

# David M. Blei

Professor

Columbia University

Department of Computer Science

Department of Statistics

Columbia Data Science Institute

New York, NY

Email: [david.blei@columbia.edu](mailto:david.blei@columbia.edu)

Homepage: <http://www.cs.columbia.edu/~blei>

## EDUCATION

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

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

Advisor: Michael Jordan

## EMPLOYMENT

Professor, Departments of Statistics and Computer Science, Columbia University, 2014–

Associate Professor, Department of Computer Science, Princeton University, 2011–2014

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

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

Advisor: John Lafferty

## AWARDS

Guggenheim Fellowship, 2017

Fellow of the Institute of Mathematical Statistics, 2017

ICML Test of Time Award (for “Dynamic Topic Models”), 2016

Presidential Award for Outstanding Teaching, Honorable Mention, 2016

Fellow of the Association of Computing Machinery, 2015

SIGIR Test of Time Award Honorable Mention (for “Modeling Annotated Data”), 2015

ACM Prize in Computing, 2013

Blavatnik Award for Young Scientists: Faculty Winner, 2013

Presidential Early Career Award for Scientists and Engineers (PECASE), 2011

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,  
2006, 2007, 2008, 2009, 2010, 2011, 2012, 2013

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

## PUBLICATIONS

### *Journal Articles*

1. R. Ranganath and D. Blei. Correlated random measures. *Journal of the American Statistical Association*, to appear.
2. D. Blei and P. Smyth. Science and data science. *Proceedings of the National Academy of Sciences*, 114(33):8689–8692, 2017.
3. D. Blei, A. Kucukelbir, and J. McAuliffe. Variational inference: A review for statisticians. *Journal of the American Statistical Association*, 112(518): 859–877, 2017.
4. A. Kucukelbir, D. Tran, A. Gelman, and D. Blei. Automatic differentiation variational inference. *Journal of Machine Learning Research*, 18(14):1–45, 2017.
5. D. Tran and D. Blei. Comment on “Fast approximate inference for arbitrarily large semiparametric regression models via message passing.” *Journal of the American Statistical Association*, 112(517):156–158, 2017.
6. P. Gopalan, W. Hao, D. Blei, and J. Storey. Scaling probabilistic models of genetic variation to millions of humans. *Nature Genetics*, 48 (1587–1590), 2016.
7. D. Mimno, D. Blei, and B. Engelhardt. Posterior predictive checks to quantify lack-of-fit in admixture models of latent population structure. *Proceedings of the National Academy of Sciences*, 112(26), 2015.

8. A. Perotte, R. Ranganath, J. Hirsch, D. Blei, and N. Elhadad. Risk prediction for chronic kidney disease progression using heterogeneous electronic health record data and time series analysis. *Journal of the American Medical Informatics Association*, 22 (4), 2015.
9. J. Paisley, C. Wang, D. Blei, and M. Jordan. A nested HDP for hierarchical topic modeling. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.
10. G. Polatkan, M. Zhou, L. Carin, D. Blei, and I. Daubechies. A Bayesian nonparametric approach to image super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.
11. S. Gershman, P. Frazier, and D. Blei. Distance dependent infinite latent feature models. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 37 (2), 2015.
12. D. Blei. Build, Compute, Critique, Repeat: Data Analysis with Latent Variable Models. *Annual Review of Statistics and Its Application*, 1 203–232, 2014.
13. S. Gershman, D. Blei, K. Norman, and P. Sederberg. Decomposing spatiotemporal brain patterns into topographic latent sources. *NeuroImage*, 98:91–102, 2014.
14. J. Manning, R. Ranganath, K. Norman, and D. Blei. Topographic factor analysis: A Bayesian model for inferring brain networks from neural data. *PLoS ONE*, 9(5), 2014.
15. P. Gopalan and D. Blei. Efficient discovery of overlapping communities in massive networks. *Proceedings of the National Academy of Sciences*, 110 (36) 14534–14539, 2013.
16. M. Hoffman, D. Blei, C. Wang, and J. Paisley. Stochastic variational inference. *Journal of Machine Learning Research*, 14:1303–1347, 2013.
17. C. Wang and D. Blei. Variational inference in nonconjugate models. *Journal of Machine Learning Research*, 14:1005–1031, 2013.
18. P. DiMaggio, M. Nag, and D. Blei. Exploiting affinities between topic modeling and the sociological perspective on culture: Application to newspaper coverage of U.S. government arts funding. *Poetics*, 41:6, 2013.
19. D. Blei. Topic modeling and digital humanities. *Journal of Digital Humanities*, 2(1), 2013.
20. D. Blei. Comment on multinomial inverse regression for text analysis. *Journal of the American Statistical Association*, 108 (503) 771–772, 2013.
21. B. Chen, G. Polatkan, G. Sapiro, D. Blei, D. Dunson, L. Carin. Deep learning with hierarchical convolutional factor analysis. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35 (8), 2013.
22. J. Paisley, C. Wang and D. Blei. The discrete infinite logistic normal distribution. *Bayesian Analysis*, 7(2):235–272, 2012.
23. D. Blei. Probabilistic topic models. *Communications of the ACM*, 55(4):77–84, 2012.

