

## Eric Poe Xing

(412) 268-2559 (Office)  
(412) 268-3431 (Fax)  
epxing@cs.cmu.edu  
<http://www.cs.cmu.edu/~epxing>

4127 Wean Hall, Machine Learning Dept.  
School of Computer Science  
Carnegie Mellon University  
Pittsburgh, PA 15213

---

### Positions

- **Assistant Professor**, Machine Learning Department & Language Technology Institute & Computer Science Department, School of Computer Science, *Carnegie Mellon University*. (September 1, 2004 – )
- **Adjunct Assistant Professor**, Department of Computational Biology, School of Medicine, *University of Pittsburgh*. (March 1, 2006 – )

### Education

- **University of California, Berkeley**, Ph.D. in Computer Science (1999–2004).  
Thesis title: *Probabilistic graphical models and algorithms for genomic analysis*  
Research advisors: Profs. Richard Karp, Michael Jordan and Stuart Russell
- **Rutgers University**, Ph.D. in Molecular Biology and Biochemistry (1994–1999).  
Thesis title: *The Molecular Mechanism of Human Esophageal Carcinogenesis*  
Research advisor: Prof. Chung S. Yang
- **Rutgers University**, M.Sc. in Computer Science (1996–1998).  
Thesis title: *Segmentation and Combinatorial Clustering Analysis of Ribosomal RNA Sequences*  
Research advisor: Prof. Casimir Kulikowski
- **Tsinghua University**, B.Sc. in Physics and Biology (dual major, 1988–1993).  
Thesis title: *On Electron-Beam-Controlled Continuously Tunable High-Pressure CO<sub>2</sub> Laser System*  
Research advisor: Prof. Jun Zhao

### Awards and Honors

- Alfred P. Sloan Research Fellowship in Computer Science, 2008-2010.
- Career Award, National Science Foundation, 2006-2011.
- Best Paper Award, SIAM International Conference on Data Mining, 2007.
- John Van Ryzin Award for best paper, International Biometric Society-ENAR Annual Meetings, 2006.
- Runner-up Best Student Paper Award, 18th Conference on Uncertainty in Artificial Intelligence, 2003.
- Regents Fellowship, UC Berkeley, 1999.
- *Anthony Lu* Best Paper Award, Rutgers University, 1999.

### Principal External Grants and Awards

1. NSF CCF-0523757 (PI): "Nonparametric Bayesian Models for Genetic Variations and Their Associations to Diseases and Population Demography", Aug 1, 2005 – Jul 31, 2008, \$ 300,000.00.
2. NSF DBI-0546594, Career Award (PI): "CAREER: Uncovering the Process and Mechanism of Regulatory Evolution – Novel Statistical Models and Computational Algorithms for Evolutionary Genomics", Mar 1, 2006 – Feb 28, 2011, \$1,312,321.00.

3. NIH 1090151 (co-PI, with R. Murphy and W. Cohen): "Probabilistic Modeling of Information from Images and Text in Online Journals", Jul 1, 2006 – Jun 30, 2009, \$791,891.00.
4. NSF DBI- 0640543 (PI, with co-PI Christos Faloutsos): "Indexing, Mining and Modeling Spatio-Temporal Patterns of Gene Expressions", Aug 15, 2007 – July 30, 2010, \$1,331,995.00.
5. NSF IIS-0713379 (PI): "Novel Statistical Models and Algorithms for Network Modeling, Mining, and Reverse Engineering", Sep 15, 2007 – Aug 30, 2010, \$429,000.00.
6. DARPA (PI): "Computer Science Futures II-Engaging Young Scholars in Computer Science", May 1, 2008 – May 1, 2009, \$180,000.00.
7. Alfred P. Sloan Foundation: "2008 Sloan Research Fellowship in Computer Science", Sep 16, 2008 – Sep 15, 2010, \$50,000.00.

## Research Interests

**Machine learning**, with emphasis on theory and algorithms for learning complex probabilistic models, learning with prior knowledge, and reasoning and decision-making in open, evolving and uncertain possible worlds. Of particular interest are:

- Variational inference and learning theory, and development of turn-key variational inference engines
- Nonparametric Bayesian analysis, infinite mixture models, algorithms and applications of Bayesian nonparametrics for data mining and object/topic/event tracking in open, evolving possible worlds
- Statistical modeling and analysis of network and relational data, especially reverse engineering and meta-analysis of temporally evolving social and biological networks
- Statistical machine learning models and algorithms for image/text/relational information retrieval
- Probabilistic and optimization-theoretic methods for semi-supervised learning and distance metric learning

**Computational biology**, with emphasis on developing probabilistic models and algorithms that address problems of practical biological and medical concerns. Problems of primary focus include:

- Modeling evolution of gene expression, cis-regulatory code and transcriptional regulatory network in metazoan organisms
- Modeling genome-microenvironment interactions in cancer development and embryogenesis via joint analysis of genomic, proteomic, cytogenetic and pathway signaling data
- Statistical inference on genetic fingerprints, pedigrees, and their associations to diseases and other complex traits; application to clinical diagnosis and forensic analysis
- Modeling substitution, recombination, selection and genome rearrangement for comparative genomic analysis
- Biological image and text mining

**Career Goal:** To unravel and model the mechanism and evolution of living systems based on mathematical principles; and to develop the theories and algorithms for general probabilistic inference and learning systems both as a tool for computational biology and as building blocks of generic intelligent systems for a wide range of engineering applications in informational retrieval, natural language processing and reasoning under uncertainty.

