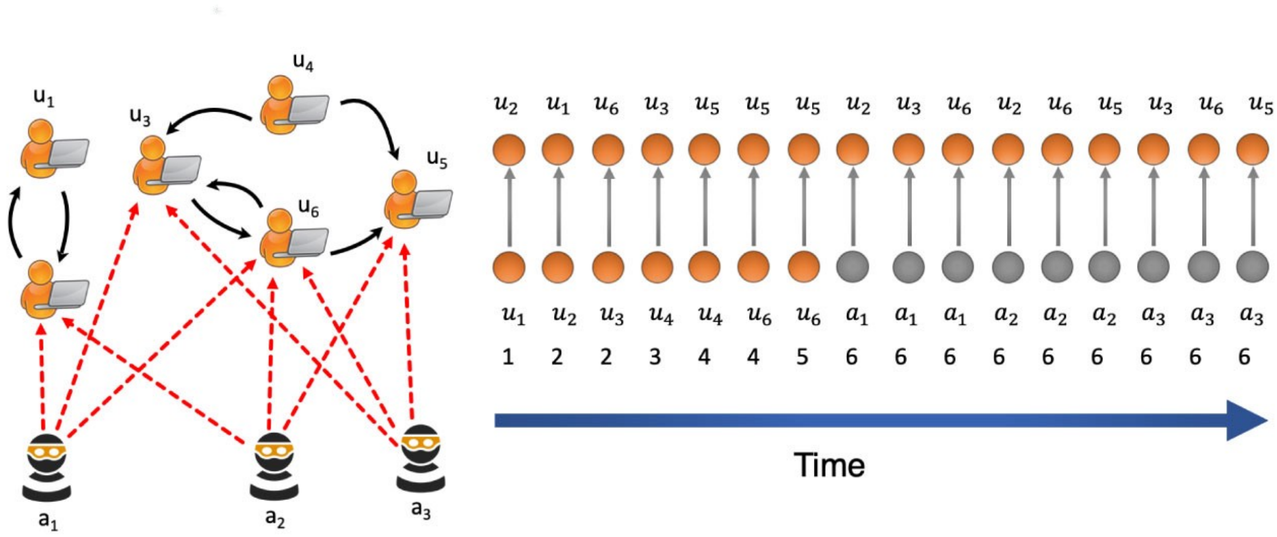
Rajvi Shah

Nov 27 · 6 min read · Listen

Save

Anomalous User Detection in Social Network using Graph Neural Networks (GNN)


u_1
u_3
u_4
u_5
u_6
a_1
a_2
a_3
u_2 u_1 u_6 u_3 u_5 u_5 u_5 u_2 u_3 u_6 u_2 u_6 u_5 u_3 u_6 u_5
u_1 u_2 u_3 u_4 u_4 u_6 u_6 a_1 a_1 a_1 a_2 a_2 a_2 a_3 a_3 a_3
1 2 2 3 4 4 5 6 6 6 6 6 6 6 6 6
Time

Reference: Link

Introduction

Graph neural networks apply the predictive power of deep learning to rich data structures that depict objects and their relationships as points connected by lines in a graph. In GNNs, data points are called nodes, which are linked by lines — called edges — with elements expressed mathematically so machine learning algorithms can make useful predictions at the level of nodes, edges or entire graphs. An expanding list of companies is applying GNNs to improve drug discovery, fraud detection and recommendation systems. These applications and many more rely on finding patterns communication. Social media is expanding & growing like something trending in the world. It allows the users to create their own profiles, communicate with the user and share their information, status, photos and videos. Social media application has become daily activities for the user and also helps in the business world, advertisement, and journalism and these are all indirectly depending upon the user's opinions. This popularity of social media became a black market that spoils the trust between users . These social media applications become illegal services for malicious users or fake users through daily conversations, comments, life tagging and sharing URLs as depicted in figure.

