

Bachelor/Master Thesis

Evaluation Of Graph Embedding Methods

Missing *link prediction* is a data mining task used to grow a social network by suggesting friends. Learning vector representation of nodes in social networks showed very promising results not only for link prediction but also for *node classification* and *network reconstruction*. Vector representations of nodes are typically learned using graph embedding methods. DeepWalk [1] and node2vec [2] are prominent examples of random walk-based graph embedding methods, while HOPE [3] and M-NMF [4] are methods based on graph adjacency matrix factorization. Each graph embedding method captures different structural properties of the network at hand and is, therefore, more or less suitable for a given data mining tasks. For example, the degree distribution, level of network sparsity, and homophily are network characteristics that might affect the performance of graph embeddings methods.

The goal of the thesis is to understand which (type of) graph embedding methods are suitable for a given data mining task and in which type of networks.

More specifically, as part of this thesis, the performance of graph embedding methods has to be evaluated with respect to two data mining tasks (i) *navigation*—finding (short) paths between nodes in a network, and (ii) *node ranking*—global node importance and importance of nodes with respect to a query.

To this end, the development of a robust evaluation framework allowing for the fair comparison of the embedding methods with respect to the two data mining tasks will be the main part of the thesis. The framework should be written in Python and easily extensible to new graph embedding methods.

Ideally, the insights of the evaluation of graph embeddings would help to develop new graph embedding methods able to support navigation and node ranking networks.

Contact:

Dr. Dimitar Dimitrov
GESIS - Leibniz Institute for Social Sciences
Unter Sachsenhausen 6-8, 50667 Cologne
E-Mail: dimitar.dimitrov@gesis.org
Phone: +49 (221) 47694-512

Prof. Dr. Stefan Dietze
Heinrich Heine University Düsseldorf - Data & Knowledge Engineering
Universitätsstr. 1 Building: 25.12 Floor/Room: 01.39, 40225 Düsseldorf
E-Mail: stefan.dietze@hhu.de
Phone: +49 (211) 81-13785

Introductory Readings:

- [1] Perozzi, Bryan, Rami Al-Rfou, and Steven Skiena. "Deepwalk: Online learning of social representations." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2014.
- [2] Grover, Aditya, and Jure Leskovec. "node2vec: Scalable feature learning for networks." *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016.
- [3] Ou, Mingdong, et al. "Asymmetric transitivity preserving graph embedding." *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 2016.
- [4] Cavallari, Sandro, et al. "Learning community embedding with community detection and node embedding on graphs." *Proceedings of the 2017 ACM on Conference on Information and Knowledge Management*. ACM, 2017.