24. S. Gershman and D. Blei. A tutorial on Bayesian nonparametric models. *Journal of Mathematical Psychology*, 56:1–12, 2012.
25. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. *Journal of Machine Learning Research*, 12:2461–2488, 2011.
26. L. Hannah, D. Blei and W. Powell. Dirichlet process mixtures of generalized linear models. *Journal of Machine Learning Research*, 12:1923–1953, 2011.
27. S. Gershman, D. Blei, F. Pereira, and K. Norman. A topographic latent source model for fMRI data. *NeuroImage*, 57:89–100, 2011.
28. D. Blei, L. Carin, and D. Dunson. Probabilistic topic models. *Signal Processing*, 27(6):55–65, 2010.
29. D. Blei, T. Griffiths, and M. Jordan. The nested Chinese restaurant process and Bayesian nonparametric inference of topic hierarchies. *Journal of the ACM*, 57(2):1–30, 2010.
30. J. Chang and D. Blei. Hierarchical relational models for document networks. *Annals of Applied Statistics*, 4(1), 2010.
31. S. Gershman, D. Blei, and Y. Niv. Context, learning and extinction. *Psychological Review*, 117(1):197–209, 2010.
32. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.
33. D. Blei and J. Lafferty. A correlated topic model of Science. *Annals of Applied Statistics*, 1(1):17–35, 2007.
34. D. Blei and S. Fienberg. Discussion of model-based clustering for social networks. *Journal of the Royal Statistical Society, Series A*, 170:332, 2007.
35. J. McAuliffe, D. Blei, and M. Jordan. Nonparametric empirical Bayes for the Dirichlet process mixture model. *Statistics and Computing*, 16(1):5–14, 2006.
36. Y. Teh, M. Jordan, M. Beal, and D. Blei. Hierarchical Dirichlet processes. *Journal of the American Statistical Association*, 101(476):1566–1581, 2006.
37. 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.
38. D. Blei and M. Jordan. Variational inference for Dirichlet process mixtures. *Journal of Bayesian Analysis*, 1(1):121–144, 2005.
39. 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.
40. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. *Journal of Machine Learning Research*, 3:993–1022, January 2003.

## Conference Articles

41. R. Ranganath, D. Tran, and D. Blei. Hierarchical implicit models and likelihood-free variational inference. *In Neural Information Processing Systems*, 2017.
42. L. Liu, F. Ruiz, and D. Blei. Context selection for embedding models. *In Neural Information Processing Systems*, 2017.
43. A. Dieng, D. Tran, R. Ranganath, J. Paisley, and D. Blei. Variational inference via  $\chi$ -upper bound minimization *In Neural Information Processing Systems*, 2017.
44. M. Rudolph, F. Ruiz, and D. Blei. Structured embedding models for grouped Data. *In Neural Information Processing Systems*, 2017.
45. A. Kucukelbir, Y. Wang, and D. Blei. Evaluating Bayesian models with posterior dispersion indices. *In International Conference on Machine Learning*, 2017.
46. Y. Wang, A. Kucukelbir, and D. Blei. Robust probabilistic modeling with Bayesian data reweighting. *In International Conference on Machine Learning*, 2017.
47. D. Tran, M. Hoffman, R. Saurous, E. Brevdo, K. Murphy, and D. Blei. Deep probabilistic programming. *In International Conference on Learning Representations*, 2017.
48. C. Naesseth, F. Ruiz, S. Linderman, and D. Blei. Reparameterization gradients through acceptance-rejection sampling algorithms. *In Artificial Intelligence and Statistics*, 2017. **Best Student Paper Award.**
49. S. Linderman, A. Miller, R. Adams, M. Johnson, D. Blei, and L. Paninski. Recurrent switching linear dynamical systems. *In Artificial Intelligence and Statistics*, 2017.
50. A. Chaney, H. Wallach, and D. Blei. Detecting and Characterizing Events. *In Empirical Methods in Natural Language Processing*, 2016.
51. F. Ruiz, M. Titsias, D. Blei. The generalized reparameterization gradient. *In Neural Information Processing Systems*, 2016.
52. R. Ranganath, D. Tran, J. Altsaar, and D. Blei. Operator variational inference. *In Neural Information Processing Systems*, 2016.
53. M. Rudolph, F. Ruiz, S. Mandt, and D. Blei. Exponential family embeddings. *In Neural Information Processing Systems*, 2016.
54. R. Ranganath, A. Perotte, N. Elhadad, and D. Blei. Deep survival analysis. *Machine Learning for Health Care*, 2016.
55. D. Liang, J. Altsaar, L. Charlin, and D. Blei. Factorization meets the item embedding: Regularizing matrix factorization with item co-occurrence. *In ACM Conference on Recommendation Systems*, 2016.
56. F. Ruiz, M. Titsias, and D. Blei. Overdispersed black-box variational inference. *In Uncertainty in Artificial Intelligence*, 2016.