Reference: Link

Traditional Approaches

Traditional approaches to anomalous user detection mainly concentrate on the explicit account information (nickname, head portrait etc.) collected from users' activities, and only treat users as isolated individuals. Furthermore, as a result of evolving techniques about generative adversarial networks(GANs), anomalous users are able to imitate human-like behaviour and disguise themselves. Hence the above static detection approaches are no longer adapted to the current situations, and the discrimination of anomalous users has become more ineffective. Recently, due to the perfect performance of graph neural networks(GNNs) in capturing the hidden connectivity in a graph structure, many GNN-based anomaly detectors have been applied to various fraud or anomaly detection scenarios. In contrast to the traditional detection methods, GNN-based approaches consider the neighbourhood information to learn the node

Relevance-aware Anomalous User Detection

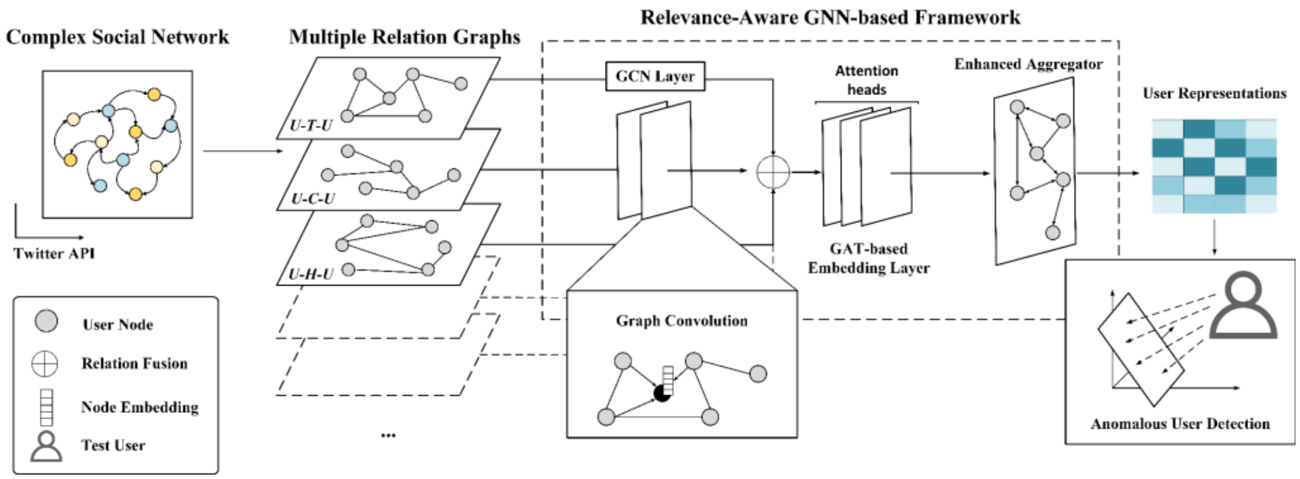
To further detect the well-disguised users in social networks, in this article, we have explained a GNN-based Relevanceaware Anomalous Users Detection (RAU-GNN) model to achieve fine-grained anomalous user detection results. We first extract the multiple relations from all users in the social networks. The relations between users could be roughly defined as an interaction, including retweets, comments or forwarding etc.

As for anomalous users, they prefer to forward similar blogs and take action at the same time. All these features can be constructed into a unified multiple-relation graph. Secondly, we leverage the relevance-aware GNN-based framework to learn the hidden representation in the constructed relation graph from users. Concretely, we adopt the GCN module to initially aggregate the structural information across different relationships, and embedded the processed fusional features to the centre nodes. Then we use the multi-head GAT module to learn the high-level embeddings and we feed the final node embeddings to the following GNN layer and aggregate all the users' information from their neighbours, in order to consolidate the previous embeddings.

The main contributions of RAU-GNN are summarized as follows:

- Extracted different relations from users and accordingly establish a multiple relation users graph network as the basis of RAU-GNN and explore the importance of different users and relations.
- It's named RAU-GNN based on a relevance-aware GNN framework, which consists of a GCN-based relation fusion layer, a GAT-based embedding layer and a final GNN aggregator respectively. The integration of these GNN layers can better learn the high-level representations and see through the well-designed disguise.
- Extensive experiments with real-world datasets are conducted to validate the effectiveness of RAU-GNN on anomalous user detection. The results demonstrate that the approach can achieve high accuracy and outperforms other classic

Concretely, our RAU-GNN contains three processes, including the construction of multiple relation graphs for users, an integrating GNN-based framework to learn the high-level representations, and a discrimination layer to detect the anomalous users. The proposed GNN-based framework integrates three layers, including a GCN-based relation fusion layer, a GAT-based embedding layer and a final GNN aggregator. The following sections will introduce more details of RAU-GNN at length.



