Temporal Patterns Discovery of Evolving Graphs for Graph Neural Network (GNN)-based Anomaly Detection in Heterogeneous Networks

Jongmo Kim[1], Kunyoung Kim[1], Gi-yoon Jeon[2], and Mye Sohn[1*]
[1]Department of Industrial Engineering, Sungkyunkwan University, Suwon, South Korea
{dignityc, kimkun0, myesohn}@skku.edu
[2]R&D Institute, Agency for Defense Development, Seoul, South Korea
gyjeon@add.re.kr

Received: July 31, 2021; Accepted: December 29, 2021; Published: February 28, 2022

Abstract

This paper proposes a new method named evolving-graph generation framework to simultaneously solve the complexity and dynamic nature of the attribute networks that can occur in graph-based anomaly detection with Graph Neural Networks (GNN). The proposed framework consists of two components. The first component is a feature selection method that hybridizes filter-based and wrapper-based techniques to reduce the snapshots. The second component is an association method based on temporal patterns for the snapshots using the subgraph embedding technique and gaussian-base KL divergence. At the time, the association method finds intra-snapshots and inter-snapshots associations. As a result, we can obtain an evolving graph that is simplified and temporal patterns-enhanced from original networks. It is used an input graph for a GNN-based anomaly detection model. To show the superiority of the proposed framework, we conduct experiments and evaluations on 8 real-world datasets with anomaly labels with comparative state-of-the-art models of graph-based anomaly detection. We show that the proposed framework outperforms state-of-the-art methods in the accuracy and stability of training with the trend of decreasing train loss.

Keyword: Graph-based Anomaly Detection, Evolving Graphs, GNN, Attributed Networks, Heterogeneous Networks

1 Introduction

With the rapid development of communications and digital technologies, the connections among different types of networks such as social networks, financial networks, and/or health networks are increasing significantly [1]. So, it has caused an explosion in the volume of stream data over the networks [2]. the threats by fraudsters and attackers on the networks can spread across real-world applications [3]. For example, the fraudsters who have been stealing personal information on social media may try to access his/her bank account through the connections. Furthermore, they may attempt to manipulate the users' accounts to perform malicious activities such as DDoS, which attack by fabricating of the stream data [4]. To counter the threats, the graph-based anomaly detection is emerged [3]. In recent, the most widely used techniques for graph-based anomaly detection are data mining and machine learning [5]. In more detail, research on the graph-based anomaly detection is classified into two categories according

Journal of Internet Services and Information Security (JISIS), volume: 12, number: 1 (February), pp. 72-82
DOI:*10.22667/JISIS.2022.02.28.072*

*Corresponding author: Department of Industrial Engineering, Sungkyunkwan University, 300 Cheoncheon-dong, Jangan-Gu, Suwon, South Korea, Tel: +82-(0)31-290-7605, Fax: +82-(0)31-290-7610

to the network problem to be solved. The first one is the attribute-based approach that tries to resolve the complexity of the networks [6, 7, 8]. In general, the complexity of the networks can be explained by many types of features and structural topologies of the graphs that represent the characteristics of the networks. It means the more features and topologies there are, the greater the complexity, vice versa. The complexity of the networks may be further deepened by the prevalence of attribute networks and the inter-connections between them. So, it causes the high-dimensionality, non-linearity, and sparsity of network data. To reduce the problems, most attribute-based approaches find a low dimensional dense sub-space using Graph Neural Networks (GNN) model and graph embedding techniques. In other words, it solves the complexity of the networks by relying on the high generalization performance of the GNN. However, it isn't easy to find the best sub-space for anomaly detection with a single GNN model as complexity increases. Although some researchers have attempted to reduce the complexity by using graph embedding and clustering techniques, it is not feasible in real-world applications, as they assume the static networks to avoid non-stationary distributions [9]. The second one is the temporal-based approach that has emerged to address the dynamic nature of the networks [10, 11, 12, 13]. A dynamic network has the characteristic that its structure is not static but varies depending on time. It means that as the networks change dynamically, the patterns of the anomalies also become time-dependent. The methods for generating and updating temporal-based evolving graphs, which can flexibly represent the changes on the networks depending on time using the concept of subgraphs and their propagation, have been proposed to detect time-dependent anomalies. In other words, the temporal-based approach detects anomalies by generating the graphs for changes in the networks over time. To do so, attention-based GNN is generally adopted for learning the prediction model to avoid dependency on time intervals [11]. However, this approach has a fatal weakness to represent subgraphs of the attribute network at a specific timestamp (referred to as snapshots) using predefined features to avoid generating high-dimensional sparse feature sets. So, the risk of omitting essential information may occur for a sophisticated predictive model that can detect rare and complex anomalies. Even though many researchers attempt to overcome the shortcomings of these approaches, there is still no research trying to solve both problems like complexity and dynamic nature of the attribute networks simultaneously.