57. R. Ranganath, D. Tran, and D. Blei. Hierarchical variational models. In *International Conference on Machine Learning*, 2016.
58. A. Schein, M. Zhou, D. Blei, and H. Wallach. Bayesian Poisson Tucker decomposition for learning the structure of international relations. In *International Conference on Machine Learning*, 2016.
59. S. Mandt, M. Hoffman, and D. Blei. A variational analysis of stochastic gradient algorithms. In *International Conference on Machine Learning*, 2016.
60. D. Tran, R. Ranganath, and D. Blei. The variational Gaussian process. In *International Conference on Learning and Representation*, 2016.
61. D. Liang, L. Charlin, J. McInerney, D. Blei. Modeling user exposure in recommendation. In *International World Wide Web Conference*, 2016.
62. M. Rudolph, J. Ellis, and D. Blei. Objective variables for probabilistic revenue maximization in second-price auctions with reserve. In *International World Wide Web Conference*, 2016.
63. S. Mandt, J. McInerney, F. Abrol, R. Ranganath, and D. Blei. Multicanonical stochastic variational inference. In *Artificial Intelligence and Statistics*, 2016.
64. D. Tran, D. Blei, and E. Airoldi. Variational inference with copula augmentation. In *Neural Information Processing Systems*, 2015.
65. A. Kucukelbir, R. Ranganath, A. Gelman, and D. Blei. Automatic variational inference in Stan. In *Neural Information Processing Systems*, 2015.
66. J. McInerney, R. Ranganath, and D. Blei. The population posterior and Bayesian inference on streams. In *Neural Information Processing Systems*, 2015.
67. L. Charlin, R. Ranganath, J. McInerney, and D. Blei. Dynamic Poisson factorization. In *ACM Conference on Recommendation Systems*, 2015.
68. A. Chaney and D. Blei and T. Ellassi-Rad. A probabilistic model for using social networks in personalized item recommendation. In *ACM Conference on Recommendation Systems*, 2015.
69. P. Gopalan, J. Hofman, and D. Blei. Scalable recommendation with hierarchical Poisson factorization. In *Uncertainty in Artificial Intelligence*, 2015.
70. R. Ranganath, A. Perotte, N. Elhadad, and D. Blei. The survival filter: Joint survival analysis with a latent time series. In *Uncertainty in Artificial Intelligence*, 2015.
71. A. Kucukelbir and D. Blei. Population empirical Bayes. In *Uncertainty in Artificial Intelligence*, 2015.
72. A. Schein, J. Paisley, D. Blei, and H. Wallach. Bayesian Poisson tensor factorization for inferring multilateral relations from sparse dyadic event counts. In *Knowledge Discovery and Data Mining*, 2015.
73. M. Hoffman and D. Blei. Structured stochastic variational inference. In *Artificial Intelligence and Statistics*, 2015.

74. R. Ranganath, L. Tang, L. Charlin, and D. Blei. Deep exponential families. In *Artificial Intelligence and Statistics*, 2015.
75. N. Houlsby and D. Blei. A filtering approach to stochastic variational inference. In *Neural Information Processing Systems*, 2014.
76. S. Mandt and D. Blei. Smoothed gradients for stochastic variational inference. In *Neural Information Processing Systems*, 2014.
77. P. Gopalan, L. Charlin, and D. Blei. Content based recommendations with Poisson factorization. In *Neural Information Processing Systems*, 2014.
78. R. Ranganath, S. Gerrish, and D. Blei. Black box variational inference. In *Artificial Intelligence and Statistics*, 2014.
79. P. Gopalan, F. Ruiz, R. Ranganath, and D. Blei. Bayesian nonparametric Poisson factorization for recommendation systems. In *Artificial Intelligence and Statistics*, 2014.
80. M. Rabinovich and D. Blei. The inverse regression topic model. In *International Conference on Machine Learning*, 2014.
81. P. Gopalan, C. Wang and D. Blei. Modeling overlapping communities with node popularities. In *Neural Information Processing Systems*, 2013.
82. D. Kim, P. Gopalan, D. Blei, and E. Sudderth. Efficient online inference for Bayesian nonparametric relational models. In *Neural Information Processing Systems*, 2013.
83. R. Ranganath, C. Wang and D. Blei. An adaptive learning rate for stochastic variational inference. In *International Conference on Machine Learning*, 2013.
84. P. Gopalan, D. Mimno, S. Gerrish, M. Freedman, and D. Blei. Scalable inference of overlapping communities. In *Neural Information Processing Systems*, 2012.
85. S. Gerrish and D. Blei. How they vote: Issue-adjusted models of legislative behavior. In *Neural Information Processing Systems*, 2012.
86. C. Wang and D. Blei. Truncation-free online variational inference for Bayesian nonparametric models. In *Neural Information Processing Systems*, 2012.
87. J. Paisley, D. Blei and M. Jordan. Variational Bayesian inference with stochastic search. In *International Conference On Machine Learning*, 2012.
88. D. Mimno, M. Hoffman and D. Blei. Sparse stochastic inference for latent Dirichlet allocation. In *International Conference On Machine Learning*, 2012.
89. S. Gershman, M. Hoffman and D. Blei. Nonparametric variational inference. In *International Conference On Machine Learning*, 2012.
90. A. Chaney and D. Blei. Visualizing topic models. In *International AAAI Conference on Weblogs and Social Media*, 2012.

