Implicit Anomaly Subgraph Detection (IASD) in Multi-Domain Attribute Networks

Ying Sun[1,2]
[1]Tianjin University
[2]Inner Mongolia University of Science and Technology
yingsun@tju.edu.cn

Abstract

Anomaly subgraph detection is a vital task in various real applications. However, with the advancement of AI technology, it faces new challenges: 1) Anomaly features are often deeply hidden within large datasets, and 2) Anomaly detection approaches are required to unveil the mechanisms behind anomaly generation. Our study focuses on detecting hidden anomaly subgraphs within big data and offering improved explanations for the root cause of anomalies by integrating multi-domain datasets.

1 Introduction

Anomaly subgraph refers to a subset of nodes and edges that exhibit unusual structural patterns or feature attributes within a graph. Anomaly subgraph detection plays a crucial role in various real-world applications, including social networks, public security, financial risk management, healthcare, and more. Existing research on anomaly detection primarily focuses on two fundamental aspects: the accuracy of anomaly identification and the capability to explain detected anomalies[Akoglu *et al.*, 2015]. With the surge in numerous Internet applications and the widespread adoption of big data, anomaly detection faces new challenges: (1)**Hidden anomalous features**. Attribute dimensions of nodes or edges in datasets are continuously growing, which further complicates the identification of anomalous features due to the curse of dimensionality. The scarcity of anomalies compared to normal data in large datasets undermines the performance and accuracy of anomaly detection models. Heterogeneity in data from diverse fields poses a challenge to the limited adaptability of anomaly detection models to varied data distributions. Data quality defects such as missing values, noise, and errors within large datasets can significantly impact the accuracy and stability of anomaly detection models. (2) **Anomaly explanation**. Anomaly detection can facilitate the explanation of the root cause of anomalies, especially when integrating multiple sources of data.

Due to variations in attribute values observed across multi-source datasets, the same anomalous event is likely to manifest differently in different network applications. In extreme scenarios, anomalous features may not be discernible within the target dataset; nonetheless, it is imperative to identify anomalies in such datasets. For example, infrastructure or geographical network data may not contain attributes related to criminal behavior. However, it remains essential to identify activity trajectories and hot spots of criminal activity using this dataset. Similarly, computer network traffic data may not contain information about organized financial fraud gangs. In certain scenarios, it's essential to analyze the behavior of individual scammers within the dataset.

Traditional anomaly subgraph detection approaches often utilize graph traversal techniques, e.g., [Wu *et al.*, 2019; Chen and Neill, 2014]. However, their computing performance is often constrained when tasked with exploring large datasets. Furthermore, none of them explicitly addresses the absence of anomalous features within datasets. Recently, most anomaly detection approaches based on deep learning models have focused on detecting anomalies in large datasets rather than specifically targeting anomalous subgraphs. Only a few of these approaches address anomalous subgraph detection explicitly[Ma *et al.*, 2023; Zhang and Zhao, 2022].

Our research focuses on detecting implicit anomaly subgraphs within complex networks. We aim to address the challenges posed by the lack of anomaly features in large datasets during detection. This problem is referred to as "Implicit Anomaly Subgraph Detection (IASD)." To address the IASD problem, our methodology relies on multiple domain datasets that incorporate auxiliary attributes to aid in the detection process within the target dataset. Therefore, we construct a complex network model from multiple large datasets, capturing intricate relationships among entities within real-world complex systems. We propose our novel anomaly-alignment IASD research framework on two kinds of network models: a multi-layer network and multiple networks to represent multiple domain data. One network or layer is derived from the target dataset, devoid of anomaly features, while the others originate from auxiliary attribute datasets that encompass the anomaly features. In our anomaly-alignment IASD framework, we integrate techniques for extracting anomalous features, anomaly detection, and network alignment. Technically, we investigate both non-deep learning and deep learning techniques separately. In the realm of non-deep learning techniques, we propose two algorithms: the Anomaly Subgraph Detection with Feature Transfer (ASDFT) algorithm, which operates on a multi-layer network, and the Anomaly

Alignment in attributed networks (AAAN) algorithm, which operates on the multiple network model. Since our framework fuses other auxiliary attributes and aligns them with the target network, both ASDFT and AAAN exhibit exceptional performance in IASD as well as outstanding capabilities for explaining anomalies. Furthermore, we are working on IASD within large datasets utilizing deep learning techniques. Our novel deep learning-based algorithm, Anomaly Subgraph Detection through High-order Sample Contrastive Learning (ASD-HC), has been proposed to implement IASD on a single network model.

We will provide detailed descriptions of our work in Section 2.

