

## Purnamrita Sarkar

Machine Learning Department  
Carnegie Mellon University  
5000 Forbes Ave.  
Pittsburgh, PA 15213

412-855-8298 (cell)  
412-268-2022 (office)  
Email: [psarkar@cs.cmu.edu](mailto:psarkar@cs.cmu.edu)  
URL: <http://www.cs.cmu.edu/~psarkar>

### RESEARCH INTERESTS

Link prediction in social networks, personalized graph search techniques, keyword search in databases, collaborative filtering and fraud detection are important practical problems that requires ranking entities based on graph theoretic measures of similarity to a given query. These graphs are generally very large in size and so designing fast and memory-efficient algorithms for query processing is of key importance. Currently I am working towards developing new *incremental* algorithms with formally proven properties for ranking problems on graphs of up to a billion nodes on a single CPU machine.

I have also worked on tractable generative models which capture the evolution of complex networks over time. In this project we generalized a successful static model of relationships into a dynamic model that accounts for friendships drifting over time. We also developed tractable learning algorithms that can scale to large social networks. More details about these research projects can be found at

[http://www.cs.cmu.edu/~psarkar/app/sarkar\\_research\\_stmt.pdf](http://www.cs.cmu.edu/~psarkar/app/sarkar_research_stmt.pdf)

### EDUCATION

**Ph.D. candidate** School of Computer Science, Carnegie Mellon University (Aug 2004–present)

Area: Machine Learning  
Advisor: Prof. Andrew W. Moore  
Expected graduation date: May, 2010

**M.S.** School of Computer Science, Carnegie Mellon University (Aug 2004–Aug 2006)

Area: Knowledge Discovery and Data Mining

**B. Tech.** Indian Institute of Technology, Kharagpur (May 2000–May 2004)

Major: Computer Science and Engineering

### WORK EXPERIENCE

**Aug 2004–Present:** Graduate Research Assistant, The Auton Lab, Carnegie Mellon University,

- I work with Prof. Andrew 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

- 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 Tech

- 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

### Book chapters

- P. Sarkar and A. W. Moore. “Role of Random Walks in Ranking Applications and Social Network Analysis” To Appear in *Social Network Data Analytics*. Ed. Charu Aggarwal, Springer (2010)

### Journal papers

- P. Sarkar and A. W. Moore. “Dynamic Social Network Analysis using Latent Space Models.” *ACM SIGKDD Explorations, Special Issue on Link Mining*, (2005).
- Jure Leskovec, Purnamrita Sarkar, Carlos Guestrin “Modeling Link Qualities in a Sensor Network.” *Informatica*, 29(4): 445-452 2005

### Major Conference papers

- 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 25rd 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.

### Selected Manuscripts

- P. Sarkar, A. W. Moore “Fast Nearest-neighbor Search in Disk-resident Graphs” Submitted, 2010. Also published as a technical report (CMU-ML-10-100).
- P. Sarkar, D. Chakrabarti, A. W. Moore “Theoretical Justification of Popular Heuristics for Link Prediction” In preparation, 2010.

### Other Conferences and Workshop Papers

- Purnamrita Sarkar, Lujie Chen, Artur 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 (Winner of the Best Paper Award)
- Purnamrita Sarkar, Geoffrey J. Gordon “RandomWalks with Random Projections” *Workshop on Analyzing Networks and Learning with Graphs, at the Advances in Neural Information Processing Systems (NIPS) 2009*
- Purnamrita Sarkar, Lujie Chen, Artur Dubrawski “Dynamic Network Model for Predicting Occurrences of Salmonella at Food Facilities.” *Biosurveillance and Biosecurity, International Workshop (BioSecure)*, 2008.
- Purnamrita Sarkar, Sajid M. Siddiqi, and Geoffrey J. 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.

## **Selected Talks**

**Invited talk at Joint Statistical Meeting 2010** : Vancouver, Canada. “Probabilistic Modeling of Dynamic Networks using Latent Space Models”, July 31-August 5, 2010.

**The International World Wide Web Conference, Data Mining Track** : Madrid, Spain. “Fast Dynamic Reranking in Large Graphs.” April 2009.

**Heinz School of Public Policy and Management, CMU** : Pittsburgh, PA. “Social Networks: A General Overview.” April 2009.

**International Conference on Machine Learning** : Helsinki, Finland. “Fast Incremental Proximity Search in Large Graphs.” July 2008.

**Carnegie Mellon University** : Pittsburgh, PA. “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** : San Juan, Puerto Rico. “A Tractable Approach to Finding Closest Truncated-hitting-time Neighbors in Large Graphs.” March 2007.

**Social Net Mid-Year Workshop, at SAMSI**<sup>1</sup>: Hosted by Carnegie Mellon University, Pittsburgh, PA. “Dynamic Social Network Analysis using Latent Space Models.” March 2006.

## **PROFESSIONAL SERVICES**

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.

## **TEACHING EXPERIENCE**

**Machine Learning (For Graduate Students)** (2007 spring).

**Machine Learning (For Advanced Undergraduates and Masters Students)** (2009 spring).

## **REFERENCES**

Provided on request.

---

<sup>1</sup>Statistical and Applied Mathematical Sciences Institute