NetSciX2026: International School and Conference on Network Science
February 17^{th}-20^{th}, 2026 - Auckland, New Zealand

Comparison of link prediction methods on synthetic and real networks

Keywords: link,prediction,methods,comparison,synthetic,real,networks

Extended Abstract

The problem of link prediction consists in using existing graph information (nodes and links) to predict unseen or missing links. Link prediction is an important and widely studied research topic in different disciplines – including complex networks and machine learning [1–4].

Solving this problem in complex networks requires algorithms that are able to explore statistical regularities in the existing network data. We investigate the interplay between algorithm efficiency and network structures through the introduction of suitably-designed synthetic graphs [5]. We apply these ideas to a real problem of link prediction in scientific paper citation graphs applying the same algorithms to this problem.

As a result of this work we propose a family of random graphs that incorporates both micro-scale motifs and meso-scale communities, two ubiquitous structures in complex networks. One contribution is the derivation of theoretical upper bounds for link prediction performance in our synthetic graphs, allowing us to estimate the predictability of the task and obtain an improved assessment of the performance of any method 1. Our results on the performance of classical methods (e.g., Stochastic Block Models [6], Node2Vec [7], GraphSage [8]) show that the performance of all methods correlate with the theoretical predictability, that no single method is universally superior, and that each of the methods exploit different characteristics known to exist in large classes of networks. Our findings underline the need for careful consideration of graph structure when selecting a link prediction method and emphasize the value of comparing performance against synthetic benchmarks.

We provide open-source code [9] for generating these synthetic graphs, enabling further research on link prediction methods.

References

[1] Aaron Clauset, Cristopher Moore, and M. E. J. Newman. "Hierarchical structure and the prediction of missing links in networks". In: *Nature* 453.7191 (2008). Publisher: Nature Publishing Group, pp. 98–101.

[2] Tao Zhou, Linyuan Lü, and Yi-Cheng Zhang. "Predicting missing links via local information". In: *The European Physical Journal B* 71 (2009), pp. 623–630.

[3] Zonghan Wu et al. "A comprehensive survey on graph neural networks". In: *IEEE transactions on neural networks and learning systems* 32.1 (2020), pp. 4–24.

[4] Xin Li and Hsinchun Chen. "Recommendation as link prediction in bipartite graphs: A graph kernel-based machine learning approach". In: *Decision Support Systems* 54.2 (2013), pp. 880–890.

[5] Alexey Vlaskin and Eduardo G Altmann. "Synthetic graphs for link prediction benchmarking". In: *Journal of Physics: Complexity* 6.1 (2025), p. 015004. URL: https://dx.doi.org/10.1088/2632-072X/ada07f.

[6] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. "Stochastic blockmodels: First steps". In: *Social Networks* 5.2 (1983), pp. 109–137.

[7] Aditya Grover and Jure Leskovec. "node2vec: Scalable feature learning for networks". In: *Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining*. 2016, pp. 855–864.

[8] Will Hamilton, Zhitao Ying, and Jure Leskovec. "Inductive representation learning on large graphs". In: *Advances in neural information processing systems* 30 (2017).

[9] Alexey Vlaskin. *Code and generated graphs repository for the paper.* Permanent link: `http://doi.org/10.5281/zenodo.14263197`. URL: `https://github.com/avlaskin/synthetic-graphs-for-lp`.

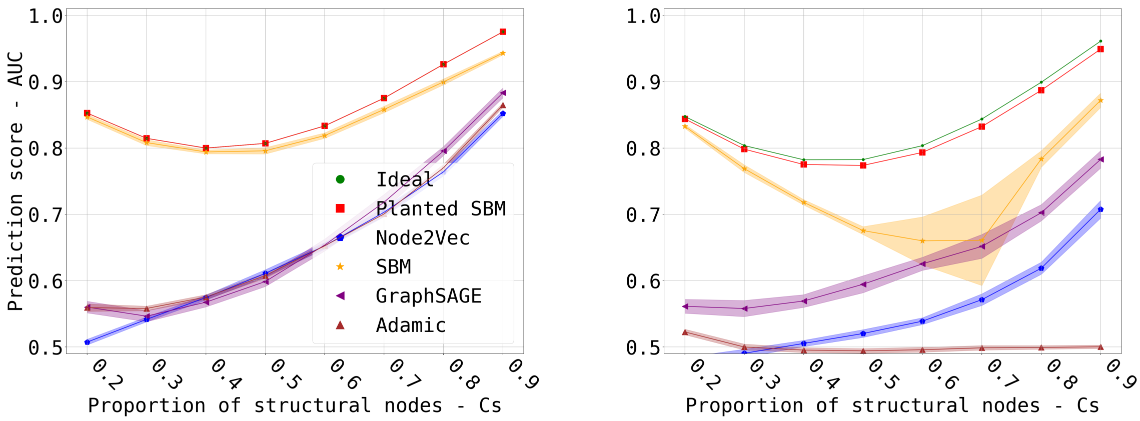

Figure 1: Performance of link prediction methods for graphs with increasing fraction C_S of structure nodes. Graphs with $N = 3200$ nodes were build using two types of structures: cliques (left panel) and 2d lattices (right) and bridges. $k = 8$, so clique has 8 nodes each and lattice has $8 \times 8 = 64$ nodes each. Probability of bridge $\alpha = 12/(N \cdot C_S)$.