This paper proposes a novel method named Temporal Enhanced Evolving-graph Generation (T2EG) framework to simultaneously solve the complexity and dynamic characteristics of networks that can occur in graph-based anomaly detection. The evolving-graph generation framework consists of two components. The first component is a feature selection method that hybridizes filter-based and wrapper-based techniques to reduce the snapshots. At the time, the filter-based techniques are used to select the most appropriate subset of nodes needed to detect the anomalies from the attribute networks. The wrapper-based technique supports the selection of relevant attributes to the anomalous patterns. The second component is an association method based on temporal patterns for snapshots using the subgraph embedding technique and gaussian-base KL divergence. It is performed to find the association of the intra-snapshot and the inter-snapshot. At the time, the intra-snapshot is to find similar nodes in terms of information for the anomaly detection with embedding vectors and a prediction model. The inter-snapshot predicts the temporally changed data for a specific node in identified snapshots using gaussian-base KL divergence.

This paper is organized as follow. Section 2 review the related works of the graph-based anomaly detection. Section 3 offers the overall framework and their details. Section 4 evaluates the superiority of the proposed framework. Finally, Section 5

2 Related Works

Graph-based anomaly detection can be classified into two approaches depending on whether it focuses on the complexity or time series of the attribute network. In the attribute-based approach, most studies use deep learning models such as Graph Neural Networks (GNN) and Graph Convolutional Networks (GCN) to deal with topological structures and features of complex attribute networks. The issue of these studies is that the deep learning model recognizes the patterns of the complex and diverse subgraphs included in attribute networks and effectively solves the problems of sparsity and nonlinearity of the data [6, 7, 8, 14, 15]. The GATAE method focused on the attention-based networks for learning the models with flexibility to alleviate the over-smoothing problem [7]. On the other hand, the GraphAnoGAN have proposed a method to generate the anomalous snapshots from the global topology to provide enriched dataset produced by the GAN. It aims to resolve the lack of label information in the graph-based anomaly detection [6]. In addition, studies have been conducted to improve the performance of the GNN or GCN models, such as simplifying the topology of a graph by fuzzy clustering or utilizing external knowledge [14, 15].

In the temporal-based approach, many studies focused to represent time-dependent graphs from attribute networks using propagation or snapshots in a flexible way [10, 11, 12, 13]. The challenges of temporal-based approach are to discover the temporal patterns underlying in a large graph and to resolve the insufficient labelled data in time-series set. The AddGraph have been proposed to provide a semi-supervised learning framework based on extended GCN and attention-based GRU by combining hidden states between snapshots [11]. The StrGNN have been proposed to detect unusual subgraph focusing the edges with the temporal dependency [10].

3 The overall framework

The framework consists of two modules. The first module extracts reduced snapshots with filter-based and wrapper-based feature selection techniques to resolve the complexity of the attribute networks in graph-based anomaly detection. The second module discovers and represents the temporal patterns by finding two link predictions based on graph embedding and gaussian-based KL divergence to handle dynamic nature of graph in anomaly detection. The overall structure of the framework is depicted in Figure 1.

3.0.1 Snapshots Extraction with Feature Selection Methods Module

As mentioned, the ‘snapshot’ is usually defined as any sub-graph of a graph at a specific timestamp (hereafter, temporal context). However, not all snapshots contain the necessary information for detecting anomalies. In addition, non-informative (or redundant) snapshots worsen the performance of anomaly detection by increasing the dimensionality and sparseness of the training data. Therefore, it is critical that significant snapshots, which are composed of relevant attributes and nodes with respect to anomalies are identified from the attribute networks.

To do so, we propose a novel method for extracting the significant snapshots using filter-based and wrapper-based feature selection techniques, simultaneously. In the proposed method, the filter technique is used to select the nodes with the high relevance to the anomalies. By applying the filter technique, the computational complexity for large-scale of attribute networks can be significantly reduced. Next, for each node, the wrapper technique is applied to find the best minimal subset of the attributes that are required for distinguishing the node from other nodes in terms of anomaly detection. Because it is important to find the appropriate subset of the attributes directly related to the essential information of anomaly detection, we used the wrapper technique despite the heavy computational burden. For all