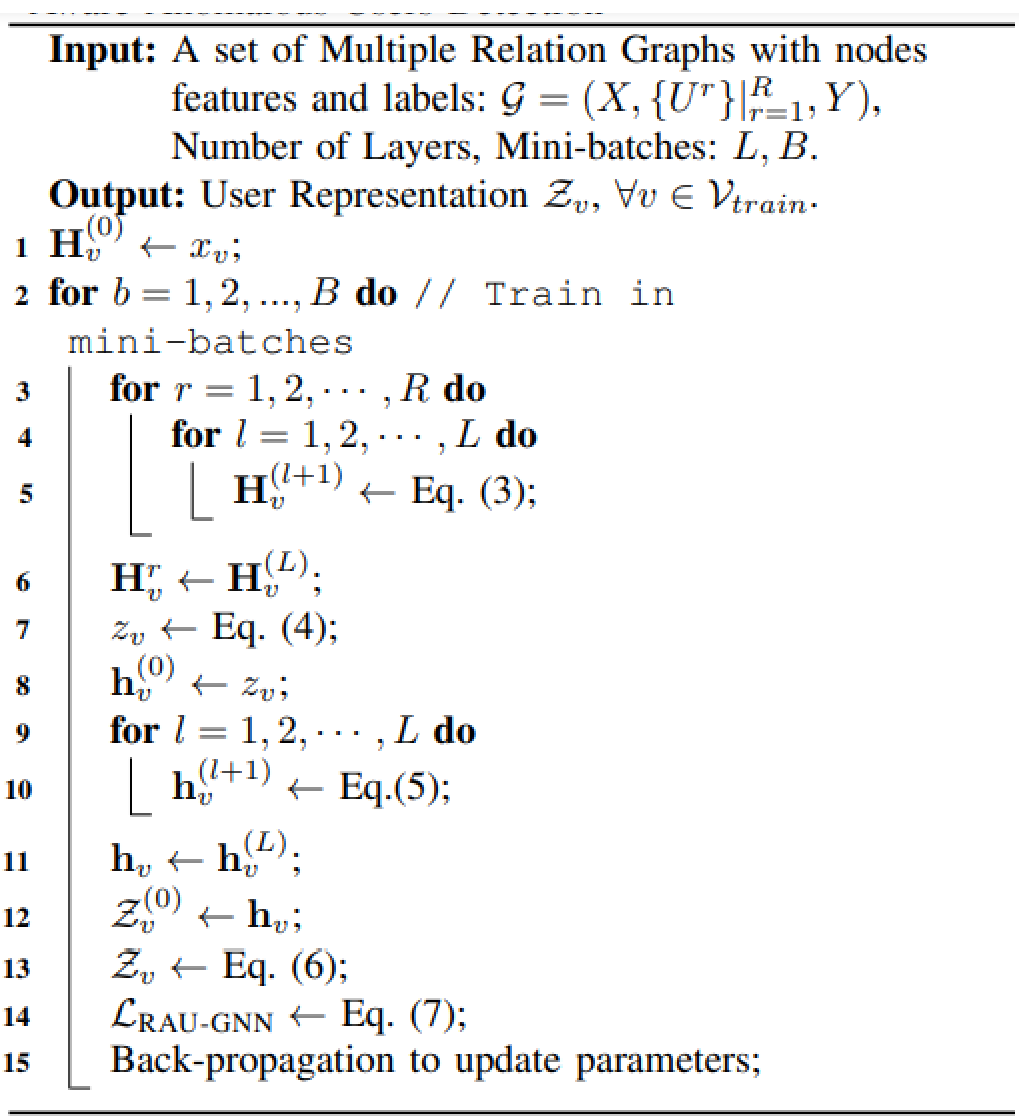
Reference: Link

To begin with, we extract the features of users from user objects by Twitter API. The rich extracted features serve as the initial node features in the social network. During the preprocessing, we aim to unify all the users and their relations and break them down into a set of relation-based homogeneous graphs. First, we extract the user-oriented elements, such as tweets, comments etc. and organize them in a unified manner to represent the different relations. For example, we extract a tweet that reveals User1 comments on tweets of other users. We denote User1 as v1 and add an edge between v1 and the extracted tweet comments. We repeat the same process for all the users in social networks and obtain a complex heterogeneous graph.