91. J. Paisley, D. Blei, and M. Jordan. Stick-breaking beta processes and the Poisson process. In *Artificial Intelligence and Statistics*, 2012.
92. 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.
93. C. Wang and D. Blei. Collaborative topic modeling for recommending scientific articles. In *Knowledge Discovery and Data Mining*, 2011. **Best Student Paper Award.**
94. D. Mimno and D. Blei. Bayesian checking for topic models. In *Empirical Methods in Natural Language Processing*, 2011.
95. S. Gerrish and D. Blei. Predicting legislative roll call from text. In *International Conference on Machine Learning*, 2011. **Distinguished Application Paper Award.**
96. J. Paisley, D. Blei, and L. Carin. Variational inference for stick-breaking beta process priors. In *International Conference on Machine Learning*, 2011.
97. 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.**
98. C. Wang, J. Paisley and D. Blei. Online variational inference for the hierarchical Dirichlet process. In *Artificial Intelligence and Statistics*, 2011.
99. M. Hoffman, D. Blei, and F. Bach. On-line learning for latent Dirichlet allocation. In *Neural Information Processing Systems*, 2010.
100. L. Hannah, W. Powell, and D. Blei. Nonparametric density estimation for stochastic optimization with an observable state variable. In *Neural Information Processing Systems*, 2010.
101. D. Blei and P. Frazier. Distance dependent Chinese restaurant processes. In *International Conference on Machine Learning*, 2010.
102. S. Gerrish and D. Blei. A language-based approach to measuring scholarly impact. In *International Conference on Machine Learning*, 2010.
103. M. Hoffman, D. Blei, and P. Cook. Bayesian nonparametric matrix factorization for recorded music. In *International Conference on Machine Learning*, 2010.
104. 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.
105. L. Hannah, D. Blei, and W. Powell. Dirichlet process mixtures of generalized linear models. In *Artificial Intelligence and Statistics*, 2010.
106. 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.
107. 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.



108. S. Cohen, D. Blei, and N. Smith. Variational inference for adaptor grammars. In *North American Chapter of the Association for Computational Linguistics*, 2010.
109. C. Wang and D. Blei. Decoupling sparsity and smoothness in the discrete hierarchical Dirichlet process. In *Neural Information Processing Systems*, 2009.
110. C. Wang and D. Blei. Variational inference for the nested Chinese restaurant process. In *Neural Information Processing Systems*, 2009.
111. 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.
112. 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.**
113. J. Chang, J. Boyd-Graber, and D. Blei. Connections between the lines: Augmenting social networks with text. In *Knowledge Discovery and Data Mining*, 2009.
114. J. Boyd-Graber and D. Blei. Multilingual topic models for unaligned text. In *Uncertainty in Artificial Intelligence*, 2009.
115. J. Chang and D. Blei. Relational topic models for document networks. In *Artificial Intelligence and Statistics*, 2009.
116. C. Wang, B. Thiesson, C. Meek, and D. Blei. Markov topic models. In *Artificial Intelligence and Statistics*, 2009.
117. 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.
118. 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.**
119. 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.
120. C. Wang, D. Blei, and L. Fei-Fei. Simultaneous image classification and annotation. In *Computer Vision and Pattern Recognition*, 2009.
121. I. Mukherjee and D. Blei. Relative performance guarantees for approximate inference in latent Dirichlet allocation. In *Neural Information Processing Systems*, 2008.
122. J. Boyd-Graber and D. Blei. Syntactic topic models. In *Neural Information Processing Systems*, 2008.
123. E. Airoldi, D. Blei, S. Fienberg, and E. Xing. Mixed membership stochastic blockmodels. In *Neural Information Processing Systems*, 2008.
124. C. Wang, D. Blei, and D. Heckerman. Continuous time dynamic topic models. In *Uncertainty in Artificial Intelligence (UAI)*, 2008.

