

# Stephen H. Bach

## Curriculum Vitae

bach@cs.stanford.edu  
http://stephenbach.net

**Research** I investigate **machine learning** methods for incorporating high-level knowledge into statistical models. My approach is to develop new methods that blend statistical methods with rule-based systems to achieve benefits from each. Topics I have worked on include weak supervision, statistical relational learning, and information extraction.

**Positions** **Postdoctoral Scholar, 2015–Present**  
**Stanford University**  
Computer Science Department  
Advisor: Christopher Ré

**Education** **Ph.D., Computer Science, 2015**  
**University of Maryland, College Park**  
Advisor: Lise Getoor  
Dissertation: Hinge-Loss Markov Random Fields and Probabilistic Soft Logic:  
A Scalable Approach to Structured Prediction  
Committee: Rama Chellapa, Hal Daumé III, Larry Davis, Kevin Murphy  
**Larry S. Davis Doctoral Dissertation Award**

**B.S., Computer Science and Mathematics (double major), 2010**  
**Georgetown University**  
Advisor: Mark Maloof  
Magna Cum Laude

## Publications

### Journal Papers

- S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. Hinge-loss Markov random fields and probabilistic soft logic. *Journal of Machine Learning Research (JMLR)*, 18(109):1–67, 2017.
- A. J. Ratner, S. H. Bach, H. E. Ehrenberg, J. Fries, S. Wu, and C. Ré. Snorkel: Rapid training data creation with weak supervision. *PVLDB*, 11(3):269–282, 2017.
- G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Soft quantification in statistical relational learning. *Machine Learning*, 2017.

### Peer-Reviewed Conference Papers

- S. H. Bach, B. He, A. J. Ratner, and C. Ré. Learning the structure of generative models without labeled data. In *International Conference on Machine Learning (ICML)*, 2017.
- H. Lakkaraju, S. H. Bach, and J. Leskovec. Interpretable decision sets: A joint framework for description and prediction. In *ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)*, 2016.
- S. H. Bach\*, B. Huang\*, J. Boyd-Graber, and L. Getoor. Paired-dual learning for fast training of latent variable hinge-loss MRFs. In *International Conference on Machine Learning (ICML)*, 2015.
- S. H. Bach, B. Huang, and L. Getoor. Unifying local consistency and MAX SAT relaxations for scalable inference with rounding guarantees. In *Artificial Intelligence and Statistics (AISTATS)*, 2015.  
**Selected for oral presentation, 6% of submitted papers (27/442).**
- G. Farnadi, S. H. Bach, M. Blondeel, M.-F. Moens, L. Getoor, and M. De Cock. Statistical relational learning with soft quantifiers. In *International Conference on Inductive Logic Programming (ILP)*, 2015.  
**Best Student Paper Award.**

- S. H. Bach, B. Huang, B. London, and L. Getoor. Hinge-loss Markov random fields: Convex inference for structured prediction. In *Uncertainty in Artificial Intelligence (UAI)*, 2013.
- S. H. Bach, M. Broecheler, L. Getoor, and D. P. O’Leary. Scaling MPE inference for constrained continuous Markov random fields with consensus optimization. In *Advances in Neural Information Processing Systems (NIPS)*, 2012.
- S. H. Bach and M. A. Maloof. A Bayesian approach to concept drift. In *Advances in Neural Information Processing Systems (NIPS)*, 2010.
- S. H. Bach\* and M. A. Maloof\*. Paired learners for concept drift. In *IEEE International Conference on Data Mining (ICDM)*, 2008.

### Demonstrations

- A. J. Ratner, S. H. Bach, H. E. Ehrenberg, and C. Ré. Snorkel: Fast training set generation for information extraction. ACM SIGMOD Conference on Management of Data (SIGMOD), 2017.

