

# David M. Blei

Princeton University  
Department of Computer Science  
35 Olden Street  
Princeton, NJ 08540

Phone: (609) 258-9907  
Fax: (609) 258-1771  
Email: [blei@cs.princeton.edu](mailto:blei@cs.princeton.edu)  
Homepage: <http://www.cs.princeton.edu/~blei>

## EDUCATION

B.Sc. (Honors) Computer Science and Mathematics, Brown University, 1997.

Ph.D. Computer Science, University of California Berkeley, 2004.

## EMPLOYMENT

Associate Professor, Department of Computer Science, Princeton University, 2011–

Assistant Professor, Department of Computer Science, Princeton University, 2006–2011

Postdoctoral Fellow, Department of Machine Learning, Carnegie Mellon University, 2004–2006

## AWARDS

Office of Naval Research Young Investigator Award, 2011

Alfred P. Sloan Fellowship, 2010

E.L. Keyes Jr. Emerson Electric Co. Faculty Award, 2008

National Science Foundation CAREER Award, 2008

Microsoft New Faculty Fellowship Finalist, 2007

Microsoft Research Award, 2007

Google Research Award, 2006, 2007, 2010

Princeton Engineering Commendation List for Outstanding Teaching,  
Spring 2006, Fall 2006, Spring 2007, Fall 2007, Spring 2008, Spring 2009, Fall 2009

U.C. Berkeley C.V. Ramamoorthy Distinguished Research Award, 2006

Microsoft Research Graduate Fellowship, 2002

Berkeley Micro-Electronics Fellowship, 1999

Sigma Xi Scientific Honor Society, 1997

## POPULAR PRESS

“Avalanches of Words, Sifted and Sorted.” *The New York Times*. March 24, 2012.

“Organising the Web: The Science of Science.” *The Economist*. April 28, 2011.

“Statistical Time Travel Helps to Answer What-Ifs.” *Wall Street Journal*. November 12, 2009.

## PUBLICATIONS

*Refereed Journal Articles*

1. J. Paisley, C. Wang and D. Blei. The discrete infinite logistic normal distribution. *Bayesian Analysis*, to appear.
2. D. Blei. Probabilistic Topic Models. *Communications of the ACM*, 55(4):77–84.
3. S. Gershman and D. Blei. A Tutorial on Bayesian Nonparametric Models. *Journal of Mathematical Psychology*, 56:1–12, 2012.
4. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research*, 12:2461–2488, 2011.
5. L. Hannah, D. Blei and W. Powell. Dirichlet process mixtures of generalized linear models. *Journal of Machine Learning Research*, 12:1923–1953, 2011.
6. S. Gershman, D. Blei, F. Pereira, and K. Norman. A topographic latent source model for fMRI data. *NeuroImage*, 57:89–100, 2011.
7. D. Blei, L. Carin, and D. Dunson. Probabilistic topic models. *Signal Processing*, 27(6):55–65, 2010.
8. D. Blei, T. Griffiths, and M. Jordan. The nested Chinese restaurant process and Bayesian non-parametric inference of topic hierarchies. *Journal of the ACM*, 57(2):1–30, 2010.
9. J. Chang and D. Blei. Hierarchical relational models for document networks. *Annals of Applied Statistics*, 4(1), 2010.
10. S. Gershman, D. Blei, and Y. Niv. Context, learning and extinction. *Psychological Review*, 117(1):197–209, 2010.
11. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.
12. D. Blei and J. Lafferty. A correlated topic model of Science. *Annals of Applied Statistics*, 1(1):17–35, 2007.
13. D. Blei and S. Fienberg. Discussion of model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A*, 170:332, 2007.