125. 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.
126. M. Hoffman, P. Cook, and D. Blei. Data-driven recomposition using the hierarchical Dirichlet process hidden Markov model. In *International Computer Music Conference*, 2008.
127. M. Dudik, D. Blei, and R. Schapire. Hierarchical maximum entropy density estimation. In *Proceedings of the 28th International Conference on Machine Learning*, 2007.
128. W. Li, D. Blei, and A. McCallum. Nonparametric Bayes pachinko allocation. In *The 23rd Conference on Uncertainty in Artificial Intelligence*, 2007.
129. D. Kaplan and D. Blei. A computational approach to style in American poetry. In *IEEE Conference on Data Mining*, 2007.
130. D. Blei and J. McAuliffe. Supervised topic models. In *Neural Information Processing Systems*, 2007.
131. J. Boyd-Graber, D. Blei, and X. Zhu. A topic model for word sense disambiguation. In *Empirical Methods in Natural Language Processing*, 2007.
132. D. Blei and J. Lafferty. Correlated topic models. In *Neural Information Processing Systems*, 2006.
133. D. Blei and J. Lafferty. Dynamic topic models. In *International Conference on Machine Learning*, 2006. **ICML 2016 Test of Time Award.**
134. T. Griffiths, M. Steyvers, D. Blei, and J. Tenenbaum. Integrating topics and syntax. In *Neural Information Processing Systems*, 2005.
135. D. Blei and M. Jordan. Variational methods for the Dirichlet process. In *International Conference on Machine Learning*, 2004.
136. D. Blei and M. Jordan. Modeling annotated data. In *ACM SIGIR Conference on Research and Development in Information Retrieval*, 2003. **SIGIR Test of Time Award (Honorable Mention).**
137. 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.**
138. D. Blei, A. Ng, and M. Jordan. Latent Dirichlet allocation. In *Neural Information Processing Systems*, 2002.
139. D. Blei, J. Bagnell, and A. McCallum. Learning with scope, with application to information extraction and classification. In *Uncertainty in Artificial Intelligence*, 2002.
140. 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*

141. E. Airoldi, D. Blei, E. Erosheva, and S. Fienberg. Introduction to Mixed Membership Models and Methods. In *Handbook of Mixed-Membership Models and Their Applications*. Chapman & Hall/CRC, 2014.
142. S. Williamson, C. Wang, K. Heller, and D. Blei. Nonparametric mixed membership models using the IBP compound Dirichlet process. In K. Mengerson, C. Robert, and D. Titterington, editors, *Mixture Estimation and Applications*. John Wiley and Sons, 2011.
143. 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.
144. 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.
145. 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 Volumes*

146. E. Airoldi, D. Blei, E. Erosheva, and S. Fienberg, editors. *Handbook of Mixed-Membership Models and Their Applications*. Chapman and Hall/CRC, 2014.
147. 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.

## 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.

## AWARDED GRANTS

1. *Next-Generational Variational Methods: Active Inference, Streaming Inference, and Assessing Model Fitness* (PI). Office of Naval Research. \$1M. 2015-2020.
2. *Extracting Mathematical Knowledge from the Scientific Literature: Statistical Machine Learning Models and Tools* (Co-PI). Sloan Foundation. \$900K. 2015-2018.

3. *Deciphering the Cortex: Circuit Inference from Large-Scale Brain Activity Data* (Co-PI). Defense Advanced Research Projects Agency. \$1M. 2015-2016.
4. *Estimating Multidimensional Influence in Science and Scholarship* (PI). Templeton Foundation. \$100K. 2014-2015.
5. *The Next Generation of Probabilistic Programming: Massive Data, Data Streams, and Model Diagnostics* (PI). Defense Advanced Research Project Agency. \$1.8M. 2013-2017.
6. *BIGDATA: Discovery and Social Analytics for Large-Scale Scientific Literature* (Co-PI). National Science Foundation. \$997K. 2013-2015.
7. *Scalable Topic Modeling: Online Learning, Diagnostics, and Recommendation* (PI). Office of Naval Research. \$510K. 2011-2014.
8. *Text, Neuroimaging, and Memory: Unified Models of Corpora and Cognition* (PI). National Science Foundation. \$730K. 2010-2013.
9. *Non-Parametric Bayesian Analysis of Heterogeneous Data* (PI). Air Force Office of Scientific Research. \$360K. 2009-2012.
10. *Dynamic and Supervised Topic Models for Literature-Based Discovery* (PI). Office of Naval Research. \$300K. 2008-2011.
11. *CAREER: New Directions in Probabilistic Topic Models* (PI). National Science Foundation. \$550K. 2008-2013.
12. *Interactive Discovery and Semantic Labeling of Patterns in Spatial Data* (Co-PI). National Science Foundation. \$500K. 2009-2012.