Algorithm

To verify the effectiveness of RAU-GNN, we compare our proposed model against heterogeneous attention over different types of nodes, and it performs cross-layer messages from different types of neighbours for higher-order aggregation. GraphSAGE is an inductive framework that leverages node attribute information to efficiently generate representations on previously unseen data. FDStar is a graph convolutional network approach for fraudster detection in review systems.

GAS is a GCN-based Anti-Spam model, that captures the local and global context of a comment to detect spammers. SemiGNN is a semisupervised attentive graph neural network that utilizes multi-view labelled and unlabeled data for fraud detection. Furthermore, to inspect the validation of a relevance-aware GNN-based framework, we decompose the RAU-GNN into a plain relation RAU-GNN(PR-RAU-GNN), which directly sums up the initial embedding across relations without a GCN module, and a plain aggregation RAU-GNN(PA-RAUGNN), which removes the last enhanced GNN aggregator.



```
Input: A set of Multiple Relation Graphs with nodes
       features and labels: $\mathcal{G} = (X, \{U^r\}|_{r=1}^{R}, Y)$,
       Number of Layers, Mini-batches: $L, B$.
Output: User Representation $\mathcal{Z}_v, \forall v \in \mathcal{V}_{train}$.
$\mathbf{H}_v^{(0)} \leftarrow x_v$;
for $b = 1, 2, ..., B$ do // Train in mini-batches
    for $r = 1, 2, \cdots, R$ do
        for $l = 1, 2, \cdots, L$ do
            $\mathbf{H}_v^{(l+1)} \leftarrow$ Eq. (3);
    $\mathbf{H}_v^r \leftarrow \mathbf{H}_v^{(L)}$;
    $z_v \leftarrow$ Eq. (4);
    $\mathbf{h}_v^{(0)} \leftarrow z_v$;
    for $l = 1, 2, \cdots, L$ do
        $\mathbf{h}_v^{(l+1)} \leftarrow$ Eq.(5);
    $\mathbf{h}_v \leftarrow \mathbf{h}_v^{(L)}$;
    $\mathcal{Z}_v^{(0)} \leftarrow \mathbf{h}_v$;
    $\mathcal{Z}_v \leftarrow$ Eq. (6);
    $\mathcal{L}_{\text{RAU-GNN}} \leftarrow$ Eq. (7);
    Back-propagation to update parameters;
```

Reference: Link

Conclusion

In this blog, a new GNN-based relevanceaware anomalous user detection model based on the paper, named RAU-GNN, to effectively discriminate the well-disguised anomalous users in social networks. Firstly, RAU-GNNextracts multiple relations between users in social network, and accordingly construct the multiple user relation graphs. Secondly, the authors have designed a relevance-aware GNN framework to learn the high-level users and discriminate the anomalous users through discriminating.

GAT-based embedding layer to represent the hidden embeddings of users. Lastly, the authors feed the learned representations to the following GNN aggregator in order to get the node embedding by aggregating the final users' embeddings and develop the robustness and generalization of RAU-GNN. The experimental results show that the approach can achieve better accuracy for anomalous user detection.

References

- https://blogs.nvidia.com/blog/2022/10/24/what-are-graph-neural-networks/
- https://ieeexplore.ieee.org/document/9534136
- https://towardsdatascience.com/introducing-midas-a-new-baseline-for-anomaly-detection-in-graphs-4311b8f7a737
- https://www.researchgate.net/publication/354685122_Detecting_malicious_users_in_the_social_networks_using_machine_learning_approach