## Teaching

- **Instructor**, *Machine Learning* (15-781/10-701).  
CMU, Fall 2006, Spring 2008, Fall 2008.

This is a core-curriculum course for SCS graduate students, focusing on fundamental algorithms and theory for statistical machine learning, pattern recognition and information retrieval.

- **Co-Instructor**, *Computational Genomics* (10-810) (formally known as Computational Molecular Biology: a Machine Learning Approach).  
CMU, Spring 2005, 2006, 2007, 2009.

This course focuses on modern machine learning methodologies for computational problems in molecular biology and genetics. This is a core-curriculum course for CMU-Pitt computational biology Ph.D. program.

- **Instructor**, *Probabilistic Graphical Models* (10-708).  
CMU, Fall 2005, Fall 2007, Fall 2009.

This is an advanced machine learning course covering probabilistic graphical models for efficient inference, decision-making and learning in problems with a very large number of attributes, complex stochastic dependencies, and huge datasets.

- **Instructor**, *Advanced Topics in Graphical Models* (10-801).  
CMU, Spring 2007.

This course covers advanced topics in approximate inference, model selection, Bayesian nonparametrics, and their applications.

## Papers and Publications

### Journal Papers

#### Published

- [1] P. Ray, S. Shringarpure, M. Kolar and **E. P. Xing**, *CSMET: Comparative Genomic Motif Detection via Multi-Resolution Phylogenetic Shadowing*.  
PLOS Computational Biology, 2008 (to appear).
- [2] E. Airodi, D. Blei, S. Fienberg and **E. P. Xing**, *Mixed Membership Stochastic Blockmodels*.  
Journal of Machine Learning Research, 2008 (to appear). (Journal version of [52].)
- [3] J. Yang, R. Yan, Y. Liu, and **E. P. Xing**, *Harmonium Models for Video Classification*.  
Statistical Analysis and Data Mining, ol 1, issue 1, p23-37, 2008. (Journal version of [40].)
- [4] K-A Sohn and **E. P. Xing**, *Spectrum: Joint Bayesian Inference of Population Structure and Recombination Event*.  
Bioinformatics, 23:i479-i489, 2007. (Journal version of [34].)
- [5] **E. P. Xing**, M. Jordan and R. Sharan, *Bayesian Haplotype Inference via the Dirichlet Process*.  
Journal of Computational Biology, Volume 14, Number 3, Pp. 267-284, 2007. (Journal version of [58].)
- [6] **E. P. Xing** and K-A Sohn *Hidden Markov Dirichlet Process: Modeling Genetic Recombination in Open Ancestral Space*.  
Journal of Bayesian Analysis, Volunn 2, Number 2, 2007. (Journal version of [44].)
- [7] T. Lin, E.W. Myers and **E. P. Xing**, *Interpreting Anonymous DNA Samples From Mass Disasters — probabilistic forensic inference using genetic markers*.  
Bioinformatics, 22(14):e298-e306, 2006. (Journal version of [45].)
- [8] W. Wu, N. Dave, G.C. Tseng, T. Richards, **E. P. Xing**, and N. Kaminsky, *Comparison of normalization methods for CodeLink Bioarray data*.  
BMC Bioinformatics, vol. 6, no. 309, 2005.
- [9] W. Wu, **E. P. Xing**, C. Myers, I. Mian and M. Bissell, *Evaluation of normalization methods for cDNA*

*microarray data by k-NN classification.*

BMC Bioinformatics, vol. 6, no. 191, 2005.

- [10] **E. P. Xing** and R. Karp *MotifPrototyper: a profile Bayesian model for motif family.* Proc. Natl. Acad. Sci., vol. 101, no. 29, 10523-10528, 2004.
- [11] **E. P. Xing**, D. Wolf, I. Dubchak, S. Spengler, M. Zorn, I. Muchnik and C. Kulikowski, *Automatic discovery of sub-molecular sequence domains in multi-aligned sequences: a dynamic programming algorithm for multiple alignment segmentation.* J Theor Biol, 212(2):129-39, 2001.
- [12] **E. P. Xing** and R. Karp, *CLIFF: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts.* Bioinformatics, 17 Suppl 1:S306-15, 2001. (Journal version of [65].)
- [13] Y. Cai, G. Yang, Y. Nie, L. Wang, X. Zhao, Y. Song, D. Seril, J. Liao, **E. P. Xing** and C. Yang, *Molecular alterations of p73 in human esophageal squamous cell carcinomas: loss of heterozygosity occurs frequently; loss of imprinting and elevation of p73 expression may be related to defective p53.* Carcinogenesis, 21(4):683-9, 2000.
- [14] **E. P. Xing**, Y. Nie, Y. Song, G. Yang, Y. Cai, L. Wang and S. Yang, *Mechanisms of inactivation of p14<sup>ARF</sup>, p15<sup>INK4b</sup>, and p16<sup>INK4a</sup> genes in human esophageal squamous cell carcinoma.* Clin Cancer Res, 5(10):2704-13, 1999.
- [15] **E. P. Xing**, G. Yang, L. Wang, S. Shi and S. Yang, *Loss of heterozygosity of the Rb gene correlates with pRb protein expression and associates with p53 alteration in human esophageal cancer.* Clin Cancer Res, 5(5):1231-40, 1999.
- [16] T. Shi, G. Yang, L. Wang, Z. Xue, B. Feng, W. Ding, **E. P. Xing** and S. Yang, *Role of p53 gene mutations in human esophageal carcinogenesis: results from immunohistochemical and mutation analyses of carcinomas and nearby non-cancerous lesions.* Carcinogenesis, 20(4):591-7, 1999.
- [17] **E. P. Xing**, Y. Nie, L. Wang, G. Yang, and S. Yang, *Aberrant methylation of p16<sup>INK4a</sup> and deletion of p15<sup>INK4b</sup> are frequent events in human esophageal cancer in Linxian, China.* Carcinogenesis, 20(1):77-84, 1999.