## PROFESSIONAL ACTIVITIES

### Senior Program Committee

International Conference on Machine Learning, 2015 Program Co-chair

International Conference on Machine Learning  
(2008, 2009, 2010, 2011, 2012, 2013, 2014, 2016, 2017)

Neural Information Processing Systems (2009, 2010, 2014)

Bayesian Nonparametrics (2017)

Artificial Intelligence and Statistics (2008, 2012, 2014)

### Associate Editor and Editorial Board

Journal of the American Statistical Association (2017–present)

Journal of Machine Learning Research (2008–present)

IEEE Transactions on Pattern Analysis and Machine Intelligence (2014–2016)

Statistics and Computing (2009–2013)

Chapman Hall Series on Computer Science and Data Analysis (2008–present)

### **Journal Reviewing and Editorial Board**

Proceedings of the National Academy of Science  
Science Magazine  
Journal of Machine Learning Research  
Journal of the American Statistical Association  
Journal of the Royal Statistical Society  
Foundations and Trends in Machine Learning  
Annals of Applied Statistics  
Bayesian Analysis  
Statistics and Computing  
Machine Learning Journal  
Journal of Artificial Intelligence Research  
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

### **Conference Reviewing**

Neural Information Processing Systems (2005, 2006, 2007, 2008, 2011, 2012, 2013)  
Artificial Intelligence and Statistics (2005, 2007, 2010)  
International Conference on Machine Learning (2006, 2007)  
Uncertainty in Artificial Intelligence (2005, 2006, 2007)  
Association of Computational Linguistics (2008)  
Empirical Methods in Natural Language Processing (2007)  
Association of Artificial Intelligence (2007)  
International Joint Conference on Artificial Intelligence (2005)  
SIGIR Conference on Information Retrieval (2005)  
Knowledge Discovery and Data Mining (2005)

### **Grant Reviewing**

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

### **Columbia University**

Director of Graduate Studies, Data Science Institute

Executive Committee of the Data Science Institute  
Education Committee of the Data Science Institute  
Senior Search Committee of the Data Science Institute  
Digital Humanities Task Force  
Internal Review Committee for the Economics Department  
Search committee chair (Statistics)  
Data Science committee (Statistics)  
PhD committee (Computer Science)  
Distinguished lecture committee (Computer Science)

### **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**

“Computational Challenges in Machine Learning” (Simons Institute, 2017)  
“Mathematical Analysis of Cultural Expressive Forms” (IPAM, 2016)  
“Advanced in Approximate Bayesian Inference” (NIPS, 2015)  
“Advances in Variational Inference” (NIPS, 2014)  
“Topic Models: Computation, Application, and Evaluation” (NIPS, 2013)  
“Statistics and Machine Learning at Princeton” (Princeton University, 2011)  
“Applications of Topic Modeling” (NIPS, 2008)  
“Statistical Network Analysis” (ICML, 2005)  
“Syntax and Semantics” (NIPS, 2003)

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

Prem Gopalan (2014); Voleon Capital  
Sean Gerrish (2012); Data Scientist, Google  
Samuel Gershman (2012); Assistant Professor, Harvard University  
Gungor Polatkan (2012, co-advised with Ingrid Dubeiches); Data Scientist, Twitter  
Chong Wang (2012); Data Scientist, Baidu research  
Jonathan Chang (2011); Data Scientist, Facebook

Matthew Hoffman (2010, co-advised with Perry Cook); Data Scientist, Adobe Systems

Lauren Hannah (2010, co-advised with Warren Powell); Assistant Professor, Columbia University

Jordan Boyd-Graber (2009); Assistant Professor, University of Colorado

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

Allison Chaney (expected 2016)

Rajesh Ranganath (expected 2016)

Jaan Aaltosar (expected 2018)

Dustin Tran (expected 2018)

Adji Deng (expected 2019)

Maja Rudolph (expected 2019)

Kui Tang (expected 2019)

Yixin Wang (expected 2019)

#### **Postdoctoral Fellows**

David Mimno; Assistant Professor, Cornell University

John Paisley; Assistant Professor, Columbia University

Laurent Charlin; Assistant Professor, University of Montreal

Jeremy Manning; Assistant Professor, Dartmouth College

Stephan Mandt; Disney Research, Pittsburgh PA

James McInerney; Spotify, New York NY

Kriste Krstovski

Alp Kucukelbir

Scott Linderman

Liping Liu

Francisco Ruiz

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

Yuening Hu (University of Maryland, 2014)

Edouard Graves (INRIA, 2013)

