Purnamrita Sarkar

Department of Statistics, and
Department of Electrical Engineering and Computer Sciences
University of California, Berkeley
Berkeley, CA 94704
Email: psarkar@eecs.berkeley.edu
Webpage: http://www.cs.cmu.edu/∼psarkar

Education

Postdoctoral Scholar (Aug 2013 – present). Department of Statistics and Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
Advisors: Prof. Peter J. Bickel and Prof. Michael I. Jordan

Postdoctoral Scholar (Aug 2010 – Aug 2013). Department of Electrical Engineering and Computer Sciences, University of California, Berkeley
Advisor: Prof. Michael I. Jordan

Ph.D. (Aug 2004 – Aug 2010). School of Computer Science, Carnegie Mellon University
Area: Machine Learning
Advisor: Prof. Andrew W. Moore

M.S. (Aug 2004 – Aug 2006). School of Computer Science, Carnegie Mellon University
Area: Knowledge Discovery and Data Mining
Advisor: Prof. Andrew W. Moore

B. Tech. (May 2000 – May 2004). Indian Institute of Technology, Kharagpur
Major: Computer Science and Engineering

Work experience

Aug 2013 – present: Postdoctoral Researcher, Department of Statistics, and Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

- I work with Prof. Peter J. Bickel on theoretical and practical problems in networks. Some of the projects I am working on are:
  - Theoretical analysis of the role of normalization in spectral clustering
  - A hypothesis testing framework for learning the number of clusters

Aug 2010 – Aug 2013: Postdoctoral Researcher, Department of Electrical Engineering and Computer Sciences, University of California, Berkeley

- I worked with Prof. Michael I. Jordan on solving nonparametric statistical problems arising from large scale machine learning problems. Specific projects include:
  - Nonparametric modeling of large scale dynamic social networks
  - Bag of Little Bootstraps: An efficient variant of Bootstrap for large datasets
  - Application of bootstrap to approximate query processing in large database systems and crowdsourcing

Aug 2004 – Aug 2010: Ph.D. student, Machine Learning Department, Carnegie Mellon University

- I worked with Prof. Andrew W. Moore on current problems in statistical data mining. My focus has primarily been on efficient models and algorithms for very large graphs. Specific projects include:
  - Efficient dynamic visualization of large social networks
  - Efficient computation of nearest neighbors in random walk-based proximity measures in very large graphs

Fall 2006: Research Intern, Google Inc. Pittsburgh, PA

- I worked on random walk-based proximity measures on graphs.

Feb 2002 – Apr 2004: Undergraduate Research Assistant, Indian Institute of Technology, Kharagpur, India

- I worked with Prof. Rimli Sengupta on feature selection from DNA Microarray data.

May 2003 – Aug 2003: Exchange student, Intelligent Systems Research Group, College of Computing, Georgia Institute of Technology

- I worked with Prof. Charles Isbell in the area of gene-selection from large DNA-Microarray data, with a special focus on using techniques like Independent Components Analysis (ICA).

Publications

Thesis

- P. Sarkar. "Tractable Algorithms for Proximity Search on Large Graphs." Machine Learning Department, School of Computer Science, Carnegie Mellon University, 2010.

Book chapters

- P. Sarkar and A. W. Moore. "Role of Random Walks in Ranking Applications and Social Network Analysis." *Social Network Data Analytics. Ed. Charu Aggarwal, Springer, 2010.*

In submission

- P. Sarkar and P. J. Bickel. "Role of Normalization in Spectral Clustering for Stochastic Blockmodels." *Submitted to the Annals of Statistics with favorable first round reviews, 2013.* Preprint on arXiv:1310.1495.
- P. J. Bickel and P. Sarkar. "Hypothesis Testing for Automated Community Detection in Networks." *Submitted to the Journal of the Royal Statistical Society, Series B, 2013.* Preprint on arXiv:1311.2694.
- P. Sarkar, D. Chakrabarti, and M. I. Jordan. "Nonparametric Link Prediction in Large Scale Dynamic Networks." *Submitted to the Electronic Journal of Statistics with favorable first round reviews, 2013.* Preprint on arXiv:1109.1077.

Journal papers

- A. Kleiner, A. Talwalkar, P. Sarkar, and M. I. Jordan. "A Scalable Bootstrap for Massive Data." *Journal of the Royal Statistical Society, Series B, In press, 2013.* Preprint on arXiv:1112.5016.
- P. Sarkar and A. W. Moore. "Dynamic Social Network Analysis using Latent Space Models." *ACM SIGKDD Explorations, Special Issue on Link Mining, 2005.*
- J. Leskovec, P. Sarkar, and C. Guestrin. "Modeling Link Qualities in a Sensor Network." *Informatica, 29(4): 445-452, 2005.*

Major conference papers

- B. Trushkowsky, T. Kraska, M. J. Franklin, and P. Sarkar. "Crowdsourced Enumeration Queries." *The 29th International Conference on Data Engineering (ICDE), 2013. (**Best paper award**).*
- P. Sarkar, D. Chakrabarti, and M. I. Jordan. "Nonparametric Link Prediction in Dynamic Networks." *The 29th International Conference on Machine Learning (ICML), 2012.*
- A. Kleiner, A. Talwalkar, P. Sarkar and M. I. Jordan. "The Big Data Bootstrap." *The 29th International Conference on Machine Learning (ICML), 2012.*
- P. Sarkar, D. Chakrabarti, and A. W. Moore. "Theoretical Justification of Popular Link Prediction Heuristics." *The 23rd Annual Conference on Learning Theory (COLT), 2010 (**Best student paper award**).*
  - Invited to the Best Paper Track, *International Joint Conference on Artificial Intelligence (IJCAI), 2011.*