14. J. McAuliffe, D. Blei, and M. Jordan. Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16(1):5–14, 2006.
15. Y. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.
16. D. Blei, K. Franks, M. Jordan, and S. Mian. Statistical modeling of biomedical corpora: mining the Caenorhabditis Genetic Center Bibliography for genes related to life span. *BMC Bioinformatics*, 7(250), 2006.
17. D. Blei and M. Jordan. Variational inference for Dirichlet process mixtures. *Journal of Bayesian Analysis*, 1(1):121–144, 2005.
18. K. Barnard, P. Duygulu, N. de Freitas, D. Forsyth, D. Blei, and M. Jordan. Matching words and pictures. *Journal of Machine Learning Research*, 3:1107–1135, 2003.
19. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

#### *Refereed Conference Articles*

20. J. Paisley, D. Blei and M Jordan. Variational Bayesian inference with stochastic search. In *International Conference On Machine Learning*, 2012.
21. D. Mimno, M. Hoffman and D Blei. Sparse stochastic inference for latent Dirichlet allocation. In *International Conference On Machine Learning*, 2012.
22. S. Gershman, M. Hoffman and D Blei. Nonparametric variational inference. In *International Conference On Machine Learning*, 2012.
23. A. Chaney and D Blei. Visualizing topic models. In *International AAI Conference on Weblogs and Social Media*, 2012.
24. J. Paisley, D. Blei, and M. Jordan. Stick-breaking beta processes and the Poisson process. In *Artificial Intelligence and Statistics*, 2012.
25. S. Ghosh, A. Ungureanu, E. Sudderth, and D. Blei. A Spatial distance dependent Chinese restaurant process for image segmentation. In *Neural Information Processing Systems*, 2011.
26. C. Wang and D. Blei. Collaborative topic modeling for recommending scientific articles. In *Knowledge Discovery and Data Mining*, 2011. **Best Student Paper Award.**
27. D. Mimno and D. Blei. Bayesian checking for topic models. In *Empirical Methods in Natural Language Processing*, 2011.
28. S. Gerrish and D. Blei. Predicting legislative roll call from text. In *International Conference on Machine Learning*, 2011. **Distinguished Application Paper Award.**
29. J. Paisley, D. Blei, and L. Carin. Variational inference for stick-breaking beta process priors. In *International Conference on Machine Learning*, 2011.

30. J. Paisley, C. Wang and D. Blei. The discrete infinite logistic normal distribution for mixed-membership modeling. In *Artificial Intelligence and Statistics*, 2011. **Notable Paper Award.**
31. C. Wang, J. Paisley and D. Blei. Online variational inference for the hierarchical Dirichlet process. In *Artificial Intelligence and Statistics*, 2011.
32. M. Hoffman, D. Blei, and F. Bach. On-line learning for latent Dirichlet allocation. In *Neural Information Processing Systems*, 2010.
33. L. Hannah, W. Powell, and D. Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In *Neural Information Processing Systems*, 2010.
34. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. In *International Conference on Machine Learning*, 2010.
35. S. Gerrish and D. Blei. A language-based approach to measuring scholarly impact. In *International Conference on Machine Learning*, 2010.
36. M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning*, 2010.
37. S. Williamson, C. Wang, K. Heller, and D. Blei. The IBP compound Dirichlet process and its application to focused topic modeling. In *International Conference on Machine Learning*, 2010.
38. L. Hannah, D. Blei, and W. Powell. Dirichlet process mixtures of generalized linear models. In *Artificial Intelligence and Statistics*, 2010.
39. A. Lorbert, D. Eis, V. Kostina, D. Blei, and P. Ramadge. Exploiting covariate similarity in sparse regression via the pairwise elastic net. In *Artificial Intelligence and Statistics*, 2010.
40. J. Li, C. Wang, Y. Lim, D. Blei, and L. Fei-Fei. Building and using a semantivisual image hierarchy. In *Computer Vision and Pattern Recognition*, 2010.
41. S. Cohen, D. Blei, and N. Smith. Variational inference for adaptor grammars. In *North American Chapter of the Association for Computational Linguistics*, 2010.
42. C. Wang and D. Blei. Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process. In *Neural Information Processing Systems*, 2009.
43. C. Wang and D. Blei. Variational inference for the nested Chinese restaurant process. In *Neural Information Processing Systems*, 2009.
44. R. Socher, S. Gershman, A. Perotte, P. Sederberg, D. Blei, and K. Norman. A Bayesian analysis of dynamics in free recall. In *Neural Information Processing Systems*, 2009.
45. J. Chang, J. Boyd-Graber, S. Gerrish, C. Wang, and D. Blei. Reading tea leaves: How humans interpret topic models. In *Neural Information Processing Systems*, 2009. **Honorable Mention: Best Student Paper Award.**
46. J. Chang, J. Boyd-Graber, and D. Blei. Connections between the lines: Augmenting social networks with text. In *Knowledge Discovery and Data Mining*, 2009.