Mohammad Emtiyaz Khan (University of British Columbia, 2012)

Alex Lorbert (Princeton, 2012)

David Mimno (University of Massachusetts, 2012)

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 (University of Massachusetts, 2007)  
Miroslav Dudik (Princeton, 2007)

### **Professional Memberships**

Association of Computing Machinery  
Institute for Mathematical Statistics  
American Statistical Association  
Bernoulli Society

### **Advising and Consulting**

Scientific Advisor, Undecidable Labs (2015–present)  
Scientific Advisor, Gamalon Technologies (2015–present)  
Scientific Advisor, Recruit Artificial Intelligence Laboratories (2015–present)  
Scientific Advisor, Liftlighter (2015–present)  
Scientific Advisor, MyRoll (2015–2016)  
Scientific Advisor, VoxGov (2014–2016)  
Steering Committee, Declassification Engine (2014–present)  
Scientific Advisor, Applied Communications Sciences (2012–2013)  
Consulting Researcher, Microsoft Research (2013, 2014)  
External Reviewer, NYU Data Sciences PhD program (2013)  
Scientific Advisor, Chomp (2011–2012, Acquired by Apple)

## **INVITED TALKS**

### **Year 2017**

1. University of California Los Angeles, Statistics Seminar
2. European Association of Computational Linguistics, Keynote Lecture
3. University of Toronto, AI Institute Lecture
4. University of Michigan, Data Science Seminar
5. University of Michigan, Statistics Seminar
6. University of Edinburgh, Distinguished Lecture



7. Workshop for Young European Statisticians, Invited Tutorial

**Year 2016**

8. University of California Berkeley, Neyman Seminar
9. University of California Berkeley, Focused Research Group
10. University of Tokyo, Machine Learning Seminar
11. Stonybrook University Computer Science, Distinguished Lecture
12. Keynote Speaker: Recruit Tokyo Data Science Conference
13. Columbia University, Sustainable Development Seminar
14. AT&T Research, Statistics Seminar
15. Columbia University, Biostatistics Seminar
16. Broad Institute (Cambridge, MA)
17. MIT Statistics Seminar (Cambridge, MA)
18. IPAM Workshop on the Mathematical Analysis of Cultural Expressive Forms (Los Angeles, CA)
19. AIG Data Science Seminar (New York, NY)
20. Isaac Newton Institute Workshop on the Limits of Graph Statistics (Cambridge, UK)
21. Cambridge University Machine Learning Seminar
22. Microsoft Research Conference on Machine Learning and Economics (New York, NY)
23. University of Illinois Distinguished Lecture in Computer Science
24. University of Chicago conference: "Machine Learning: What's In It For Economics?"
25. Sandia National Laboratories Distinguished Lecture (Albuquerque, NM)
26. Latent Variables 2016, Plenary Speaker (Columbia, SC)
27. Spotify (New York, NY)
28. Princeton Day of Statistics (Princeton, NJ)
29. Two Sigma (New York, NY)
30. Invited NIPS Tutorial on Variational Inference (Barcelona, Spain)
31. NIPS Workshop on Bayesian Deep Learning (Barcelona, Spain)
32. NIPS Workshop on Causal Inference (Barcelona, Spain)

**Year 2015**

33. 2015 Joint Statistical Meetings, Session on Large-Scale Inference, Seattle
34. Center for Mathematical Research, Montreal Canada
35. Princeton Conference on Text Analysis and the Social Sciences
36. Facebook Artificial Intelligence Research, New York City
37. 10th Conference on Bayesian Nonparametric Statistics
38. Amazon Inc., Machine Learning @ Amazon

39. University of Chicago, Statistics Colloquium
40. University of Connecticut, Statistics Colloquium
41. Brown University, Computer Science Colloquium
42. Brown University, Applied Mathematics Colloquium
43. Rutgers University, Innovations in Statistics and Data Analysis
44. Office of Naval Research, Naval Future Force
45. Microsoft Research, Data Science Seminar
46. NIPS Workshop on Black Box Inference
47. Indiana University, Distinguished Lecture
48. New York University, Text as Data Series
49. Columbia University, IGERT Distinguished Speaker
50. University of Massachusetts, Data Science Distinguished Lecture
51. University of Massachusetts, Machine Learning Colloquium
52. Johns Hopkins University, Applied Mathematics Seminar

**Year 2014**

53. Keynote Speaker, IBM Research Colloquium on Cognitive Computing (Haifa)
54. Keynote Speaker, IBM Research Machine Learning Seminar (Haifa)
55. Keynote Speaker, DIMACS Mixer
56. George Mason University, Distinguished Lecture
57. University of Washington, Distinguished Lecture
58. Simons Foundation, Frontiers of Data Science
59. Microsoft Research, Redmond
60. Keynote speaker, Uncertainty in Artificial Intelligence
61. Data, Society, and Inference Seminar at Stanford University
62. IPAM Workshop on “Stochastic Gradient Methods”