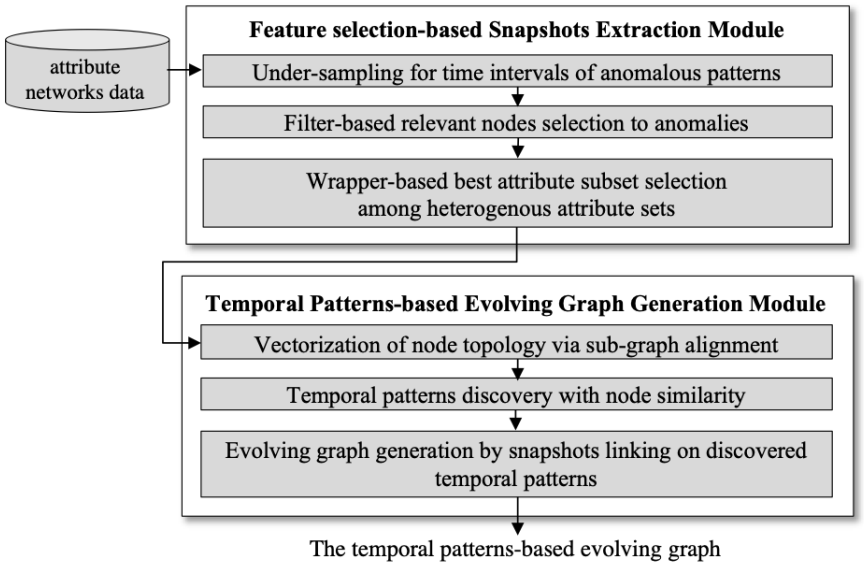

Figure 1: The framework of evolving graph generation based on temporal patterns discovery for attention-based GNN learning in graph-based anomaly detection

timestamps in a time interval on each anomalous pattern, the wrapper-based attributes selection and the filter-based nodes selection are performed, sequentially. Finally, we can get the snapshots at each timestamp. The illustrative example of the snapshots is depicted in Figure 2.

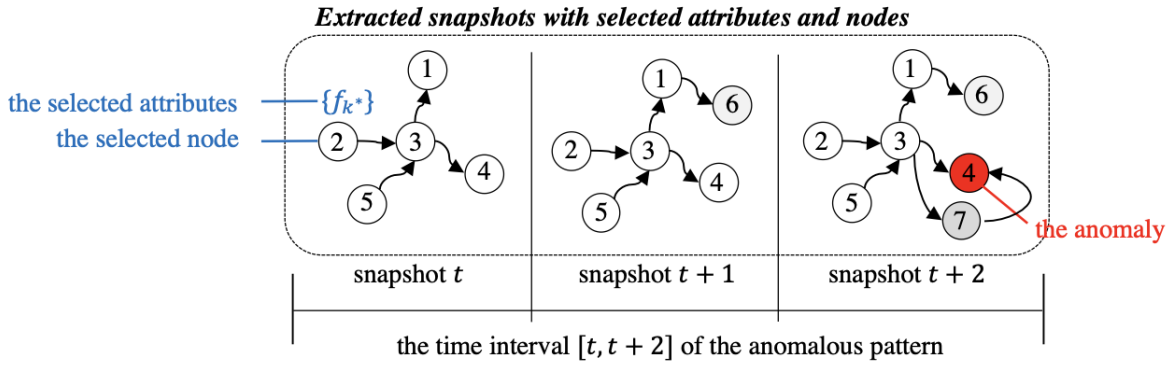

Figure 2: Extracted snapshots with selected attributes and nodes in the time interval of the anomalous pattern

As a data source, the proposed framework uses the attribute networks with label information for the anomalies. Additionally, the attribute networks are components of the associated attribute network, which are fully connected graphs of the heterogeneous attribute networks. The associated attribute network is defined as follows.

Definition 1 The associated attribute network (G) is a global graph combined of the heterogeneous attribute networks, and is represented by nodes, edges, and attributes just like the attribute networks. It

is simply represented by

$$G = \{V, E, X, Y\} \tag{1}$$

where V is a node set ($V = \{v_1, v_2, \ldots, v_i, \ldots, v_n\}$), E is an edge set ($E = \{e_1, e_2, \ldots, e_j, \ldots, e_m\}$), X is a attribute set of attribute set for each node ($X = \{x_1, x_2, \ldots, x_i, \ldots, x_n\}$, $x_i = \{t_i, f_{ik} \mid 1 \leq k\}$, t_i is a time vector to represent timestamps, and f_{ik} is a k-th attribute vector for v_i), and Y is a label indicating whether the node is abnormal or normal. ($Y = \{y_1, y_2, \ldots, y_i, \ldots, y_n\}$) with $y_i \in \{0, 1\}$ (0 and 1 denote normal and anomaly, respectively).
Generally, the anomalous patterns inevitably cause an imbalance problem because it appears with a very low frequency compared to the normal patterns. To resolve the imbalance, we perform the under-sampling for the time interval related to the normal patterns on the G. At this time, the under-sampling target is the time intervals, not typical classes. This means that a novel under-sampling method is required. So, we propose the following sampling method. This method first finds all the time intervals for the anomaly patterns and uses these intervals to randomly select the time intervals for the normal patterns. As apply the time intervals of both the normal and anomaly patterns, it can generate the balanced set of G (named bG).