47. J. Boyd-Graber and D. Blei. Multilingual topic models for unaligned text. In *Uncertainty in Artificial Intelligence*, 2009.
48. J. Chang and D. Blei. Relational topic models for document networks. In *Artificial Intelligence and Statistics*, 2009.
49. C. Wang, B. Thiesson, C. Meek, and D. Blei. Markov topic models. In *Artificial Intelligence and Statistics*, 2009.
50. M. Hoffman, D. Blei, and P. Cook. Finding latent sources in recorded music with a shift-invariant HDP. In *International Conference on Digital Audio Effects*, 2009.
51. M. Hoffman, D. Blei, and P. Cook. Easy as CBA: A simple probabilistic model for tagging music. In *International Conference on Music Information Retrieval*, 2009. **Best Student Paper Award.**
52. M. Hoffman, P. Cook, and D. Blei. Bayesian spectral matching: Turning young MC into MC hammer via MCMC sampling. In *International Computer Music Conference*, 2009.
53. C. Wang, D. Blei, and L. Fei-Fei. Simultaneous image classification and annotation. In *Computer Vision and Pattern Recognition*, 2009.
54. I. Mukherjee and D. Blei. Relative performance guarantees for approximate inference in latent Dirichlet allocation. In *Neural Information Processing Systems*, 2008.
55. J. Boyd-Graber and D. Blei. Syntactic topic models. In *Neural Information Processing Systems*, 2008.
56. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. In *Neural Information Processing Systems*, 2008.
57. C. Wang, D. Blei, and D. Heckerman. Continuous time dynamic topic models. In *Uncertainty in Artificial Intelligence (UAI)*, 2008.
58. M. Hoffman, D. Blei, and P. Cook. Content-based musical similarity computation using the hierarchical Dirichlet process. In *International Conference on Music Information Retrieval*, 2008.
59. M. Hoffman, P. Cook, and D. Blei. Data-driven recomposition using the hierarchical Dirichlet process hidden Markov model. In *International Computer Music Conference*, 2008.
60. M. Dudik, D. Blei, and R. Schapire. Hierarchical maximum entropy density estimation. In *Proceedings of the 28th International Conference on Machine Learning*, 2007.
61. W. Li, D. Blei, and A. McCallum. Nonparametric Bayes pachinko allocation. In *The 23rd Conference on Uncertainty in Artificial Intelligence*, 2007.
62. D. Kaplan and D. Blei. A computational approach to style in American poetry. In *IEEE Conference on Data Mining*, 2007.
63. D. Blei and J. McAuliffe. Supervised topic models. In *Neural Information Processing Systems*, 2007.

64. J. Boyd-Graber, D. Blei, and X. Zhu. A topic model for word sense disambiguation. In *Empirical Methods in Natural Language Processing*, 2007.
65. D. Blei and J. Lafferty. Correlated topic models. In *Neural Information Processing Systems*, 2006.
66. D. Blei and J. Lafferty. Dynamic topic models. In *International Conference on Machine Learning*, 2006.
67. T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum. Integrating topics and syntax. In *Neural Information Processing Systems*, 2005.
68. D. Blei and M. Jordan. Variational methods for the Dirichlet process. In *International Conference on Machine Learning*, 2004.
69. D. Blei and M. Jordan. Modeling annotated data. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003.
70. D. Blei, T. Griffiths, M. Jordan, and J. Tenenbaum. Hierarchical topic models and the nested Chinese restaurant process. In *Neural Information Processing Systems*, 2003. **Best Student Paper Award.**
71. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. In *Neural Information Processing Systems*, 2002.
72. D. Blei, J. Bagnell, and A. McCallum. Learning with scope, with application to information extraction and classification. In *Uncertainty in Artificial Intelligence*, 2002.
73. D. Blei and P. Moreno. Topic segmentation with an aspect hidden Markov model. In *ACM SIGIR conference on Research and Development in Information Retrieval*, 2001.

