.
- X. Liu, P. Cao, D. Zhao, and A. Banerjee, "Multi-task Spare Group Lasso for Characterizing Alzheimer's Disease," Workshop on Data Mining for Medicine and Healthcare (DMMH), SIAM International Conference on Data Mining (SDM), 2016.
- T. Zhou, H. Shan, A. Banerjee, and G. Sapiro, "Kernelized Probabilistic Matrix Factorization: Exploiting Graphs and Side Information," SIAM International Conference on Data Mining (SDM), 2012.
- 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," International Conference on Machine Learning (ICML), Pittsburgh, 2006.
- A. Banerjee, C. Krumpelman, S. Basu, R. Mooney, and J. Ghosh, "Model-based Overlapping Clustering," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2005.

- A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. Modha, "A Generalized Maximum Entropy Approach to Bregman Co-clustering and Matrix Approximation," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2004.
- A. Banerjee and J. Langford, "An Objective Evaluation Criterion for Clustering," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2004.
- A. Banerjee, X. Guo, and H. Wang, "Optimal Bregman Prediction and Jensens Equality," International Symposium on Information Theory (ISIT), 2004.
- A. Banerjee, S. Merugu, I. Dhillon, and J. Ghosh, "Clustering with Bregman Divergences," SIAM International Conference on Data Mining (SDM), 2004.
- A. Banerjee, I. Dhillon, J. Ghosh, and S. Sra, "Generative Model-based Clustering of Directional Data," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2003.

Tutorials:

- A. Banerjee and C. Monteleoni, "Climate Change: Challenges for Machine Learning," Neural Information Processing Systems (NIPS), 2014.
- A. Banerjee, "Introduction to Machine Learning," NASA Conference on Intelligent Data Understanding (CIDU), 2011.
- A. Banerjee, "Introduction to Graphical Models for Data Mining," ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2010.
- 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), 2008.
- A. Banerjee, V. Chandola, A. Lazarevic, V. Kumar, and J. Srivastava, "Anomaly Detection: A Tutorial," SIAM Data Mining Conference (SDM), 2008.

Other Key Activities & Accomplishments

- Project with Max-Planck Institute for Biogeochemistry, developed and delivered a plant trait gap-filling package which is being incorporated into TRY db, the largest database of plant information, 2014.
- 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

Semester	Year	Course
Spring	2016	CSci 5512: Artificial Intelligence II
Fall	2015	CSci 5525: Machine Learning
Spring	2015	CSci 4041: Algorithms and Data Structures
Fall	2014	CSci 5525: Machine Learning
Spring	2014	CSci 8980: Machine Learning at Large Scale and High Dimensions
Fall	2013	CSci 5525: Machine Learning
Spring	2013	CSci 5512: Artificial Intelligence II
Fall	2012	CSci 5525: Machine Learning
Spring	2012	CSci 5512: Artificial Intelligence II
Fall	2011	CSci 4041: Algorithms and Data Structures
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 8980: Machine Learning at Large Scale and High Dimensions, offered Spring 2014
  Machine learning relies heavily on large scale optimization to train models and often such models are high-dimensional, having millions of parameters. The course, designed for advanced graduate students, explored recent advances in large scale optimization and estimation high-dimensional statistical machine learning models.
- 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 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 year in Fall.
- 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

- Alex Cook, Honors Thesis, 2014.
- Thomas Gebhart, UROP, Fall, 2014.
- Garrison Kubis, UROP, Fall, 2011.
- Tinghui Zhou, UROP, Fall, 2010.
- Kim Kawatra, UROP, Spring, 2010.
- Charles Curtsinger, UROP, Fall, 2007.

Graduate Student Activities

Masters Theses Directed

- Robert Giaquinto, MS (Plan B), 2016.
  Title: Gaussian Markov Random Fields for Modeling Spatiotemporal Dependencies in Air Pollution
- Saravana Balasubramanian, MS (Plan B), 2013.
  Title: An Analysis of the Anticor Algorithm
- Subrahmanya Bhat, MS (Plan B), 2010.
  Title: Probabilistic Graph Partitioning for Topic Modeling in Text Streams
- Roman Briskine, MS (Plan B), 2008.
  Title: Clustering based Meta-prediction of Phosphorylation Sites

Masters Student Advisees (Current)

- Hardik Goel
- Shayesteh Kiaei

Doctoral Dissertations Directed

- Nicholas Johnson: PhD, 2016
  Title: Structured Online Learning with Full and Bandit Information
  First & Current Appointment: University of Pennsylvania.
- Igor Melnyk: PhD, 2016
  Title: Dynamic Bayesian Networks: Estimation, Inference, and Applications
  First & Current Appointment: IBM Research.
- André Ricardo Gonçalves: PhD, 2016 (at University of Campinas, Brazil), co-advised with Fernando Jose Von Zuben
  Title: Sparse and Structural Multitask Learning
  First & Current Appointment: Center for Research and Development in Telecommunications (CPqD), Brazil.

