References

[1] E. M. Airoldi, D. M. Blei, S. E. Fienberg, and E. P. Xing. Mixed-membership stochastic blockmodels. *Journal of Machine Learning Research*, 9:1981–2014, 2008.

[2] A. Banerjee, I. Dhillon, J. Ghosh, S. Merugu, and D. S. Modha. A generalized maximum entropy approach to bregman co-clustering and matrix approximation. In *Proceedings of the tenth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 509–514. ACM, 2004.

[3] A. Beutel, K. Murray, C. Faloutsos, and A. J. Smola. CoBaFi: Collaborative Bayesian Filtering. In *World Wide Web Conference*, pages 97–108, 2014.

[4] A. Gittens and M. W. Mahoney. Revisiting the nyström method for improved large-scale machine learning. In *ICML (3)*, pages 567–575, 2013.

[5] D. Görür, F. Jäkel, and C. E. Rasmussen. A choice model with infinitely many latent features. In *Proceedings of the 23rd international conference on Machine learning*, pages 361–368. ACM, 2006.

[6] T. Griffiths and Z. Ghahramani. The indian buffet process: An introduction and review. *Journal of Machine Learning Research*, 12:1185âĹŠ1224, 2011.

[7] N. Halko, P. Martinsson, and J. A. Tropp. Finding structure with randomness: Stochastic algorithms for constructing approximate matrix decompositions, 2009. oai:arXiv.org:0909.4061.

[8] J. A. Hartigan. Direct clustering of a data matrix. *Journal of the american statistical association*, 67(337):123–129, 1972.

[9] Y. Koren. Factorization meets the neighborhood: a multifaceted collaborative filtering model. In *Knowledge discovery and data mining KDD*, pages 426–434, 2008.

[10] Y. Koren, R. Bell, and C. Volinsky. Matrix factorization techniques for recommender systems. *IEEE Computer*, 42(8):30–37, 2009.

[11] D. Koutra, U. Kang, J. Vreeken, and C. Faloutsos. *VOG: Summarizing and Understanding Large Graphs*, chapter 11, pages 91–99. SIAM, 2014.

[12] A. Kyrola, G. Blelloch, and C. Guestrin. GraphChi: Large-scale graph computation on just a pc. In *OSDI*, Hollywood, CA, 2012.

[13] J. Lee, S. Kim, G. Lebanon, and Y. Singer. Local low-rank matrix approximation. In *Proceedings of The 30th International Conference on Machine Learning*, pages 82–90, 2013.

[14] M. Leeuwen, J. Vreeken, and A. Siebes. Identifying the components. *Data Mining and Knowledge Discovery*, 19(2):176–193, 2009.

[15] M. Li, G. L. Miller, and R. Peng. Iterative row sampling. In *FOCS 2013*, pages 127–136, 2013.

[16] L. W. Mackey, M. I. Jordan, and A. Talwalkar. Divide-and-conquer matrix factorization. In *Advances in Neural Information Processing Systems*, pages 1134–1142, 2011.

[17] K. Palla, D. Knowles, and Z. Ghahramani. An infinite latent attribute model for network data. In *International Conference on Machine Learning*, 2012.

[18] E. E. Papalexakis, N. D. Sidiropoulos, and R. Bro. From k-means to higher-way co-clustering: multilinear decomposition with sparse latent factors. *Signal Processing, IEEE Transactions on*, 61(2):493–506, 2013.

[19] I. Porteous, E. Bart, and M. Welling. Multi-HDP: A non parametric bayesian model for tensor factorization. In D. Fox and C. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence*, pages 1487–1490. AAAI Press, 2008.

[20] J. Rissanen. Modeling by shortest data description. *Automatica*, 14(5):465–471, 1978.

[21] H. Shan and A. Banerjee. Bayesian co-clustering. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*, pages 530–539. IEEE, 2008.

[22] M. Turk and A. Pentland. Face recognition using eigenfaces. In *Proceedings CVPR*, pages 586–591, Hawaii, June 1991.

[23] J. Vreeken, M. Leeuwen, and A. Siebes. Krimp: mining itemsets that compress. *Data Mining and Knowledge Discovery*, 23(1):169–214, 2011.

[24] P. Wang, K. B. Laskey, C. Domeniconi, and M. I. Jordan. Nonparametric bayesian co-clustering ensembles. In *SDM*, pages 331–342. SIAM, 2011.