### *Book Chapters*

74. D. Blei and J. Lafferty. Topic models. In A. Srivastava and M. Sahami, editors, *Text Mining: Classification, Clustering, and Applications*. Chapman & Hall/CRC Data Mining and Knowledge Discovery Series, 2009.
75. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Combining stochastic block models and mixed membership for statistical network analysis. In *Statistical Network Analysis: Models, Issues and New Directions*, Lecture Notes in Computer Science, pages 57–74. Springer-Verlag, 2007.
76. D. Blei, A. Ng, and M. Jordan. Hierarchical Bayesian models for applications in information retrieval. In J. Bernardo, J. Berger, A. Dawid, D. Heckerman, A. Smith, and M. West, editors, *Bayesian Statistics 7*, volume 7, pages 25–44. Oxford University Press, 2003.

### *Edited Volume*

77. E. Airoldi, D. Blei, S. Fienberg, A. Goldenberg, E. Xing, and A. Zheng, editors. *Statistical Network Analysis: Models, Issues and New Directions*. Lecture Notes in Computer Science. Springer-Verlag, 2007.

## AWARDED GRANTS

1. *Scalable Topic Modeling: Online Learning, Diagnostics, and Recommendation*. PI: David M. Blei. Office of Naval Research. \$510K. 2011-2014.
2. *Dynamic and Supervised Topic Models for Literature-Based Discovery*. PI: David M. Blei. Office of Naval Research. \$300K. 2008-2011.
3. *Non-Parametric Bayesian Analysis of Heterogeneous Data*. PI: David M. Blei. Air Force Office of Scientific Research. \$360K. 2009-2012.
4. *CAREER: New Directions in Probabilistic Topic Models*. National Science Foundation. PI: David M. Blei. \$550K. 2008-2013.
5. *Text, Neuroimaging, and Memory: Unified Models of Corpora and Cognition*. PI: David M. Blei. National Science Foundation. \$730K. 2010-2013.
6. *Interactive Discovery and Semantic Labeling of Patterns in Spatial Data* PI: Thomas Funkhouser. Co-PI: David M. Blei. National Science Foundation. \$500K. 2009-2012.

## PROFESSIONAL ACTIVITIES

### Senior Program Committee

Neural Information Processing Systems (2009, 2010)  
International Conference on Machine Learning (2008, 2009, 2010, 2011, 2012)  
Artificial Intelligence and Statistics (2008, 2012)

### Editorial Board

Journal of Machine Learning Research  
Machine Learning Journal  
Journal of Artificial Intelligence Research  
Statistics and Computing  
Chapman Hall Series on Computer Science and Data Analysis

### Conference Reviewing

Neural Information Processing Systems (2005, 2006, 2007, 2008, 2011)  
International Conference on Machine Learning (2006, 2007)  
Artificial Intelligence and Statistics (2005, 2007, 2010)  
Uncertainty in Artificial Intelligence (2005, 2006, 2007)  
International Joint Conference on Artificial Intelligence (2005)  
Association of Artificial Intelligence (2007)  
SIGIR Conference on Information Retrieval (2005)

Knowledge Discovery and Data Mining (2005)  
Empirical Methods in Natural Language Processing (2007)  
Association of Computational Linguistics (2008)

### **Journal Reviewing**

Proceedings of the National Academy of Science  
Journal of Machine Learning Research  
Machine Learning Journal  
Journal of the American Statistical Association  
Journal of the Royal Statistical Society,  
Annals of Applied Statistics  
Bayesian Analysis  
Statistics and Computing  
IEEE Transactions on Pattern Analysis and Machine Intelligence  
IEEE Transactions on Neural Networks  
IEEE Transactions on Audio, Speech, and Language Processing  
International Journal on Very Large Data Bases  
ACM Transactions on Knowledge Discovery from Data