- Soumyadeep Chatterjee, PhD, 2015
  Title: High Dimensional Statistical Models: Applications to Climate
  First Appointment: Yahoo
  Current Appointment: Quora.
- Huahua Wang, PhD, 2014
  Title: Large Scale Optimization for Machine Learning
  First & Current Appointment: Uber ATC.
- Karthik Subbian, PhD, 2014, co-advised with Jaideep Srivastava
  Title: Scalable Analysis of Information Flow in Networks
  First & Current Appointment: Facebook.
- Puja Das, PhD, 2014
  Title: Online Convex Optimization and its application to Online Portfolio Selection
  First & Current Appointment: Apple.
- Qiang Fu, PhD, 2014
  Title: Efficient Inference Algorithms for Some Probabilistic Graphical Models
  SFirst & Current Appointment: Google.
- Nishith Pathak, co-advised with Jaideep Srivastava, PhD, 2013
  Title: Analyzing Information Flow in Social Networks for Knowledge Discovery
  First Appointment: Ninja Metrics.
- Hanhuai Shan, PhD, 2012
  Title: Probabilistic Models for Multi-relational Data Analysis
  First & Current Appointment: Microsoft.
- Amrudin Agovic, co-advised with Maria Gini, PhD, 2011
  Title: Predictive Modeling using Dimensionality Reduction and Dependency Structures
  First & Current 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)
  Current Appointment: State University of New York (SUNY) at Buffalo.

Doctoral Student Advisees (Current)

- Sheng Chen
- Konstantina Christakopoulou
- Farideh Fazayeli
- Robert Giaquinto
- Jamal Golmohammady
- Qilong Gu
- Sijie He
- Vidyashankar Sivakumar
- Amir Asiaee Taheri
- Yingxue Zhou

Postdoctoral fellows supervised

- Paolo Codenotti, 2011-2013.
  First & Current Appointment: Google.

PhD Thesis Committees

Abhirup Mallik, Aditya Pal, Akshay Soni, Ajay Joshi, Andrew Exley, Anoop Cherian, Aziz Mohaisen, Bret Borghetti, Bridget McInnes, Candace Fuller, Chris Kauffman, Daniel Acuna, David Anastasiu, Evan Ribnick, Feng Cai, Fikri Goksu, Gaurav Pandey, Gyorgy Simon, Gowtham Atluri, James Faghmous, Jaya Kawale, Jesse Vig, Jen Teshera-Levye, John Goes, Jie Chen, Joshua Vander Hook, Lane Schwartz, Mohamed Elidrisi, Muhammad Ahmad, Mazier Sanjabi, Micahel Ekstrand, Milan Shetti, Nan Jing, Nathaniel Bird, Nikolas Trawny, Kou-Wei Hsu, Pablo Spreechman, Pedro Forero, Shaozhe Tao, Seongwook Jeong, Shilad Sen, Stefan Atev, Steven Damer, Swayambhoo Jain, Tarciso Leao, Thanh Ngo, Varun Mittal, Xun Zhou, Zachary Crockett.

SERVICE AND PUBLIC OUTREACH

Service To The Discipline/Profession/Interdisciplinary Areas

Journal Reviewer Experience

Associate Editor:

- IEEE Transactions on Knowledge and Data Engineeting (TKDE), 2014–
- Journal of Aerospace Information Systems (JAIS), 2012–2016.