The goal of filter-based relevant node selection is to find nodes that are related to anomalies. The nodes in the bG have various types such as numerical and categorical. So, as an evaluation function, we adopt a mutual information entropy that shows a versatile performance to mixed data of the nodes. I addition, we apply the minimum Redundancy and Maximum Relevance (mRMR) for filtering the redundant nodes. As a result, we finally obtain the selected node set V^- ($V^- = \{v_{i^-} \mid \forall i^-\}$, $|V^-| < n$). The information entropy-based mRMR evaluation function ($I - mRMR$) for the node selection is defined as follows.

$$I - mRMR = \max_{i \in \{i \mid \forall i\} - S\}} \left(I(x_i; y_i) - \frac{1}{|S|} \int_{s \in S} I(x_i; x_s) \right), \; x_{i,s}, y_i \in bG \tag{2}$$

where s is an index of selected node, S is a set of indices s ($S = \{s \mid \forall S\}, |S| < n$), and $I(a;b)$ is a mutual information function to measure mutual dependence between random variables.

In the final step of this module, the best subset of the attributes selects from the set of selected nodes (V^-). To do so, we apply the wrapper method that consists of two prediction models: Support Vector Classifier (SVC) and Random Forest Classifier (RFC). Since the two classifiers analyze and process the dimensions of data from different perspectives, constructing a wrapper using these two classifiers enables unbiased attributes selection.

3.1 Temporal and Attention-based GNN Learning Module

It is essential to capture the dynamic nature of the networks, in other words, temporal patterns in anomaly detection. To capture the temporal patterns, we need to discover the temporal patterns underlying among snapshots and represent them in graph structure. However, it is very difficult to discover and represent informative temporal patterns related to anomalies because the snapshots and the dynamic nature of the networks are not matched exactly. To reduce the difficulty, we propose two novel link predictions that can support discover and represent the temporal patterns.

The first one is link prediction of intra-snapshot. The targets of the intra-link prediction are two nodes in a snapshot. For the two nodes, it predicts whether the topology and information of two nodes are similar in perspective of the anomalies. In a snapshot, they are different nodes, but if they are predicted to be similar, a similar link is established between them. It can contribute to reducing the heterogeneity of the attribute networks and increasing the complexity of the graph. For predicting the similar links, we propose subgraph-based graph embedding technique.

The second one is link prediction of inter-snapshot. It predicts whether the information of the same node will change in time-varying snapshots. If there is no change, then no time-dependent information has occurred in terms of the graph. On the other hand, if there is a change in information, it is predicted that an anomaly is likely to occur, and these nodes on two snapshots are connected and named as an evolving link. By the evolving links, the temporal pattern of anomalies is revealed more explicitly. Furthermore, it can predict to detect anomalies with very high time-dependency, such as long-term anomalies. At this time, the evolving link is predicted according to the degree of change in the value of the attribute of a specific node over time. In addition, the degree of change is estimated according to the Gaussian distribution. Finally, we can get the snapshots at each timestamp. The generated evolving graph with two types predicted links is depicted in Figure 3.

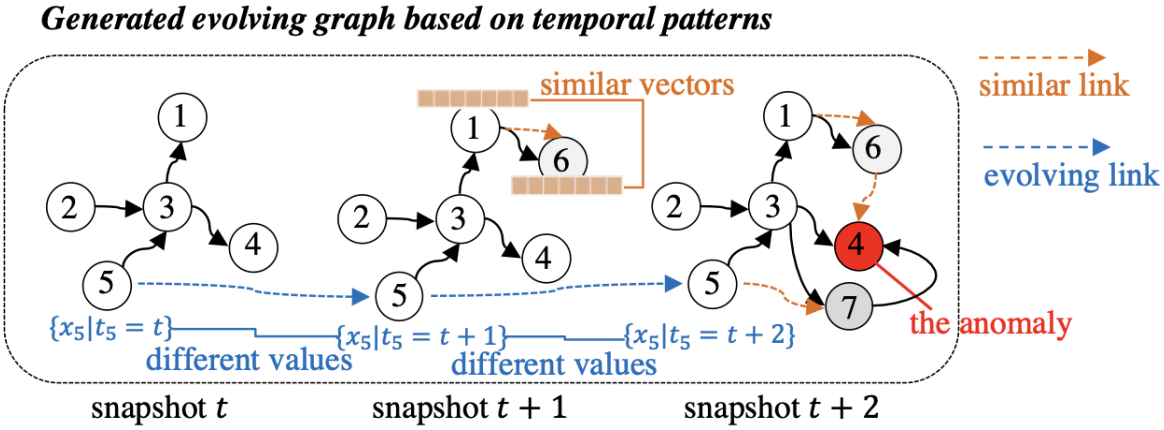

Figure 3: Generated evolving graph based on temporal patterns with similar links between different identified nodes and evolving links between same node in different snapshots with changed values of attributes