#### Invited

- [18] **E. P. Xing**, W. Wu, M. Jordan and R. Karp, *LOGOS: A modular Bayesian model for de novo motif detection.* Journal of Bioinformatics and Computational Biology, 2(1), 127-154, 2004. Invited to a special issue devoted to CSB2003 (expanded from [60] and peer reviewed again).

#### Submitted

- [19] J. Zhu, **E. P. Xing**, and B. Zhang *Maximum Entropy Discrimination Markov Network.* Journal of Machine Learning Research (under review), 2008.
- [20] **E. P. Xing** and K-A Sohn *Bayesian Multi-Population Haplotype Inference via a Hierarchical Dirichlet Process Mixture.* Annals of Applied Statistics (under review), 2008.

#### Books and Book Chapters

- [21] W. Wu and **E. P. Xing**, A Survey of cDNA Microarray Normalization and a Comparison by k-NN Classification, in S. Phillip ed., *Methods in Microarray Normalization*, p81-120, CRC Press, 2008.
- [22] E. Airodi, D. Blei, S. Fienberg, A. Goldenberg, **E. P. Xing**, and A. Zheng, Eds. *Statistical Network*

*Analysis: Models, Issues & New Directions*, Lecture Notes in Computer Science, volume no. 4503. Springer-Verlag, 2007.

- [23] **E. P. Xing**, Feature Selection in Microarray Analysis, in D. Berrar, W. Dubitzky and M. Granzow eds., *A Practical Approach to Microarray Data Analysis*, Kluwer, 2002.

## Refereed Conference Papers

### Published

- [24] R. Nallapati, A. Ahmed, **E. P. Xing**, and W. Cohen, *Joint Latent Topic Models for text and citations*. Proceedings of The Fourteen ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2008. (KDD '08).
- [25] S. Kim and **E. P. Xing**, *Sparse Feature Learning in High-Dimensional Space via Block Regularized Regression*. Proceedings of the 24th International Conference on Conference on Uncertainty in Artificial Intelligence, 2008. (UAI '08).
- [26] J. Zhu, **E. P. Xing**, and B. Zhang, *Laplace Maximum Margin Markov Networks*. Proceedings of the 25th International Conference on Machine Learning, 2008. (ICML '08).
- [27] A. Martins, M. Figueiredo, P. Aguiar, N.A. Smith, and **E. P. Xing**, *Nonextensive Entropic Kernels*. Proceedings of the 25th International Conference on Machine Learning, 2008. (ICML '08).
- [28] S. Shringarpure and **E. P. Xing**, *mStruct: A New Admixture Model for Inference of Population Structure in Light of Both Genetic Admixing and Allele Mutations*. Proceedings of the 25th International Conference on Machine Learning, 2008. (ICML '08).
- [29] A. Ahmed and **E. P. Xing**, *Dynamic Non-Parametric Mixture Models and the Recurrent Chinese Restaurant Process*. Proceedings of The Eighth SIAM International Conference on Data Mining, 2008. (SDM '08).
- [30] Z. Guo, Z. Zhang, **E. P. Xing** and C. Faloutsos, *Semi-supervised Learning Based on Semiparametric Regularization*. Proceedings of The Eighth SIAM International Conference on Data Mining, 2008. (SDM '08).
- [31] T. Lin, P. Ray, G. K. Sandve, S. Uguroglu, and **E. P. Xing**, *BayCis: a Bayesian hierarchical HMM for cis-regulatory module decoding in metazoan genomes*. Proceedings of the Twelfth Annual International Conference on Research in Computational Molecular Biology, 2008. (RECOMB '08).
- [32] B. Zhao and **E. P. Xing** *HM-BiTAM: Bilingual Topic Exploration, Word Alignment, and Translation*. Advances in Neural Information Processing Systems 21 (eds. D. Koller and Y. Singer), MIT Press, Cambridge, MA, 2008. (NIPS '07).
- [33] L. Chang, N. Pollard, T. Michell and **E. P. Xing** *Feature selection for grasp recognition from optical markers*. Proceedings of the 2007 IEEE/RSJ Intl. Conference on Intelligent Robots and Systems, 2007. (IROS '07).
- [34] K-A Sohn and **E. P. Xing**, *Spectrum: Joint Bayesian Inference of Population Structure and Recombination Event*. Proceedings of the Fifteenth International Conference on Intelligence Systems for Molecular Biology, 2007. (ISMB '07)
- [35] F. Guo, S. Hanneke, W. Fu and **E. P. Xing**, *Recovering Temporally Rewiring Networks: A model-based approach*.

