Towards Automated Self-Supervised Learning for Truly Unsupervised Graph Anomaly Detection (Abstract Reprint)

Zhong Li[1], **Yuhang Wang**[1], **Matthijs van Leeuwen**[1]

[1]Leiden Institute of Advanced Computer Science (LIACS), Leiden University, Leiden, The Netherlands

Abstract Reprint. This is an abstract reprint of the journal article by Li, Wang, and van Leeuwen (2025).

Abstract

Self-supervised learning (SSL) is an emerging paradigm that exploits supervisory signals generated from the data itself, and many recent studies have leveraged SSL to conduct graph anomaly detection. However, we empirically found that three important factors can substantially impact detection performance across datasets: (1) the specific SSL strategy employed; (2) the tuning of the strategys hyperparameters; and (3) the allocation of combination weights when using multiple strategies. Most SSL-based graph anomaly detection methods circumvent these issues by arbitrarily or selectively (i.e., guided by label information) choosing SSL strategies, hyperparameter settings, and combination weights. While an arbitrary choice may lead to subpar performance, using label information in an unsupervised setting is label information leakage and leads to severe overestimation of a methods performance. Leakage has been criticized as 'one of the top ten data mining mistakes', yet many recent studies on SSL-based graph anomaly detection have been using label information to select hyperparameters. To mitigate this issue, we propose to use an internal evaluation strategy (with theoretical analysis) to select hyperparameters in SSL for unsupervised anomaly detection. We perform extensive experiments using 10 recent SSL-based graph anomaly detection algorithms on various benchmark datasets, demonstrating both the prior issues with hyperparameter selection and the effectiveness of our proposed strategy.

References

Li, Z.; Wang, Y.; and van Leeuwen, M. 2025. Towards automated self-supervised learning for truly unsupervised graph anomaly detection. *Data Mining and Knowledge Discovery*, 39: 44.