First, it projects the nodes into low-dimensional vectors using a subgraph embedding technique. At this time, the subgraph embedding is performed on all nodes of attribute networks, not on each snapshot to find the generalized sub-space. Next, the prediction model is trained with the different vectors of same node in different snapshots to capture the equivalence relationship. Finally, the trained model predicts the similar link between different two nodes with labels Y. For the subgraph embedding, the anomalous score function is defined as follows.

$$||\sum_i a^T x_i - \sum_j a^T x_j Y^T||_2^2 \tag{3}$$

To find the evolving links between two snapshots, we estimate the degree of changes for the same node during a certain time interval using Gaussian kernel. In contrast to the commonly used Gaussian distribution to evaluate the stability of a node state, we use the Gaussian distribution to statistically estimate the degree of change in the state of a specific node. First, for a specific node, distribution fitting is performed using a Gaussian kernel for x_i included in $t \pm \psi$ from the G. Then, according to the fitted distribution, the degree of change is predicted according to how far the μ value from the $\{x_i|t_i{=}t^*\}$ of the next snapshot at t^*. The Gaussian kernel for fitting is defined as follows.

$$N(x_i \mid \psi) = \frac{1}{(2\pi\sigma^2)^{1/2}} \exp\left\{-\frac{1}{2\sigma^2}(x_i - \mu)^2\right\}, \; x_i \in \{x_i| \; |t_i| \leq \; \psi\} \tag{4}$$

Table 1: The summary of multi-relational datasets from Yelp and Amazon review datasets

Datasets	#Nodes	#Anomalies (%)	#Edges	Avg. Interactions	Degree	Density
Yelp-all	45,954	6,677(6.88%)	3,846,979	83.7136	167.4748	0.003645
Yelp-rur			49,315	1.0731	4.1387	0.000173
Yelp-rtr			573,616	12.4823	25.2516	0.000556
Yelp-rsr			3,402,743	74.0467	148.2225	0.003228
Amazon-all	11,944	821(14.54%)	4,398,392	368.2511	736.5023	0.061668
Amazon-upu			175,608	14.7026	34.3521	0.003360
Amazon-usu			3,566,479	298.6	601.7343	0.050766
Amazon-uvu			1,036,737	86.7998	174.7850	0.014735

According to the fitted Gaussian distribution $\left(x_i \mid \mu, \sigma^2\right)$ for all nodes, the Gaussian distance using cross entropy (or KL divergence) of two nodes is calculated as follows.

$$KL(x_i) = \int p(x_i \mid t_i = t^*) \log(\frac{p(x_i \mid t_i = t^*)}{N(x_i|\psi)}) dx_i \tag{5}$$

As a result, based on predicted similar and evolving links, snapshots are composed to generate the evolving graph. Using this generated evolving graph and label information for anomalies, the GNN is trained to obtain a model capable of performing graph-based anomaly detection.

4 Performance evaluation

We conduct several experiments on real-world datasets compared to state-of-the-art graph-based anomaly detection methods.

4.1 Experimental Setup

Datasets. We prove a superiority of the proposed framework on real-world datasets (*Yelp* and *Amazon*) usually used in graph-based anomaly detection [16]. First, we use multi-relational datasets from Yelp review dataset including spam and legitimate reviews (annotated as anomalies) with hand-crafted 32 attributes. The multi-relational datasets from Yelp consist of 'Yelp-all', 'Yelp-rur', 'Yelp-rtr', and 'Yelp-rsr'. The 'Yelp-all' contains all interactions found in Yelp review dataset. The 'Yelp-rur' only has interactions between users and posted reviews by the same user. The 'Yelp-rtr' represents connections between two reviews posted for the same product in a month. The 'Yelp-rsr' connects interactions between reviews posted on same product with same star rating (5-star metric). Second, we use multi-relational datasets from Amazon review dataset including helpful and fraudulent reviews (annotated as anomalies). Similar to the Yelp dataset, it has three multi-relational datasets with 25 attributes. The 'Amazon-all' contains all interactions in the Amazon review dataset. The 'Amazon-upu' connects users posting reviews on same product at least one. The 'Amazon-usu' represents the connections between users having at least one review with same star rating in a week. The 'Aamzon-uvu' connects users posting similar reviews by top 5% similarity of review texts [17]. The summary of multi-relational datasets is described in Table 1.

Comparative Models. We selected the following comparative models to show how outperforms the proposed framework is compared to other models. Graph Convolutional Networks (GCN) is selected as basic models of GNN. The GCN has an advantage in finding structural patterns of dense subgraphs.