- Proceedings of the 24th International Conference on Machine Learning (ed. Zoubin Ghahramani), Omni Press, 321-329, 2007. (ICML '07)
- [36] Z. Guo, Z. Zhang, **E. P. Xing**, and C. Faloutsos, *Enhanced Max Margin Learning on Multimodal Data Mining in a Multimedia Database*. Proceedings of The Thirteenth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2007. (KDD '07)
- [37] L. Gu, **E. P. Xing**, and T. Kanade, *Learning GMRF Structures for Spatial Priors*. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2007. (CVPR '07)
- [38] Z. Guo, Z. Zhang, **E. P. Xing**, and C. Faloutsos, *A Max Margin Framework on Image Annotation and Multimodal Image Retrieval*. Proceedings of IEEE International Conference on Multimedia & Expo, 2007. (ICME '07)
- [39] A. Ahmed and **E. P. Xing**, *On tight approximate inference of logistic-normal admixture model*. Proceedings of the Eleventh International Conference on Artificial Intelligence and Statistics, 2007. (AISTATS '07)
- [40] J. Yang, Y. Liu, **E. P. Xing** and A. Hauptmann, *Harmonium-Based Models for Semantic Video Representation and Classification*. Proceedings of The Seventh SIAM International Conference on Data Mining, 2007. (SDM '07). **Recipient of the BEST PAPER Award.**
- [41] H. Kamisetty, **E. P. Xing** and C. J. Langmead, *Free Energy Estimates of All-atom Protein Structures Using Generalized Belief Propagation*. Proceedings of the Eleventh Annual International Conference on Research in Computational Molecular Biology, 2007. (RECOMB '07)
- [42] Y. Shi, F. Guo, W. Wu and **E. P. Xing**, *GIMscan: A New Statistical Method for Analyzing Whole-Genome Array CGH Data*. Proceedings of the Eleventh Annual International Conference on Research in Computational Molecular Biology, 2007. (RECOMB '07)
- [43] F. Guo, W. Fu, Y. Shi and **E. P. Xing** *Reverse engineering temporally rewiring gene networks*. Workshop on New Problems and Methods in Computational Biology, Conference on Neural Information Processing Systems. 2006.
- [44] K-A Sohn and **E. P. Xing** *Hidden Markov Dirichlet Process: Modeling Genetic Recombination in Open Ancestral Space*. Advances in Neural Information Processing Systems 20 (eds. Y. Weiss and B. Schölkopf and J. Platt), MIT Press, Cambridge, MA, 2007. (NIPS '06).
- [45] T. Lin, E.W. Myers and **E. P. Xing**, *Interpreting Anonymous DNA Samples From Mass Disasters — probabilistic forensic inference using genetic markers*. Proceedings of the Fourteenth International Conference on Intelligent Systems for Molecular Biology, 2006. (ISMB '06)
- [46] J-Y. Pang, A. Balan, **E. P. Xing**, A. Traina and C. Faloutsos, *Automatic Mining of Fruit Fly Embryo Images*. Proceedings of The Twelfth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 2006. (KDD '06)
- [47] B. Zhao and **E. P. Xing**, *BiTAM: Bilingual Topic AdMixture Models for Word Alignment*. Proceedings of the 44th Annual Meeting of the Association for Computational Linguistics, 969-976, 2006. (ACL '06)