### Workshop Papers

- S. H. Bach, B. Huang, and L. Getoor. Rounding guarantees for message-passing MAP inference with logical dependencies. In *NIPS Workshop on Discrete and Combinatorial Problems in Machine Learning (DISCML)*, 2014.
- S. H. Bach, B. Huang, and L. Getoor. Probabilistic soft logic for social good. In *KDD Workshop on Data Science for Social Good*, 2014.
- G. Farnadi, S. H. Bach, M. Moens, L. Getoor, and M. De Cock. Extending PSL with fuzzy quantifiers. In *International Workshop on Statistical Relational Artificial Intelligence (StaRAI)*, 2014.
- S. H. Bach, B. Huang, and L. Getoor. Large-margin structured learning for link ranking. In *NIPS Workshop on Frontiers of Network Analysis: Methods, Models, and Applications*, 2013.  
**Best Student Paper Award.**
- S. H. Bach, B. Huang, and L. Getoor. Learning latent groups with hinge-loss Markov random fields. In *ICML Workshop on Inferning: Interactions between Inference and Learning*, 2013.
- B. London, S. Khamis, S. H. Bach, B. Huang, L. Getoor, and L. Davis. Collective activity detection using hinge-loss Markov random fields. In *CVPR Workshop on Structured Prediction: Tractability, Learning and Inference*, 2013.
- A. Kimmig, S. H. Bach, M. Broecheler, B. Huang, and L. Getoor. A short introduction to probabilistic soft logic. In *NIPS Workshop on Probabilistic Programming: Foundations and Applications*, 2012.
- B. Huang, S. H. Bach, E. Norris, J. Pujara, and L. Getoor. Social group modeling with probabilistic soft logic. In *NIPS Workshop on Social Network and Social Media Analysis: Methods, Models, and Applications*, 2012.
- A. Memory, A. Kimmig, S. H. Bach, L. Raschid, and L. Getoor. Graph summarization in annotated data using probabilistic soft logic. In *Proceedings of the International Workshop on Uncertainty Reasoning for the Semantic Web (URSW)*, 2012.
- S. H. Bach, M. Broecheler, S. Kok, and L. Getoor. Decision-driven models with probabilistic soft logic. In *NIPS Workshop on Predictive Models in Personalized Medicine*, 2010.

\* Equal Contributors

### Invited Talks

- Snorkel: Creating Noisy Training Data to Overcome Machine Learning’s Biggest Bottleneck*  
 Berkeley AI Research Lab, University of California, Berkeley, July 10 2017.  
 SIGMOD Workshop on Data Management for End-to-End Machine Learning (DEEM), May 14 2017.
- Probabilistic Soft Logic: Fast and Accurate Models for Big Graphs*  
 Computer Science Department, University of California, Santa Cruz, March 4 2016.

InfoLab, Stanford University, February 11 2015.  
San Diego Supercomputer Center, University of California, San Diego, February 9 2015.  
Charles River Analytics, Cambridge, MA, August 20 2014.

## Teaching Experience

### Teaching Assistant

Introduction to Artificial Intelligence  
Fall 2011  
Computer Science Department, University of Maryland, College Park

### Teaching Assistant

Computer Science II  
Spring 2008  
Computer Science Department, Georgetown University

## Honors and Awards

### Larry S. Davis Doctoral Dissertation Award

2015, Computer Science Department, University of Maryland, College Park

### John D. Gannon Award for Outstanding Graduate Student

2013, Computer Science Department, University of Maryland, College Park

### Computer Science Award

2010, Georgetown University

## Professional Activities

### Workshop Organizer

2017 NIPS Workshop on Learning with Limited Labeled Data: Weak Supervision and Beyond

### National Science Foundation Peer Review Panelist

Division of Information and Intelligent Systems (IIS)

### Conference Program Committee / Reviewer

ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD)  
International Conference on Learning Representations (ICLR)  
International Conference on Machine Learning (ICML)  
International Joint Conference on Artificial Intelligence (IJCAI)  
International World Wide Web Conference (WWW)  
Neural Information Processing Systems (NIPS)  
Uncertainty in Artificial Intelligence (UAI)

### Journal Reviewer

ACM Transactions on Knowledge Discovery in Data (TKDD)  
Data Mining and Knowledge Discovery  
IEEE Transactions on Knowledge and Data Engineering (TKDE)  
IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)  
Journal of Machine Learning Research (JMLR)  
Science Advances  
Statistical Analysis and Data Mining

### Workshop Program Committee

Automatic Knowledge Base Construction (AKBC)  
Data Driven Discovery of Models (D3M)  
Knowledge Base Construction, Reasoning and Mining (KBCOM)  
Mining and Learning with Graphs (MLG)

## References

Christopher Ré  
Associate Professor of Computer Science  
Stanford University  
chrismre@stanford.edu

Lise Getoor  
Professor of Computer Science  
University of California, Santa Cruz  
getoor@soe.ucsc.edu

Mike Cafarella  
Associate Professor of Computer Science and Engineering  
University of Michigan  
michjc@umich.edu

Peter Bailis  
Assistant Professor of Computer Science  
Stanford University  
pbailis@cs.stanford.edu

December 4, 2017