Table 2: The results of accuracy of the proposed T2EG framework and comparative models

Datasets	GCN	GEM-2	GEM-4	GraphSage	Player2vec	T2EG
Yelp-all	85.46	85.58	85.64	87.20	49.93	83.37
Yelp-rur	85.47	85.80	85.80	85.47	44.72	83.52
Yelp-rtr	85.47	81.96	81.96	86.54	47.13	83.45
Yelp-rsr	85.47	84.98	84.66	85.47	48.20	83.20
Amazon-all	93.14	93.09	93.09	97.21	44.76	97.44
Amazon-upu	93.87	92.26	93.09	97.13	41.34	97.38
Amazon-usu	93.14	93.31	93.09	97.08	45.71	93.38
Amazon-uvu	94.19	91.63	93.09	97.17	49.00	97.31

GEM is a GCN-based model that has been proposed to detect malicious accounts in heterogeneous graphs [18]. The implemented GEM models were named 'GEM-2' for 2 hops and 'GEM-4' for 4 hops according to the hops of the graph to be aggregated. GraphSage is a new model not based on GCN, it is a GNN model that concatenates without calculating the weight values of neighboring nodes [19]. Play2vec is a mixed model with GCN and Graph Attention Network (GAT) that solves limitations of multi-view heterogenous graph representation [20]. The proposed T2EG framework used the GAT model as a GNN base model, and 80% of the nodes were selected in the node selection stage. '*Relu'* was adopted for the activation function of all models, and cross entropy was used for the loss function except for the GEM. The GEM uses the novel loss function based on sigmoid function.

4.2 Evaluation

We used the accuracy and trend of train loss according to learning epochs for evaluation metrics. Since an epoch reaching convergence of the loss is different depending on the individual models, 300 epochs were configured to ensure that all models converge as much as possible. First, the test accuracy (acc) of the models according to the dataset is summarized in Table 2.

The performance of the models was slightly different depending on the interaction types of the datasets, but most showed similar performance regardless of the interaction types. In particular, most models showed good performance on the amazon datasets because the amazon dataset has dense graphs with many connections compared to the number of nodes. However, Player2vec showed very poor prediction performance for most datasets. This is because player2vec was developed for a high complex and heterogenous multi-view graph where massive features are mixed on different scales. GraphSage showed the best performance, but it took the longest learning time. Also, GEM showed moderate performance compared to learning rate, but the lower bound of the train loss was quite high. On the contrary the GCN had a deep lower bound of the train loss although the initial learning rate was slow. The trend of train loss on each epochs of the models is shown in Figure 4.

As the trend of decreasing train loss in the training phase of the models, the GEM fells into sub-optimal the fastest at very early stages. This is because the topology of a specific graph is fixed as an input graphs for the GCN-based model. In the case of the GCN, it fell to a lower loss, but the initial loss is set quite high. In other words, the neural networks of GCN is too simple to capture significant information of datasets. The GraphSage showed moderate predictive performance for overall datasets. However, it did not converge properly for 'yelp-homo' due to the complexity of dataset since the GraphSage needs a quite complex sub-graphs to be input for training. Thus, the GraphSage is week to handle too complex graphs efficiently.

The Play2vec showed a decrease trend in train loss, but there were no significant changes in acc. It means this model re-learn an already discovered pattern not to capture new patterns. In addition, the

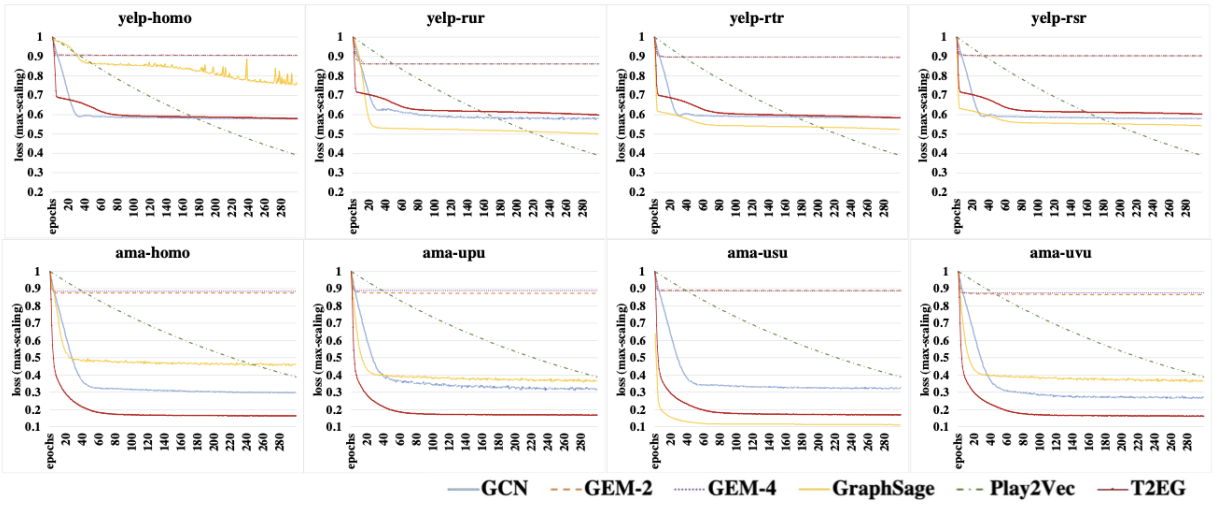

Figure 4: The train loss (max-scaling) for epochs in training of the proposed and comparative models on the experimental datasets