2 Contribution

2.1 Anomaly Subgraph Detection Within a Multi-Layer Attributed Network

In [Sun *et al.*, 2020], our approach integrates the anomaly subgraph detection methods and the network alignment techniques to design a feature transfer learning algorithm. We construct a multi-layer attributed network model tailored for multiple datasets. Our algorithm detects implicit anomaly subgraphs within the target network layer, leveraging transfer learning techniques to adapt features from the source network-layer. This allows us to detect anomalies even in cases where the target network-layer lacks anomalous features. Our experiments, conducted on several real datasets, demonstrate excellent performance in detecting anomalies relevant to real-world applications. Additionally, we conduct three case studies focusing on urban and area crime hotspot detection, special group identification, and epidemic breakout detection, all of which are pertinent to the implicit anomaly subgraph detection scenario.

2.2 Anomaly Subgraph Detection Crossing Multiple Attributed Networks

In [Sun *et al.*, 2022], we investigate the detection of the implicit connected anomaly subgraph by modeling multiple domain datasets. Our approach proposes a novel anomaly alignment diagram between two attributed networks: the auxiliary network with insufficient anomalous features and the target network with sufficient anomalous features. Extensive experiments on three real-world datasets show the effectiveness and efficiency of our algorithm. One of our case studies successfully identifies the crime hotspots in terms of city blocks from urban traffic networks, which are aligned with the criminal events reported on social networks. Another scenario involves detecting COVID-19 outbreaks using a geographical infrastructure dataset, with a traffic dataset from the same period acting as the auxiliary network. This case illustrates that population mobility plays a crucial role in the spread of COVID-19, providing the explanation of anomaly generation.

2.3 Anomaly Subgraph Detection by High-Order Sampling Contrastive Learning Model

In our third work, we focus on detecting the connected anomaly subgraph using a deep learning model to address the challenge posed by high-dimensional attributes in a large dataset and the representation of subgraphs in a deep model. Our approach comprises a novel high-order neighbor sampling module, a GCN-based contrastive learning model designed to extract anomalous features from a network with high-dimensional attributes, and a sub-optimal procedure that assesses the induced subgraph using Non-Parameter Graph Scan statistics. The evaluation experiment demonstrates the excellent performance of our approach when compared with state-of-the-art baselines across five real benchmark datasets. We further detect the maximum connected abnormal subgraph based on the obtained node's anomaly features.

3 Future Works

Building upon previous research, we will further investigate the deep learning algorithm for Implicit Anomaly Subgraph Detection (IASD). To address the challenges presented by multiple-domain big data and enhance our algorithm's ability to explain anomalies, we will expand our deep learning model to incorporate network structures that integrate multi-domain big data within our anomaly-alignment IASD research framework.

References

[Akoglu *et al.*, 2015] Leman Akoglu, Hanghang Tong, and Danai Koutra. Graph based anomaly detection and description: a survey. *Data Min. Knowl. Discov.*, 29(3):626–688, 2015.

[Chen and Neill, 2014] Feng Chen and Daniel B. Neill. Nonparametric scan statistics for event detection and forecasting in heterogeneous social media graphs. In *KDD*, pages 1166–1175. ACM, 2014.

[Ma *et al.*, 2023] Xiaoxiao Ma, Jia Wu, Shan Xue, Jian Yang, Chuan Zhou, Quan Z. Sheng, Hui Xiong, and Leman Akoglu. A comprehensive survey on graph anomaly detection with deep learning. *IEEE Trans. Knowl. Data Eng.*, 35(12):12012–12038, 2023.

[Sun *et al.*, 2020] Ying Sun, Wenjun Wang, Nannan Wu, Wei Yu, and Xue Chen. Anomaly subgraph detection with feature transfer. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management*, CIKM '20, page 1415–1424, New York, NY, USA, 2020. Association for Computing Machinery.

[Sun *et al.*, 2022] Ying Sun, Wenjun Wang, Nannan Wu, Chaochao Liu, Siddharth Bhatia, Yang Yu, and Wei Yu. AAAN: anomaly alignment in attributed networks. *Knowl. Based Syst.*, 249:108944, 2022.

[Wu *et al.*, 2019] Nannan Wu, Feng Chen, Jianxin Li, Jinpeng Huai, Baojian Zhou, Bo Li, and Naren Ramakrishnan. A nonparametric approach to uncovering connected anomalies by tree shaped priors. *IEEE Trans. Knowl. Data Eng.*, 31(10):1849–1862, 2019.

[Zhang and Zhao, 2022] Zheng Zhang and Liang Zhao. Unsupervised deep subgraph anomaly detection. In *2022 IEEE International Conference on Data Mining (ICDM)*, pages 753–762, 2022.