- [48] **E. P. Xing**, K-A. Sohn, M. Jordan and Y-W Teh, *Bayesian Multi-Population Haplotype Inference via a Hierarchical Dirichlet Process Mixture*.  
Proceedings of the 23rd International Conference on Machine Learning (eds. W. Cohen and A. Moore), ACM Press, 1049-1057, 2006. (ICML '06)
- [49] E. Airodi, D. Blei, S. Fienberg and **E. P. Xing**, *Combining Stochastic Block Models and Mixed Membership for Statistical Network Analysis*.  
Proceedings of the Workshop on Statistical Network Analysis, the 23rd International Conference on Machine Learning, 2006. (SNA-ICML '06)
- [50] S. Hanneke and **E. P. Xing**, *Discrete Temporal Models of Social Networks*.  
Proceedings of the Workshop on Statistical Network Analysis, the 23rd International Conference on Machine Learning, 2006. (SNA-ICML '06)
- [51] E. Airodi, D. Blei, S. Fienberg and **E. P. Xing**, *Latent mixed-membership allocation models of relational and multivariate attribute data*.  
Bayesian Statistics 8, Proceedings of the Valencia & ISBA Eighth World Meeting on Bayesian Statistics, 2006.
- [52] E. Airodi, D. Blei, **E. P. Xing** and S. Fienberg, *Mixed membership stochastic block models for relational data, with applications to protein-protein interactions*.  
Proceedings of International Biometric Society-ENAR Annual Meetings, 2006. **Recipient of the John Van Ryzin Award.**
- [53] F. Li, Y. Yang and **E. P. Xing**, *From Lasso regression to Feature vector machine*.  
Advances in Neural Information Processing Systems 19 (eds. Y. Weiss and B. Schölkopf and J. Platt), MIT Press, Cambridge, MA, 779–786, 2006. (NIPS '05)
- [54] E. Airoldi, D. Blei, **E. P. Xing** and S. Fienberg, *A Latent Mixed Membership Model for Relational Data*.  
Proceedings of the 3rd international workshop on Link discovery, ACM Press, New York, NY, USA, 82–89, 2005. (LinkKDD '05)
- [55] Y. Liu, **E. P. Xing** and J. Carbonell, *Predicting Protein Folds with Structural Repeats Using a Chain Graph Model*.  
Proceedings of the 22nd international conference on Machine learning (eds. L. De Raedt and S. Wrobel, ACM Press, New York, NY, USA, 513–520, 2005. (ICML '05)
- [56] **E. P. Xing**, R. Yan and A. Hauptmann, *Mining Associated Text and Images with Dual-Wing Harmoniums*.  
Proceedings of the 21st Annual Conference on Uncertainty in Artificial Intelligence (eds. F. Bacchus and T. Jaakkola), AUAI Press, Arlington, Virginia, 633–642, 2005. (UAI '05)
- [57] B. Zhao, **E. P. Xing** and A. Waibel, *Bilingual Word Spectral Clustering for Statistical Machine Translation*.  
Proceedings of the Second ACL Workshop on Effective Tools and Methodologies for Teaching NLP and CL, 2005.
- [58] **E. P. Xing**, R. Sharan and M. Jordan, *Bayesian Haplotype Inference via the Dirichlet Process*.  
Proceedings of the 21st International Conference on Machine Learning (eds. R. Greiner and D. Schuurmans), ACM Press, 879-886, 2004. (ICML '04)  
An earlier version of this paper also appeared as a book chapter in Lecture Notes in Bioinformatics, Special issue for 2nd RECOMB Satellite Workshop on Computational Methods for SNPs and Haplotypes, 2004.
- [59] **E. P. Xing**, M. Jordan and S. Russell, *Graph partition strategies for generalized mean field inference..*  
Proceedings of the 20th Annual Conference on Uncertainty in Artificial Intelligence (eds. M. Chick-

- ering and J. Halpern), AUA Press, Arlington, Virginia, 602–611, 2004. (UAI '04)
- [60] **E. P. Xing**, W. Wu, M. Jordan and R. Karp, *LOGOS: A modular Bayesian model for de novo motif detection*.  
Proceedings of the 2nd IEEE Computer Society Bioinformatics Conference, IEEE Computer Society, Washington, DC, USA, 2:266–76, 2003. (CSB '03)
- [61] **E. P. Xing**, M. Jordan and S. Russell, *A generalized mean field algorithm for variational inference in exponential families*.  
Proceedings of the 19th Annual Conference on Uncertainty in Artificial Intelligence (eds. Meek and Kjælf), Morgan Kaufmann Publishers, San Francisco, CA, 583–591, 2003. (UAI '03). **Recipient of the Runner-up Best Student Paper Award.**
- [62] **E. P. Xing**, *An expressive modular probabilistic model for de novo motif detection*.  
Workshop on Learning Graphical Models for Computational Genomics, 18th International Joint Conference on Artificial Intelligence (IJCAI '03), 2003.
- [63] **E. P. Xing**, A. Ng, M. Jordan and S. Russell, *Distance Metric Learning, with application to Clustering with side-information*.  
Advances in Neural Information Processing Systems 15 (eds. S. Becker, S. Thrun and K. Obermayer), MIT Press, Cambridge, MA, 505–512, 2003. (NIPS 02)
- [64] **E. P. Xing**, M. Jordan, R. Karp and S. Russell, *A Hierarchical Bayesian Markovian Model for Motifs in Biopolymer Sequences*.  
Advances in Neural Information Processing Systems 15 (eds. S. Becker, S. Thrun and K. Obermayer), MIT Press, Cambridge, MA, 1489–1496, 2003. (NIPS 02)
- [65] **E. P. Xing** and R. Karp, *CLIFF: clustering of high-dimensional microarray data via iterative feature filtering using normalized cuts*.  
Proceedings of the Ninth International Conference on Intelligent Systems for Molecular Biology, 2001. (ISMB '01)
- [66] **E. P. Xing**, M. Jordan and R. Karp, *Feature selection for high-dimensional genomic microarray data*.  
Proceedings of the Eighteenth International Conference on Machine Learning (eds. C. E. Brodley and A. P. Danyluk), Morgan Kaufmann Publishers Inc., San Francisco, CA, USA, 601–608, 2001. (ICML '01)
- [67] **E. P. Xing**, C. Kulikowski, I. Muchnik, I. Dubchak, D. Wolf, S. Spengler and M. Zorn, *Analysis of ribosomal RNA sequences by combinatorial clustering*.  
Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology, AAAI Press, 287–296, 1999. (ISMB '99)

**Submitted**

### **Unrefereed Technical Reports**

- [68] **E. P. Xing** and M. Jordan, *On semidefinite relaxation for normalized k-cut and connections to spectral clustering*.  
Technical Report CSD-03-1265, Computer Science Division, UC Berkeley, 2003.
- [69] **E. P. Xing**, *Dynamic Nonparametric Bayesian Models and the Birth-Death Process.*  
Technical Report CMU-CALD-05-114, Carnegie Mellon University, 2005.
- [70] **E. P. Xing**, *On Topic Evolution*.  
Technical Report CMU-CALD-05-115, Carnegie Mellon University, 2005.