linear trend of decreasing train loss means that the input data is too simpler than necessary for the model training. The proposed T2EG framework showed rather low acc for the yelp datasets, but present stable train loss decreases. Besides, for the amazon datasets, it showed better training and prediction performance than other methods. The reason of low performance for 'yelp' datasets is that the feature selection part of the proposed framework causes significant information loss due to sparsity of the interaction types of 'yelp' datasets . However, it represents high predictive performance for 'amazon' datasets with dense interactions. Consequently, the proposed framework is more suitable for heterogeneous and complex attribute networks aimed at in this paper.

5 Conclusion and Further Research

We proposed a framework to generate temporal enhanced evolving-graph for supporting GNN-based anomaly detection. The proposed framework is designed to relive the complexity and dynamic nature of attribute networks in the graph-based anomaly detection. For the complexity, we proposed the reduction method of snapshots using filter-based and wrapper-based techniques. Moreover, for the dynamic nature, the two link prediction methods based on subgraph embedding and gaussian-based KL divergence to discover and represent temporal patterns in generating evolving graph. Finally, the generate evolving graph has a simple but dense snapshot with enhanced temporal patterns for anomaly detection. It can support the improvement of prediction performance for GNN-based anomaly detection.

However, we consider the attribute networks with enriched label information for the anomalies. In real-world applications, there are only few labels for the graph-based anomaly detection. Thus, a prediction method in semi-supervised or unsupervised manner should be proposed to handle real-world problems. To do this, we will devise a method of the unsupervised evolving graph generation using Generative Adversarial Networks (GAN), auto-encoder, or metric learning (few-shot learning). In addition, a comprehensive embedding method for heterogenous attribute networks over time will be studied for integration of different types of networks into a global graph flexibly.

Acknowledgements

This research is supported by C2 integrating and interfacing technologies laboratory of Agency for Defense Development (UE201114ED).

References

[1] A. Errahmane Kiouche, S. Lagraa, K. Amrouche, and H. Seba. A simple graph embedding for anomaly detection in a stream of heterogeneous labeled graphs. *Pattern Recognition*, 112:107746, 2021.

[2] S. Thudumu, P. Branch, J. Jin, and J. Singh. A comprehensive survey of anomaly detection techniques for high dimensional big data. *Journal of Big Data*, 7(1):42, July 2020.

[3] T. Pourhabibi, K.-L. Ong, B.H. Kam, and Y.L. Boo. Fraud detection: A systematic literature review of graph-based anomaly detection approaches. *Decision Support Systems*, 133:113303, 2020.

[4] M. De Donno, N. Dragoni, A. Giaretta, and A. Spognardi. Analysis of ddos-capable iot malwares. In *Proc. of the 2017 Federated Conference on Computer Science and Information Systems (FedCSIS'17), Prague, Czech*, pages 807–816. IEEE, September 2017.

[5] Y. Wu, H.-N. Dai, and H. Tang. Graph neural networks for anomaly detection in industrial internet of things. *IEEE Internet of Things Journal*, early access, doi: 10.1109/JIOT.2021.3094295, 2021.

[6] S. Bhatia, Y. Wang, B. Hooi, and T. Chakraborty. Graphanogan: Detecting anomalous snapshots from attributed graphs. In *Proc of the 2021 Joint European Conference on Machine Learning and Knowledge Discovery in Databases (ECML PKDD'21), Part II, Bilbao, Spain*, volume 12976 of *Lecture Notes in Computer Science*, pages 36–51, Cham, September 2021. Springer International Publishing.

[7] Z. You, X. Gan, L. Fu, and Z. Wang. Gatae: Graph attention-based anomaly detection on attributed networks. In *Proc. of the 2020 IEEE/CIC International Conference on Communications in China (ICCC'20), Chongqing, China*, pages 389–394. IEEE, August 2020.

[8] K. Ding, J. Li, R. Bhanushali, and H. Liu. Deep anomaly detection on attributed networks. In *Proc. of the 2019 SIAM International Conference on Data Mining (SDM'19), Calgary, Alberta, Canada*, pages 594–602. Society for Industrial and Applied Mathematics, May 2019.