### **Grant Reviewing**

NSF Panel IIS (2008, 2009, 2010, 2012)

### **Princeton University**

Executive Committee for the Committee on Statistical Studies  
Faculty Advisory Committee on Athletics and Campus Recreation  
Advisor to Computer Science A.B. classes of 2009 and 2010  
Program in Applied and Computational Mathematics  
Princeton Institute for Computational Science and Engineering, Associated Faculty  
Center for Information Technology Policy, Affiliated Faculty  
Princeton Neuroscience Institute, Affiliated Faculty

### **Workshop organizing**

NIPS Workshop on Syntax and Semantics (2003)  
ICML Workshop on Statistical Network Analysis (2005)  
NIPS Workshop on Applications of Topic Modeling (2008)

### **Ph.D. Students**

Jordan Boyd-Graber (2009). Assistant Professor, University of Maryland

Matthew Hoffman (2010, co-advised with Perry Cook). Postdoctoral Fellow, Columbia University

Lauren Hannah (2010, co-advised with Warren Powell); Postdoctoral Fellow, Duke University

Jonathan Chang (2011). Data Scientist, Facebook

Chong Wang (expected 2012)

Gungor Polatkan (expected 2012, co-advised with Ingrid Dubeiches)

Sean Gerrish (expected 2012)

Tian Wang (expected 2015)

Allison Chaney (expected 2016)

Rajesh Ranganath (expected 2016)

#### **Other Ph.D. Thesis Committees**

David Mimno (University of Massachusetts, expected 2011)

Umar Syed (Princeton, 2010)

Melissa Carroll (Princeton, 2010)

Vasileios Kandylas (University of Pennsylvania, 2009)

Emily Fox (MIT, 2009)

Chenwei Zhu (Princeton, 2008)

Zafer Barutcuoglu (Princeton, 2008)

Katherine Heller (University College London, 2008)

Suhrid Balakrishnan (Rutgers, 2007)

Wei Li (UMass Amherst, 2007)

Miroslav Dudik (Princeton, 2007)

#### **PROFESSIONAL MEMBERSHIPS**

Association of Computing Machinery

Bernoulli Society

Institute for Mathematical Statistics

American Statistical Association

#### **INVITED TALKS**

AAAI Spring Symposium (2002)

BAE Systems (2008)

Bayesian nonparametrics workshop (2011)

Boston University (2011)

Brown University (2005)

Carnegie Mellon University (2003, 2005, 2009)  
Columbia University (2007, 2009, 2010)  
Cornell University (2007, 2010)  
DIMACS (2006, 2009)  
Duke University (2006, 2010, 2011)  
Educational Testing Service (2006, 2010)  
Google (2004, 2006, 2007)  
Harvard University (2009)  
IBM Almaden (2002)  
Institute for Pure and Applied Mathematics (2006, 2007, 2010)  
Johns Hopkins University (2006, 2012)  
Joint Statistics Meetings (2006, 2011)  
Knowledge Discovery and Data mining (2011)  
Massachusetts Institute of Technology (2003, 2007, 2008, 2009, 2012)  
Microsoft Research (2007)  
Nature Publishing Group (2008)  
New York University (2005, 2010, 2011)  
New York Academy of Sciences (2010)  
New York Machine Learning Meetup (2010)  
Political Methodology (2011)  
Princeton University (2005)  
Purdue University (2012)  
Rutgers University (2007, 2009)  
Stanford University (2011)  
U.C. Berkeley (2011)  
U.C. Irvine (2007)  
U.C. Los Angeles. (2005)  
U.C. San Diego (2005)  
U.C. Santa Cruz (2005)  
Univeristy if Illinois (2006)  
University College London (2008)  
University of Cambridge (2008)  
University of Cambridge Machine Learning Summer School (2009)  
University of Chicago (2008, 2011)

University of Connecticut (2006, 2007)  
University of Massachusetts Amherst (2004)  
University of Pennsylvania (2006, 2011)  
University of Tennessee (2011)  
University of Texas (2012)  
University of Toronto (2003)  
Yale University (2011)  
Xerox PARC (2002)