- [71] F. Guo and **E. P. Xing**, *Bayesian Exponential Family Harmoniums*.  
Technical Report CMU-ML-06-103, Carnegie Mellon University, 2006.
- [72] F. Li, Y-M. Yang and **E. P. Xing**, *Inferring regulatory networks using a hierarchical Bayesian graphical Gaussian model*.  
Technical Report CMU-ML-06-117, Carnegie Mellon University, 2006.

## Invited Talks

- [1] *Feature Selection for High-Dimensional Genomic Microarray Data in Concept Learning and Clustering Analysis*,  
Statistics in Functional Genomics, Joint Summer Research Conference of AMS/IMS/SIAM, Mount Holyoke, June 10-14, 2001.
- [2] *Feature Selection for High-Dimensional Genomic Microarray Data*,  
NIPS 2001 Workshop on Machine Learning Techniques for Bioinformatics, Whistler, Dec 8, 2001.
- [3] *Expressive Statistical Models for Motifs*,  
Intel Workshop on Machine Learning and Life Sciences, Berkeley, Nov 3-4, 2003.
- [4] *Application of nonparametric Bayesian methods in genetic inference*  
NIPS 2003 Workshop on Nonparametric Bayesian Methods and Infinite Models, Whistler, Dec 13, 2003.
- [5] *Generalized Mean Field Inference in Graphical Models*,  
The 2004 joint WNAR/IMS meeting, Albuquerque, June 27-30, 2004.
- [6] *Generalized mean field inference in graphical models, and applications to computational biology*,  
AI Seminar, Carnegie Mellon University, Pittsburgh, PA, February 3, 2004.
- [7] *Mining Associated Text and Images with Dual-Wing Harmoniums*,  
UC BERKELEY CIS (Center for Intelligent Systems) SEMINAR, Berkeley, CA, April 20, 2005.
- [8] *In silico motif detection under complex genomic and evolutionary context - new Bayesian models motivated from biological principles*,  
UC Irvine ICS Seminar CIS, Irvine, CA, April 22, 2005.
- [9] *Probabilistic Graphical Models and Algorithms for Genomic Analysis*,  
Department of Molecular and Computational Biology, University of Southern California, Los Angeles, CA, April 23, 2005.
- [10] *Variational methods for inference in graphical models*,  
Workshop on Random Graphs & Stochastic Computation, Statistical and Applied Mathematical Sciences Institute, Research Triangle Park, NC, June 13-14, 2005.
- [11] *Nonparametric Bayesian Models for Haplotype Inference*,  
Section on Bayesian Statistical Science, The Joint Statistical Meetings, Minneapolis, Aug 6-11, 2005.
- [12] *Combinatorial and Statistical Approaches to Analyzing Biological Networks*,  
Tutorial Program, 2005 IEEE Computational Systems Bioinformatics Conference, Stanford, Aug 8-11, 2005.
- [13] *Mining Associated Text and Images with Dual-Wing Harmoniums and A Latent Mixed Membership Model for Relational Data*,  
Computer Science Department, Stanford University, Stanford, CA, Aug 12, 2005.
- [14] *In silico detection of cis-regulatory elements under complex genomic and evolutionary context: a probabilistic graphical model approach*,  
DIMACS Workshop on Machine Learning Approaches for Understanding Gene Regulation, Rutgers

University, Aug 15-17, 2005.

- [15] *How many founders shall we assume for haplotype reconstruction?*  
Center for Information Theory and Its Applications, Inaugural Workshop, University of California, San Diego, February 6-10, 2006.
- [16] *How many founders shall we assume for haplotype reconstruction? — on coalescence, Dirichlet processes, and nonparametric Bayes,*  
Invited Lecture (hosted by Prof. Andrew Yao), Computer Science Department, Tsinghua University, Beijing, China, April 25, 2006.
- [17] *Machine Learning, and the Role of Machine Learning in Computational Biology*  
Keynote, Symposium of the 80th Anniversary Of the Physics Department, Tsinghua University, Beijing, China, April 29, 2006.
- [18] *How many founders shall we assume for haplotype reconstruction? — on coalescence, Dirichlet processes, and nonparametric Bayes,*  
Invited Lecture, Workshop on Learning with Nonparametric Bayesian Methods, International Conference on Machine Learning, Pittsburgh, Pennsylvania, 25-29 June, 2006.
- [19] *Reasoning in open possible worlds: on A New Class of Nonparametric Bayesian Models for Haplotype Phasing, LD Modeling and Demographic Inference in Open Ancestral Space,*  
Department of Computer Science, National University of Singapore, Singapore, 16 November, 2006.
- [20] *A New Class of Nonparametric Bayesian Models for Haplotype Phasing, LD Modeling and Demographic Inference in Open Ancestral Space,*  
College of Life Sciences, Fudan University, Shanghai, China, 12 November, 2006.
- [21] *Modeling and reasoning the temporal evolution of networks*  
The 2007 Information Theory and Applications Workshop, University of California, San Diego, Jan 29- Feb 2, 2007.
- [22] *A Hidden Markov Dirichlet Process Model for Joint Inference of Population Structure, Linkage Disequilibrium, and Recombination Hotspots,*  
Invited talk, International Biometric Society-ENAR Annual Meetings, Atlanta, Georgia, 2007.
- [23] *Statistical network analysis and inference: methods and applications,*  
UC BERKELEY CIS (Center for Intelligent Systems) SEMINAR, Berkeley, CA, April 19, 2007.
- [24] *Probabilistic Graphical Models and Algorithms for Integrative Bioinformatics,*  
Keynote, Workshop on Bioinformatics, Graybill Conference (on Statistics and Probability) VI, Colorado State University, Fort Collins, Colorado, June 12-13, 2007.
- [25] *Probabilistic Graphical Models and Algorithms for Integrative Bioinformatics,*  
Computational Biology Seminar, Department of Molecular and Computational Biology, University of Southern California, Los Angeles, California, October 4, 2007.
- [26] *Statistical network analysis and inference,*  
Invited talk, Workshop on Gene Co-Expression Network Analysis and its Applications in Systems Biology, 8th International Conference on Systems Biology, Long Beach, California, October 5, 2007.
- [27] *Probabilistic Graphical Models and Algorithms for Integrative Bioinformatics,*  
Computer Science Department Colloquia, Princeton University, Princeton, New Jersey, October 10, 2007.
- [28] *Statistical Network Analysis and Inference: Methods and Applications,*  
Invited talk, Workshop III: Social Data Mining and Knowledge Building, Institute for Pure and Applied Mathematics (IPAM), Los Angeles, California, November 5-9, 2007.
- [29] *Nonparametric Bayesian Methods for Genetic Inference,*