[9] Y.-J. Park, K. Shin, and K.-M. Kim. Hop sampling: A simple regularized graph learning for non-stationary environments. arXiv:2006.14897v2, August 2020. `https://doi.org/10.48550/arXiv.2006.14897` [Online; Accessed on February 10, 2022].

[10] L. Cai, Z. Chen, C. Luo, J. Gui, J. Ni, D. Li, and H. Chen. Structural temporal graph neural networks for anomaly detection in dynamic graphs. In *Proc. of the 30th ACM International Conference on Information & Knowledge Management (CIKM'21), Virtual Event*, pages 3747–3756. ACM, October 2021.

[11] L. Zheng, Z. Li, J. Li, Z. Li, and J. Gao. Addgraph: Anomaly detection in dynamic graph using attention-based temporal gcn. In *Proc. of the 28th International Joint Conference on Artificial Intelligence (IJCAI'19), Macao, China*, pages 4419–4425. International Joint Conferences on Artificial Intelligence, August 2019.

[12] H. Zhao, Y. Wang, J. Duan, C. Huang, D. Cao, Y. Tong, B. Xu, J. Bai, J. Tong, and Q. Zhang. Multivariate time-series anomaly detection via graph attention network. In *Proc. of the 2020 IEEE International Conference on Data Mining (ICDM'20), Sorrento, Italy*, pages 841–850. IEEE, November 2020.

[13] D. Zhou, L. Zheng, J. Han, and J. He. A data-driven graph generative model for temporal interaction networks. In *Proc. of the 26th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining (KDD'20), Virtual Event*, pages 401–411. ACM, July 2020.

[14] S. Velampalli and W. Eberle. Novel graph based anomaly detection using background knowledge. In *Proc. of the 30th International Florida Artificial Intelligence Research Society Conference, Marco Island, Florida, USA*, pages 538–543. The AAAI Press, May 2017.

[15] Ç. Ateş, S. Özdel, and E. Anarım. Graph–based anomaly detection using fuzzy clustering. In *Proc. of the 2019 International Conference on Intelligent and Fuzzy Systems (INFUS'19), Istanbul, Turkey*, volume 1029 of *Advances in Intelligent Systems and Computing*, pages 338–345. Springer, Cham, July 2019.

[16] S. Rayana and L. Akoglu. Collective opinion spam detection: Bridging review networks and metadata. In *Proc. of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (KDD'15), Sydney, NSW, Australia*, pages 985–994. ACM, August 2015.

[17] Y. Dou, Z. Liu, L. Sun, Y. Deng, H. Peng, and P.S. Yu. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proc. of the 29th ACM International Conference on Information & Knowledge Management (CIKM'20), Virtual Event*, pages 315–324. ACM, October 2020.

[18] Z. Liu, C. Chen, X. Yang, J. Zhou, X. Li, and L. Song. Heterogeneous graph neural networks for malicious account detection. In *Proc. of the 27th ACM International Conference on Information and Knowledge Management (CIKM'18), Torino, Italy*, pages 2077–2085. ACM, October 2018.

[19] W.L. Hamilton, R. Ying, and J. Leskovec. Inductive representation learning on large graphs. In *Proc. of the 31st International Conference on Neural Information Processing Systems (NIPS'17), Long Beach, CA, USA*, pages 1025–1035. Curran Associates Inc., December 2017.

[20] Y. Zhang, Y. Fan, Y. Ye, L. Zhao, and C. Shi. Key player identification in underground forums over attributed heterogeneous information network embedding framework. In *Proc. of the 28th ACM International Conference on Information and Knowledge Management (CIKM'19), Beijing, China*, pages 549–558. ACM, November 2019.

Author Biography

Jongmo Kim is a student in Ph. D. course in the Department of Industrial Engineering at Sungkyunkwan University. He received his bachelor's degree from a Sungkyunkwan University. His main interests are ontology, semantic web, Linked Open Data, web service composition, and Web-of-Things.

Kunyoung Kim is a student in a Ph. D. course in the Department of Industrial Engineering at the Sungkyunkwan University. He received his bachelor's degree from the Sungkyunkwan University. His main interests are topic modeling, recommendation systems, big data analytics on social media semantic web, and web documents mining.

Gi-yoon Jeon is a research fellow in the R&D institute of Agency for Defense Development. He received his bechelor's and doctor's degree from the Dongguk University. His main interests are C4I systems, Human-Robot Interaction, IoT/IoB, Artificial Intelligence, and Visualization.

Mye Sohn is a professor in the Department of Systems Management Engineering at Sungkyunkwan University. She received her MS and Ph. D. from the Korea Advanced Institute of Science and Technology (KAIST). Her main interest is the machine learning, ontology, Web-of-Things, Semantic Web, and so on.