**Year 2013**

63. Andresseen-Horowitz Academic Summit (Menlo Park, California)
64. Applied Communications Sciences (New Jersey)
65. Bloomberg LLC Distinguished Lecture (New York, New York)
66. City University of New York Computer Science Colloquium
67. Columbia University Data Sciences Institute
68. Duke University Machine Learning Seminar
69. Google Tech Talk (Mountain View, California)
70. INRIA Machine Learning Colloquium (Paris, France)

71. Microsoft Research New York
72. NIPS Workshop on “Probabilistic Modeling of Big Data” (Stateline, Nevada)
73. Stanford University Statistics Seminar
74. Temple University Computer Science Colloquium
75. Xerox Research 20th Anniversary Distinguished Lecture (Grenoble, France)

**Year 2012**

76. Carnegie-Mellon University Machine Learning Department
77. Harvard University Computer Science Colloquium
78. Harvard University Machine Learning Seminar
79. International Conference on Machine Learning, Invited Tutorial (Edinburgh, Scotland)
80. Johns Hopkins University Computer Science Colloquium
81. Machine Learning Summer School, Invited Lecture Series (Kyoto, Japan)
82. Massachusetts Institute of Technology
83. Jamon Lecture (Stateline, Nevada)
84. New York University Machine Learning Seminar
85. Purdue University Computer Science Colloquium
86. University of California San Diego Computer Science Colloquium
87. University of Texas Austin Statistics Seminar

**Year 2011**

88. Boston University Computer Engineering Colloquium
89. Conference on Political Methodology (Princeton, New Jersey)
90. Duke University Statistics Seminar
91. IBM Watson (Yorktown Heights, NY)
92. ISBA Workshop on Bayesian Nonparametrics (Veracruz, Mexico)
93. Joint Statistical Meetings (Miami, Florida)
94. Knowledge Discovery and Data Mining, Invited Tutorial (San Diego, California)
95. New York University Statistics Seminar
96. Stanford University Machine Learning Seminar
97. Stanford University Statistics Seminar
98. University of California Berkeley Neyman Seminar
99. University of Chicago
100. University of Pennsylvania Computer Science Colloquium
101. University of Tennessee Computer Science Colloquium
102. Yale University Applied Mathematics

103. Yale University Statistics

**2010**

104. Columbia University Computer Science Colloquium

105. Cornell University Computer Science Colloquium

106. Duke University Statistics Seminar

107. Educational Testing Service Seminar (Princeton, New Jersey)

108. Institute for Pure and Applied Mathematics (Los Angeles, California)

109. New York Academy of Sciences Machine Learning Symposium

110. New York Machine Learning Meetup

111. New York University Computer Science Colloquium

**2009**

112. Carnegie Mellon University Machine Learning Seminar

113. Carnegie Mellon University Statistics Seminar

114. Center for Discrete Mathematics and Theoretical Computer Science (Rutgers)

115. Columbia University Statistics Seminar

116. Machine Learning Summer School (Cambridge, England)

117. New Directions in Analyzing Text as Data (Cambridge, Massachusetts)

118. Rutgers University Statistics Seminar

**Before 2009**

119. AAI Spring Symposium (2002)

120. BAE Systems (2008)

121. Brown University (2005)

122. Carnegie Mellon University (2003)

123. Columbia University (2007)

124. Cornell University (2007)

125. Center for Discrete Mathematics and Theoretical Computer Science (2008)

126. Educational Testing Services (2006)

127. Duke University (2006)

128. Google Research (2004)

129. Google Research (2006)

130. Google Research (2007)

131. The Hebrew University (2008)

132. IBM Almaden (2002)

133. Institute for Pure and Applied Mathematics (2006)

134. Johns Hopkins University (2006)
135. Joint Statistical Meetings (2006)
136. Massachusetts Institute of Technology (2003)
137. Massachusetts Institute of Technology (2007)
138. Massachusetts Institute of Technology (2008)
139. Microsoft Research Redmond (2007)
140. Nature Publishing Group (2008)
141. New York University Computer Science (2005)
142. Princeton University Computer Science (2005)
143. Rutgers University (2007)
144. University of California Irvine (2007)
145. University of California Los Angeles (2005)
146. University of California San Diego (2005)
147. University of California Santa Cruz (2005)
148. University of Illinois Champagne-Urbana (2006)
149. University College London (2008)
150. University of Cambridge (2008)
151. University of Connecticut (2006)
152. University of Connecticut (2007)
153. University of Massachusetts Amherst (2004)
154. University of Pennsylvania (2006)
155. University of Toronto (2003)
156. Xerox PARC (2002)