Computational Biology Seminar, Department of Integrative Biology, UC Berkeley, Berkeley, California, November 7, 2007.

- [30] *Nonparametric Bayesian Methods for Genetic Inference*,  
Biostatistics Seminar, Department of Statistics, Stanford University, Stanford, California, November 8, 2007.
- [31] *Probabilistic Graphical Models — theory, algorithm, and application*,  
Keynote, The Sixth International Conference on Machine Learning and Applications (ICMLA'07),  
Cincinnati, Ohio, USA, December 13-15, 2007 .
- [32] *Nonparametric Bayesian Methods for Genetic Inference*,  
Statistics Seminar, Department of Statistics, University of Chicago, Chicago, Illinois, USA, March 10, 2008.
- [33] *Computational analysis of eukaryotic transcriptional regulatory sequence and its evolution*,  
Natural History Seminars, Department of Ecology and Evolution, University of Chicago, Chicago, Illinois, USA, March 11, 2008.
- [34] *Statistical Network Analysis and Inference: Methods and Applications*,  
Yahoo!-Dais Seminar, Department of Computer Science, University of Illinois, Urbana-Champaign, Illinois, USA, March 12, 2008.
- [35] *Nonparametric Bayesian Methods for Genetic Inference*,  
Computational Biology Seminar, Department of Computer Science, University of Illinois, Urbana-Champaign, Illinois, USA, March 13, 2008.
- [36] *Nonparametric Bayesian Methods for Genetic Inference*,  
Bioinformatics Seminar Series, CSAIL, MIT, Boston, Massachusetts, USA, April 9, 2008.
- [37] *Title: TBA*,  
Invited Speaker, Workshop on Future Directions in High-Dimensional Data Analysis: New Methodologies: New Data Types and New Applications, Isaac Newton Institute for Mathematical Sciences, Cambridge, UK, April 23-27, 2008.

## Professional Service

- **Invited panelist** of
  - DARPA CS Futures II Meeting, 2007.
- **Organizer or Co-Organizer** for
  - Workshop on Statistical Models of Networks. Advances in Neural Information Processing Systems 21, NIPS-07 (2007)
  - Workshop on Learning in Structured Output Spaces. The 24th International Conference on Machine Learning, ICML-07 (2007)
  - Institute of Mathematical Statistics (IMS) Session on Dynamic Network Models. International Biometric Society-ENAR Annual Meetings, Atlanta, Georgia, 2007
  - Workshop on Learning in Structured Output Spaces. The 23rd International Conference on Machine Learning, ICML-06 (2006)
  - Workshop on Statistical Network Analysis: Models, Issues and New Directions. The 23rd International Conference on Machine Learning, ICML-06 (2006)
- **Chair, co-Chair, or Senior Program Committee** member for
  - Tutorial Chair: The 7th Asia Pacific Bioinformatics Conference, APBC09 (2009)
  - The Fourteenth ACM SIGKDD International Conference on Knowledge Discovery and Data

Mining, KDD-08 (2008)

- The Twelfth Annual International Conference on Research in Computational Molecular Biology, RECOMB-08 (2008).
- The Twenty-Fourth International Conference on Machine Learning, ICML-07 (2007)