- P. Sarkar and A. W. Moore. "Fast Nearest-neighbor Search in Disk-resident Graphs." *The 16th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), 2010*.
- P. Sarkar and A. W. Moore. "Fast Dynamic Reranking in Large Graphs." *The 18th International World Wide Web Conference (WWW), Data Mining Track, 2009*.
- P. Sarkar, A. W. Moore and A. Prakash. "Fast Incremental Proximity Search in Large Graphs." *The 25th International Conference on Machine Learning (ICML), 2008*.
- P. Sarkar and A. W. Moore. "A Tractable Approach to Finding Closest Truncated-commute-time Neighbors in Large Graphs." *The 23rd Conference on Uncertainty in Artificial Intelligence (UAI), 2007*.
- P. Sarkar, S. Siddiqi and G. Gordon. "A Latent Space Approach to Dynamic Embedding of Co-occurrence Data." *Eleventh International Conference on Artificial Intelligence and Statistics (AISTATS), 2007*.
- P. Sarkar and A. W. Moore. "Dynamic Social Network Analysis using Latent Space Models." *Advances in Neural Information Processing Systems (NIPS), 2005*.

Other conference and workshop papers

- P. Sarkar, L. Chen, and A. Dubrawski. "Trade-offs between Agility and Reliability of Predictions in Dynamic Social Networks Used to Model Risk of Microbial Contamination of Food." *International Conference on Social Networks Analysis and Mining*, 2009 (***Best Paper Award***)
- P. Sarkar and G. Gordon. "Random Walks with Random Projections." *Workshop on Analyzing Networks and Learning with Graphs, at the Advances in Neural Information Processing Systems (NIPS), 2009.*
- P. Sarkar, L. Chen, and A. Dubrawski. "Dynamic Network Model for Predicting Occurrences of Salmonella at Food Facilities." *Biosurveillance and Biosecurity, International Workshop (BioSecure), 2008*.
- P. Sarkar, S. M. Siddiqi, and G. Gordon. "Approximate Kalman Filters for Embedding Author-Word Co-occurrence Data over Time." *Workshop on Statistical Network Analysis at the 23rd International Conference on Machine Learning, 2006*.

Awards and honors

Best paper award. The 29th International Conference on Data Engineering (ICDE), 2013

Best student paper award. The 23rd Annual Conference on Learning Theory (COLT), 2010

Best paper award. International Conference on Social Networks Analysis and Mining (ASONAM), 2009

Director's commendation for academic excellence for Spring semester, 2005 and Fall semester, 2006, at IIT Kharagpur

Teaching experience

Machine Learning (For Graduate Students) : Teaching assistant. Spring 2007.

Machine Learning (For Advanced Undergraduates and Masters Students) : Teaching assistant. Spring 2009.

Analytics and Business Intelligence (For Graduate Students) : Substitute lecturer. Spring 2010.

Grants

Helped Prof. Michael I. Jordan in the preparation of grant titled "Statistical Machine Learning in the Distributed Computing Setting", sponsored by the Information Technology Laboratory (ITL) Program at NIST, awarded in 2010.

Talks

Invited talk at the Workshop on Large Graphs: Modeling, Algorithms, and Applications, IMA: "Nonparametric Link Prediction", Oct, 2011.

Invited talk at Complex Network Modeling Workshop, SAMSI: "Theoretical Justification of Popular Link Prediction Heuristics", Oct, 2010.

Invited talk at Joint Statistical Meetings: "Probabilistic Modeling of Dynamic Networks using Latent Space Models", August, 2010.

The International World Wide Web Conference, Data Mining Track: "Fast Dynamic Reranking in Large Graphs." April 2009.

Heinz School of Public Policy and Management, CMU: "Social Networks: A General Overview." April 2009.

International Conference on Machine Learning: "Fast Incremental Proximity Search in Large Graphs." July 2008.

Carnegie Mellon University: "Random Walks on Graphs: A General Overview." October 2007.

Google Research, New York: "A Tractable Approach to Finding Closest Truncated-hitting-time Neighbors in Large Graphs." June 2007.

The Snowbird Workshop: "A Tractable Approach to Finding Closest Truncated-hitting-time Neighbors in Large Graphs." March 2007.

Social Net Mid-Year Workshop, SAMSI: "Dynamic Social Network Analysis using Latent Space Models." March 2006.

Professional services

Reviewed for International Conference of Machine Learning (ICML), ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD), Conference of Learning Theory (COLT), ACM Symposium on the Theory of Computing (STOC), Journal of Machine Learning Research (JMLR), Annals of Statistics, Journal of the Royal Statistical Society, Series C

Served as a reviewer on NSF Data Mining proposal review panel, 2012.

Program Committee member for KDD 2011.

Program Committee member for ICML 2010.

Served as a volunteer for ICML 2008.

Served as a student member of the graduate admissions committee, 2006-2008.

Served as a member of the local organizing committee for ICML 2006.