

Annotated Bibliography for Senior Thesis

Ziv Epstein
`ziv.epstein@pomona.edu`

September 30, 2016

1 The sources

We begin with the two foundation papers introducing the NNMF concept and standard algorithms. Lee and Seung [5] were the first to introduce this idea, and to propose that topic modelling could be thought of as a matrix factorization problem. Ho [3] expands on this notion by elaborating on optimization schemas and corresponding algorithms that we will be taking advantage of. We then consider the work of Griffiths and Tenenbaum [2], who extened the notion of topic models to a hierarchical domain. We aim to replicate this structure but using an NNMF implementation instead of Latent Dirichet Allocation (LDA).

Our method to do is construct a distance matrix from the topic representations generated from NNMF. With this graph, we will employ community detection algorithms [1] to build the hierarchy. In particular, there already exist methods for generating a hierarchical community structure within complex networks [4] that we will take advantage of to build our topic model.

References

- [1] Santo Fortunato. Community detection in graphs. *Physics reports*, 486(3):75–174, 2010.
- [2] DMBTL Griffiths and MIJJB Tenenbaum. Hierarchical topic models and the nested chinese restaurant process. *Advances in neural information processing systems*, 16:17, 2004.
- [3] Ngoc-Diep Ho. *Nonnegative matrix factorization algorithms and applications*. PhD thesis, ÉCOLE POLYTECHNIQUE, 2008.
- [4] Andrea Lancichinetti, Santo Fortunato, and János Kertész. Detecting the overlapping and hierarchical community structure in complex networks. *New Journal of Physics*, 11(3):033015, 2009.
- [5] Daniel D Lee and H Sebastian Seung. Learning the parts of objects by non-negative matrix factorization. *Nature*, 401(6755):788–791, 1999.