Benchettara N, Kanawati R, Rouveirol C. 2010. A supervised machine learning link prediction approach for academic collaboration recommendation. ACM Conference on Recommender Systems, Recsys, Barcelona, Spain, September. DBLP, 2010:253-256.

Doslu M, Bingol H O. 2016. *Context sensitive article ranking with citation context analysis*. Scientometrics, 2016, 108(2):653-671.

Le Quoc and Mikolov Tomas. 2014. *Distributed Representations of Sentences and Documents.* In Proceedings of the 31st International Conference on Machine Learning, Beijing, China.

Lim K W, Buntine W. 2016. *Bibliographic analysis on research publications using authors, categorical labels and the citation network.* Machine Learning, 2016, 103(2):185-213.

Mikolov Tomas, Sutskever Ilya, Chen Kai, Corrado Greg, and Dean Jeffrey. 2013. *Distributed representations of phrases and their compositionality.* In Advances on Neural Information Processing Systems.

Rendle Steffen, Freudenthaler Christoph, Gantner Zeno, Schmidt-Thieme Lars. 2009. *BPR: Bayesian Personalized Ranking from Implicit Feedback.* Appears in Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence.

Staudt C, Schumm A, Meyerhenke H, et al. 2013. Static and Dynamic Aspects of Scientific Collaboration Networks. IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining. IEEE, 2013:522-526.

Tang Jian, Qu Meng, Wang Mingzhe, Zhang Ming, Yan Jun, Mei Qiaozhu. 2015. *LINE: Large-scale Information Network Embedding.* WWW 2015, Florence, Italy.

Wang X, Sukthankar G. 2013. Link prediction in multi-relational collaboration networks. Ieee/acm International Conference on Advances in Social Networks Analysis and Mining. ACM:1445-1447.

Yang Weiwei, Boyd-Graber Jordan, Resnik Philip. 2016. *A Discriminative Topic Model using Document Network Structure.* Meeting of the Association for Computational Linguistics. 2016:686-696.

Yu X, Gu Q, Zhou M, et al. 2012. Citation Prediction in Heterogeneous Bibliographic Networks. Proceedings of the 2012 SIAM International Conference on Data Mining.

Liben-Nowell D, Kleinberg J. 2007. The link-prediction problem for social networks. Journal of the Association for Information Science and Technology, 58(7):1019–1031.

In scholarly papers, network analysis has become a prominent tool. Bibliographic networks have attracted interest both as social networks and as knowledge networks: they display the social structure of academia, while their bibliographic aspect allows conclusions about the structure of scientific knowledge. Recently, link prediction has become a popular task since it can be used for recommending potential research collaborations, citations and authors, which is the first and essential step for nearly all scientific research disciplines.

Document Embedding is a general method to preprocessing documents in Machine Learning, Text Mining and Natural Language Processing. Mapping variable-length, sparse word sequences into fixed-length semantic and social space is necessary for many Language model and Neural Network. (e.g. CNN). Embedding documents reflects potential semantic and social relationship between documents. Therefore, Link prediction can be effectively solved by document embedding according to the relevance of the corresponding embeddings.

Most previous works embedded documents by either text or citation network. Text embedding is the most popular method to map word sequences into vectors and one of the most common embedding methods is Bag-of-Word. Despite its popularity, BOW casts off the order of words and thus cannot reflect the genuine semantics of word sequences. Moreover, sparsity of its embedding space brings huge noise into analysis. In recent years, a variety of neural network methods (Mikolov, 2013; Le and Mikolov, 2014) have been proposed to map words, paragraphs or documents to dense vectors which are considered to embody their latent semantics.

Leveraging Citation Network (Lim and Buntine, 2016; Yang et al., 2016) is also an effective way to reflect social relationship between documents. In web pages, papers and articles, links/citations between documents can directly express relationship between documents. Many general network embedding methods, including BPRMF (Rendle et al., 2009), LINE (Tang, 2015), are widely used for document embedding.

However, few Document Embedding methods try to combine text embedding and network structure of scholarly papers while both of them can reflect one aspect of relationship between literatures. What’s more, as directional text, citation context is seldom used for document embedding generation. In this study, we proposed to lever-age citation context to induce documents and weight network. After that, we compare different embedding methods to assess contribution of different network structure and text-based embedding. Also, we proposed to combine text features and network structure to generate document embedding. Comparison results for link prediction on ACL Anthology Net-work (AAN) datasets show that our proposed method can significantly improve the predication performance.

The contributions of this paper are summarized as follows:

1) We proposed new methods to leverage citation context.

2) We proposed to generate document embedding by leveraging both text embedding and network structure.

3) We compared various network generation and text-based embedding methods in our link prediction task of AAN dataset.