Reviewer:

- ACM Transactions on Knowledge Discovery from Data
- Applied Intelligence
- Bernoulli
- Biometrika
- Communications of the ACM
- Communications in Statistics - Simulation and Computation
- Data & Knowledge Engineering
- Data Mining and Knowledge Discovery
- Geoinformatica
- Geophysical Research Letters
- IEEE Geoscience and Remote Sensing Letters
- IEEE Transactions on Information Theory
- IEEE Transactions on Pattern Analysis and Machine Intelligence
- IEEE Transactions on Neural Networks
- IEEE Transactions on Neural Networks and Learning Systems
- IEEE Transactions on Knowledge and Data Engineering
- IEEE Transactions on System, Man and Cybernetics
- IEEE Transactions on Robotics
- Journal of Machine Learning Research
- Machine Learning Journal
- Nature Scientific Reports
- Neurocomputing

- PLOS One
- SIAM Journal on Scientific Computing
- Statistical Analysis and Data Mining
- Statistics and Computing
- Water Resources Research

Review panels for external funding agencies, foundations, etc.

- National Science Foundation (NSF) Panels, 2009, 2010, 2011, 2012, 2013, 2014, 2015, 2016.
- Romanian National Research Council (CNCS), 2012.
- Austrian Science Foundation (FWF), 2012.
- Israeli Science Foundation (ISF) Reviewer, 2010.
- National Aeronautics and Space Administration (NASA) Reviewer, 2004, 2011.

Organization of conferences, workshops, panels, symposia

Chair/Co-Chair:

- Workshop Co-chair, International Workshop on Climate Informatics, 2016.
- Workshop Co-chair, KDD Workshop on Data Science for Social Good, 2014.
- Program Co-chair, SIAM International Conference on Data Mining (SDM), 2014.
- Program Co-chair, International Workshop on Climate Informatics, 2012.
- 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/Area Chair:

- World Wide Web (WWW) Conference, 2017.
- SIAM International Conference on Data Mining (SDM), 2017.
- AAAI Conference on Artificial Intelligence (AAAI), 2017.
- SIAM International Conference on Data Mining (SDM), 2016.
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2015.
- International Conference on Machine Learning (ICML), 2015.
- IEEE International Conference on Data Mining (ICDM), 2015.
- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2014.
- International Conference on Machine Learning (ICML), 2014.

- ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2013.
- International Conference on Machine Learning (ICML), 2013.
- IEEE International Conference on Data Mining (ICDM), 2013.
- SIAM International Conference on Data Mining (SDM), 2013.
- ACM International Conference on Web Search and Data Mining (WSDM), 2013.
- International Conference on Machine Learning (ICML), 2012.
- SIAM Conference on Data Mining (SDM), 2012.
- ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD), 2011.
- IEEE International Conference on Data Mining (ICDM), 2011.
- SIAM Conference on Data Mining (SDM), 2010.

Program Committee:

- International Conference on Machine Learning ('07-'11,'16).
- Advances in Neural Information Processing Systems ('06-'11,'13-'16)
- IEEE International Conference on Data Mining ('05,'07,'09-'12,'14).
- ACM International Conference on Knowledge Discovery and Data Mining (KDD'08, KDD'09, KDD'10).
- SIAM Conference on Data Mining (SDM'06, SDM'07,SDM'10).
- International Conference on Artificial Intelligence and Statistics ('09,'12-'16).
- National Conference on Artificial Intelligence ('06,'07,'15,'16).

Other Activities:

- Invited panelist on "Big Data and Deep Learning: Made for Each Other?" organized by Wei Wang, UCLA, at the SIAM International Conference on Data Mining (SDM'16).
- Invited panelist on "Encoding climate knowledge into climate learning" at the Workshop on Climate Informatics (CI), 2015.
- 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).
- Workshop Reviewing: ACM Workshop on Knowledge Discovery from Sensor Data (Sensor-KDD 2007), ACM workshop on Social Network Analysis (SNAKDD 2007, 2008), AAAI Spring Symposium on Social Information Processing (2008), Workshop on Parallel and Distributed Computing for Machine Learning and Inference Problems (ParLearning 2012), KDD Workshop on Mining and Learning from Graphs (MLG 2013, 2016), KDD Workshop on Outlier Detection and Description on Demand (ODD 2013, NIPS2016), Workshop on Decision Science for Food, Energy and Water (DSFEW 2016).

Department/Unit Service

2016-17	Recruitment Committee, Evaluation Committee (Untenured faculty)
2015-16	Chair, Graduate Admissions Committee
2014-15	Chair, Graduate Admissions Committee
2013-14	Chair, Graduate Admissions Committee
2012-13	PhD & WPE Evaluation, Space Committee
2011-12	Awards Committee
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