• **Program Committee** member for

- The 7th Asia Pacific Bioinformatics Conference, APBC-09 (2009)
- The 24th International Conference on Conference on Uncertainty in Artificial Intelligence, UAI'08 (2008)
- The Twenty-Third National Conference on Artificial Intelligence, AAI-08; and AAI-08 Nectar track (2008)
- European Conference on Computer Vision, ECCV-08 (2008)
- IEEE Conference on Computer Vision and Pattern Recognition Program, CVPR-08 (2008)
- The NIPS workshop on Machine Learning in Computational Biology, NIPS-07 (2007)
- Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning, EMNLP-CoNLL-07 (2007)
- The 11th IEEE International Conference on Computer Vision, ICCV-07 (2007)
- IEEE Conference on Computer Vision and Pattern Recognition Program, CVPR-07 (2007)
- SIAM International Conference on Data Mining (2007)
- Workshop on Multimodal Information Retrieval. The Twentieth International Joint Conference of Artificial Intelligence, IJCAI-07 (2007)
- Workshop on Learning with Nonparametric Bayesian Methods. The Twenty-Third International Conference on Machine Learning, ICML-06 (2006)
- The Twenty-Third International Conference on Machine Learning, ICML-06 (2006)
- The Twenty-First National Conference on Artificial Intelligence, AAI-06 (2006)
- The Fourth Asia-Pacific Bioinformatics Conference, APBC06 (2006)
- The Sixteenth and Seventeenth International Conference on Genome Informatics (2005, 2006,2007)
- The Tenth and Eleventh International Conference on Artificial Intelligence and Statistics, AIS-TAT (2005, 2007)
- The First, Second, and Third Annual RECOMB Satellite Workshop on Regulatory Genomics (2004, 2005, 2006)

• **Reviewer** for

- *American Journal of Human Genetics*,
- *Proc. Natl. Acad. Sci.*,
- *PLOS Computational Biology*,
- *ACM Transactions on Knowledge Discovery from Data*,
- *Bioinformatics*,
- *BMC Bioinformatics*,
- *International Journal of Computer Vision*,
- *Journal of Computational Biology*,
- *Journal of Machine Learning Research*,

- *Journal of Artificial Intelligence Research*,
- *IEEE Transactions on Information Theory*,
- *Genome Research*,
- *Knowledge and Information Systems*,
- *Machine Learning*,
- *Nucleic Acid Research*,
- *Social Networks*,
- Annual Conference on Advances in Neural Information Processing Systems (*NIPS*),
- Annual Conference on Uncertainty in Artificial Intelligence (*UAI*),
- Annual Conference on International Conference on Machine Learning (*ICML*),
- Annual IEEE Conference on Computer Vision and Pattern Recognition (*CVPR*),
- Annual Conference on Research in Computational Molecular Biology (*RECOMB*),
- Annual Conference on Intelligent Systems for Molecular Biology (*ISMB*),
- Annual Pacific Symposium on Biocomputing (*PSB*),
- National Conference on Artificial Intelligence (*AAAI*).
- **Grant Panelist** for
  - Biological Databases & Informatics, National Science Foundation, 2005
  - Information & Knowledge Management panel, IIS, National Science Foundation, 2005
  - Plant Genome Research Program, National Science Foundation, 2006
  - NSF Career Panel, 2007
  - NSF IIS Panel, 2008
- **Grant Reviewer** for
  - Austrian Science Fund (FWF)
  - Canada Foundation for Innovation (CFI)
  - Israel Science Foundation
  - National Science Foundation
  - The Research Grants Council (RGC) of Hong Kong
  - The Wellcome Trust
- **Professional organizations:**
  - Institute of Mathematical Statistics (IMS),
  - Institute of Electrical and Electronics Engineers (IEEE),
  - International Society for Bayesian Analysis (ISBA),
  - American Association for Artificial Intelligence (AAAI),
  - American Association for Cancer Research (AACR),
  - International Society for Computational Biology (ISCB).

## University Services (A partial listing)

- Annual Machine Learning Summer School, co-organizer (2005, 2006), Machine Learning Department, CMU.
- Faculty Search Committee, member (2006, 2007, 2008), Machine Learning Department, CMU.
- Admissions Committee, member (2006), Machine Learning Department, CMU.
- Admissions Committee, member (2005), Language Technology Institute, CMU.
- Admissions Committee, member (2006), chair (2007, 2008), Joint CMU-Pitt Ph.D. Program in Computational Biology.
- Curriculum Committee, member (2006, 2007), Joint CMU-Pitt Ph.D. Program in Computational Biology.
- ACM Doctoral Dissertation Award Committee, member (2007), SCS, CMU.
- New Collaborations Competition, Reviewer (2007), Language Technology Institute, CMU.

## Advising

Current students and Postdocs:

**Graduate Student:** Advising 7 Ph.D. students: Amr Ahmed (LTI), Wenjie Fu (CSD), Steve Hanneke (ML), Mladen Kolar (LTI), Kriti Puniyani (LTI), Kyung-Ah Sohn (CSD), Suyash Shringarpure (ML); co-advising 5 graduate students: Fan Guo (CSD), Judie Howrylak (M.D./Ph.D.), Hetunandan Kamichetty (CSD), Andre Martins (LTI), Pradipta Ray (CompBio).

**Post Doctoral Fellow:** Dr. Seyoung Kim.

Students graduated:

Henry Lin (LTI, M.S. 2006), Bing Zhao (LTI, Ph.D. 2007).

Served or serving on the thesis committee of:

Edoardo Airoldi (CSD), Jason Ernest (ML), Weihao Lin (LTI), Yan Liu (LTI), Yong Lu (CSD), Pradeep Ravikumar (ML), Indrayana Rustandi (CSD), Chenhe Yuan (Pitt, CS).