Topic-Link LDA: Joint Models of Topic and Author Community

Liu, Y., Niculescu-Mizil, A., & Gryc, W. (2009, June). Topic-link LDA: joint models of topic and author community. In *proceedings of the 26th annual international conference on machine learning* (pp. 665-672). ACM.

Current solutions to both topic modeling and community discovery have one major drawback: they treat all links (or missing links) between documents the same, which usually is not true in practice

To the best of our best knowledge, there is very limited work to jointly model underlying topics, author community, and link formation in one unified model.

Assumption, i.e. a citation between two documents is not purely due to content similarity.

topic-Link LDA model, we aim to quantify the

effect of topic similarity and community similarity to

the formation of a link. Therefore the model has three

major components with

we make the simple assumption that all

documents share the same Dirichlet parameter so that

inference can be made feasible for large-scale datasets.

In addition, we argue that if we are only interested in

documents within one specific domain (e.g. politics),

the simplification might be reasonable

Block-LDA: Jointly modeling entity-annotated text and entity-entity links

Balasubramanyan, R., & Cohen, W. W. (2011, April). Block-LDA: Jointly modeling entity-annotated text and entity-entity links. In *Proceedings of the 2011 SIAM International Conference on Data Mining* (pp. 450-461). Society for Industrial and Applied Mathematics.

**Minimum Spanning Tree Based Clustering Algorithms**

The MST clustering algorithm is known to be capable of detecting clusters with irregular boundaries. Unlike traditional clustering algorithms, the MST clustering algorithm does not assume a spherical shaped clustering structure of the underlying data.

Once the MST is built for a given input, there are two different ways to produce a group of clusters. If the number of clusters k is given in advance, the simplest way to obtain k

clusters is to sort the edges of the minimum spanning tree in descending order of their weights, and remove the edges with the first k−1 heaviest weights. undesired clustering structures and an unnecessarily large number of clusters, commonly faced by the EMST

Grygorash, O., Zhou, Y., & Jorgensen, Z. (2006, November). Minimum spanning tree based clustering algorithms. In *Tools with Artificial Intelligence, 2006. ICTAI'06. 18th IEEE International Conference on* (pp. 73-81). IEEE.

[**Document clustering based on non-negative matrix factorization**](http://dl.acm.org/citation.cfm?id=860485)

Xu, W., Liu, X., & Gong, Y. Document clustering based on non-negative matrix factorization. In *Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 267-273). ACM.

\bibitem{ros}

Rosvall, Martin, and Carl T. Bergstrom. "Maps of random walks on complex networks reveal community structure." Proceedings of the National Academy of Sciences 105.4 (2008): 1118-1123.

\bibitem{pon}

Pons, Pascal, and Matthieu Latapy. "Computing communities in large networks using random walks." International Symposium on Computer and Information Sciences. Springer Berlin Heidelberg, 2005.

\bibitem{clau}

Clauset, Aaron, Mark EJ Newman, and Cristopher Moore. "Finding community structure in very large networks." Physical review E 70, no. 6 (2004): 066111.

\bibitem{newn}

M. E. J. Newman, Fast algorithm for detecting community structure in networks. Phys. Rev. E 69, 066133 (2004).

\bibitem{mnew}

Newman, Mark EJ, and Michelle Girvan. "Finding and evaluating community structure in networks." Physical review E 69.2 (2004): 026113.

\bibitem{ragh}

Raghavan, Usha Nandini, Réka Albert, and Soundar Kumara. "Near linear time algorithm to detect community structures in large-scale networks." Physical review E 76.3 (2007): 036106.

\bibitem{blon}

Blondel, and Vincent D., et al. "Fast unfolding of communities in large networks." Journal of statistical mechanics: theory and experiment 2008.10 (2008): P10008.

\bibitem{reic}

Reichardt, Jörg, and Stefan Bornholdt. "Statistical mechanics of community detection." Physical Review E 74.1 (2006): 016110.

\bibitem{mark}

Newman, Mark EJ. "Finding community structure in networks using the eigenvectors of matrices." Physical review E 74.3 (2006): 036104.

\bibitem{fah}

Fahad, A., Alshatri, N., Tari, Z., Alamri, A., Khalil, I., Zomaya, A.Y., Foufou, S. and Bouras, A., 2014. A survey of clustering algorithms for big data: Taxonomy and empirical analysis. IEEE transactions on emerging topics in computing, 2(3), pp.267-279.

\bibitem{jain}

Jain, A.K., 2010. Data clustering: 50 years beyond K-means. Pattern recognition letters, 31(8), pp.651-666.

\bibitem{bez}

Bezdek, J.C., Ehrlich, R. and Full, W., 1984. FCM: The fuzzy c-means clustering algorithm. Computers and Geosciences, 10(2-3), pp.191-203.

\bibitem{frah}

Fraley, C. and Raftery, A.E., 2002. Model-based clustering, discriminant analysis, and density estimation. Journal of the American statistical Association, 97(458), pp.611-631.

\bibitem{est}

Ester, M., Kriegel, H.P., Sander, J. and Xu, X., 1996, August. A density-based algorithm for discovering clusters in large spatial databases with noise. In KDD (Vol. 96, No. 34, pp. 226-231).

\bibitem{ank}

Ankerst, M., Breunig, M.M., Kriegel, H.P. and Sander, J., 1999, June. OPTICS: ordering points to identify the clustering structure. In ACM Sigmod record (Vol. 28, No. 2, pp. 49-60). ACM.