

# COMP5329 Deep Learning

## A neural network approach to the researcher recommender system

King Tao Ng  
king7184@uni.sydney.edu.au

Anjali Upadhyay  
aupa8388@uni.sydney.edu.au

March 29, 2019

### Abstract

The Hierarchical Word Movers Distance has been recently proposed to find potential matches for research collaborations using a sample of journals extracted from the Faculty of Engineering and Information Technologies. The Hierarchical Word Movers Distance first applies the Word Movers Distance to journals and subsequently to researchers who have published them. However, this approach does not take social relations into account. For instance, a researcher may prefer to collaborate with others from the same school. Using a calculated distance metric from the Hierarchical Word Movers Distance as a predictor variable along with other social relations extracted from researcher profile pages, we use a neural network to compute a similarity metric among pairs of researchers. Because the similarity metric is a latent variable, we use a co-authorship count as a proxy. Higher similarity scores are; the more co-authorship counts should be. This paper simply extends the capstone project we previously worked on when we used a generalised linear model (GLM) to predict the co-authorship count. The neural network approach relaxes the linearity assumption that imposes on the GLM. In this paper, we describe the neural network approach and compare its performance with the GLM. It is believed the neural network would excel if the linearity assumption is violated.

## 1 Introduction

A plenty of information that floods over the internet makes searching for researchers a difficult task. Therefore, many institutes implemented such collaborative networks on the researcher profile pages based on historical publications and grants [3]. However, such visualisation does not show much value to the researchers because they already know whom they have collaborated with in the past. An alternative approach is to suggest relevant researchers based on some keywords provided by users. For instance, Gollapalli et al.[1] computed similarity by using the Named-Entity Recognition (NER) on journals and researcher profile pages. The NER annotates names of entities from a document in which users search for. This method fails to look at semantics. We addressed this issue by introducing the Hierarchical Word Movers Distance [5, 4]. The Hierarchical Word Movers Distance first applies the Word Movers Distance [2] to journals and subsequently to researchers who have published them. On the contrary, Xu et al.[6] concluded researcher collaboration systems had been thus far broadly classified into two categories for the past few years – Semantic Analysis and Social Relations. Hence, they proposed a two-layer network approach that integrated the two together. While the aim is quite similar to Xu et al.[6], our approach is completely different. Previously, we proposed to integrate the two by means of the GLM [4]. The GLM relies on the linearity assumption to hold. It is often not the case.

## 2 Research Questions

This paper simply extends the work we previously worked on and attempts to answer the following question:

- While semantic analysis is a critical factor for the researcher recommender system, arguably, social relations are equally important. By integrating the two, can we improve researcher recommendations using the neural network?

## References

- [1] S. D. Gollapalli, P. Mitra, and C. L. Giles. Similar researcher search in academic environments. *Proceedings of the 12th ACM/IEEE-CS joint conference on Digital Libraries*, 2012. URL <https://dl.acm.org/citation.cfm?id=2232849>.
- [2] M. J. Kusner, Y. Sun, N. I. Kolkin, and K. Q. Weinberger. From word embeddings to document distances. *Proceedings of the 32<sup>nd</sup> International Conference on Machine Learning*, 37, 2015. URL <http://proceedings.mlr.press/v37/kusnerb15.pdf>.
- [3] M. E. J. Newman. Coauthorship networks and patterns of scientific collaboration. *Proceedings of the National Academy of Sciences of the United States of America*, 101(1): 5200–5205, 2004. URL <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC387296/?tool=pmcentrez>.
- [4] K. T. Ng. Academic capability mapping. 2018. URL [https://github.sydney.edu.au/king7184/Academic\\_Capability\\_Mapping/blob/master/\\_documents/thesis/thesis.pdf](https://github.sydney.edu.au/king7184/Academic_Capability_Mapping/blob/master/_documents/thesis/thesis.pdf).
- [5] C. Sun, K. T. Ng, P. Henville, and R. Marchant. Hierarchical word mover distance for collaboration recommender system. *The 16<sup>th</sup> Australasian Data Mining Conference*, 2018.
- [6] Y. Xu, X. Guo, J. Hao, J. Ma, R. Y. Lau, and W. Xu. Combining social network and semantic concept analysis for personalized academic researcher recommendation. *Decision Support Systems*, 54, 2012. URL [https://ac.els-cdn.com/S0167923612001996/1-s2.0-S0167923612001996-main.pdf?\\_tid=db93b81b-2679-4bef-a258-955706be71f5&acdnat=1540251311\\_015d33d3a2473a6c4a0cc9a9a2353146](https://ac.els-cdn.com/S0167923612001996/1-s2.0-S0167923612001996-main.pdf?_tid=db93b81b-2679-4bef-a258-955706be71f5&acdnat=1540251311_015d33d3a2473a6c4a0cc9a9a2353146).