An Anomaly Event Detection Method Based on GNN Algorithm for Multi-data Sources

Yipeng Ji[1], Jingyi Wang[1], Shaoning Li[2], Yangyang Li[3], Shenwen Lin[4], Xiong Li[5,6*]

[1]School of Computer Science and Engineering, Beihang University, Beijing, China;

[2]School of Cyber Science and Technology, Beihang University, Beijing, China;

[3]National Engineering Laboratory for Public Safety Risk Perception and Control by Big Data, Beijing, China;

[4]National Computer Network Emergency Response Technical Team of China (CNCERT), Beijing, China;

[5]Beijing Biwei Network Technology Co., Ltd., Beijing, China; [6]Beijing Xindatek Technology Co., Ltd. Beijing, China;

jiyipeng,wjy10373,19373215@buaa.edu.cn;lsw@cert.org.cn;liyangyang@cetc.com.cn;li.xiong@foxmail.com

ABSTRACT

Anomaly event detection is crucial for critical infrastructure security(transportation system, social-ecological sector, insurance service, government sector etc.) due to its ability to reveal and address the potential cyber-threats in advance by analysing the data(messages, microblogs, logs etc.) from digital systems and networks. However, the convenience and applicability of smart devices and the maturity of connected technology make the social anomaly events data multi-source and dynamic, which result in the inadaptability for multi-source data detection and thus affect the critical infrastructure security. To effectively address the proposed problems, in this paper, we design a novel anomaly detection method based on multi-source data. First, we leverage spectral clustering algorithm for feature extraction and fusion of multiple data sources. Second, by harnessing the power of deep graph neural network(Deep-GNN), we perform a fine-gained anomaly social event detection, revealing the threatening events and guarantee the critical infrastructure security. Experimental results demonstrate that our framework outperforms other baseline anomaly event detection methods and shows high tracking accuracy, strong robustness and stability.

CCS CONCEPTS

• **Security and privacy** → **Domain-specific security and privacy architectures**; • **Computing methodologies** → **Machine learning**.

KEYWORDS

Secure Critical Infrastructure, Graph Neural Networks, Anomaly Event Detection

*Jingyi Wang is the corresponding author.

1 INTRODUCTION

Secure critical infrastructure, such as transportation system, economic sector, insurance service, government sector etc., is essential for the functioning of a society and economy. As smart devices and intelligent terminal advanced, critical infrastructures and their security increasingly come to rely on embedded control systems and external environment, especially social networks, which combined with massive hidden information on potential security threats [13, 32]. Due to the nature of events in social networks that is able to indirectly reflect such latent problems, detecting social events has become a research hotspot in data mining as well as security field to solve the attacks in advance.

Social abnormal event is a phenomenon that exists in a large number of social networks[15, 18, 21], and these abnormal event contain a lot of hidden information related to reality; by in-depth analysis and timely detection, it can realize effective management of social environment. Effective and timely detection is considered as an attractive method to discover the anomalies hidden or hot topic in the complex and ever-changing of multi-source data world, but there are many challenges to implement this approach; the challenges mainly include the variety of data types, the complex data sources differentiation and dependence complexity in multi-source data [6, 12]. From theory, since anomaly detection has important theoretical and practical significance, they have an attractive research field; From the reality, anomaly detection research is closely related to our economic life and has great practical significance and economic value.

Some model and method have been used in related field, and gain a lot of useful results. The most frequently used the methods and models of social anomaly detection include the followed several types: utilizing similar distance calculation method, utilizing similar density aggregation, clustering analytical method, high-dimensional characteristic and graph network to realized anomaly detection in multi-data source. In [3, 12], the paper considers the idea of the anomaly detection based on similar distance analytical method[14] is an hypothesis; the hypothesis is that the distance between the abnormal samples and the normal samples is larger than the distance from the normal sample to the data set; the anomaly degree is defined by the distance between the sample and other

objects in the data set, and the method needs set a distance hyper-parameter and a ratio hyper-parameter to renew all the samples in the data set after a process of filter operation. The disadvantage of this method can only label binary samples in the data set and cannot useful for samples of the multi-data source directly, and the similar distance have major limitations[31].

Cluster-based method is a machine learning technique that involves grouping data points; the detailed description of the method is that given a set of data points, clustering algorithms can be used to divide each data point into a specific group[8]. In theory, data points in the same group should have similar properties, and data points in different groups should have highly different attributes. Density-Based Spatial Clustering of Applications with Noise (DB-SCAN) algorithm is an representative algorithm of anomaly detection methods based on clustering. However, the disadvantage of the method based on density is that the effectiveness of the model. The type of algorithm is heavily dependent on the effectively characterize the clustering structure of the specified data set, and the generality of the algorithm is ignored.

With the popularity of intelligent communication equipment and the development of mobile Internet technology, social networking is becoming a tool that people use frequently; it forms a very large digital world of social behavior, and there are complex dependencies and correlation between the data of the diverse sources. The abnormal events of the social network can reflect the real emotions and opinions of network users, and becomes one of the important goal of big data mining in the network information space. The primary task of event detection is how to accurately represent the characteristics of social network data, and can timely detect and analyze events in streaming online social data. Aforesaid method can not be effectively applied to detect social anomaly events in multi-source data, but graph structures can be accurately used to represent the interdependencies between objects and to build model. Graph-based anomaly detection methods have become the attention of more and more researchers, especially in multi-source data domain. Multi-source exception detection method has some advantages, and the methods used samples between different data sources to detect exceptions in a multi-source data set. In these methods, the anomaly object was defined as if a sample is assigned to different clusters in different data block. In literature [16], the author considered the relationship between the sample in the specified data block and the possible clustering structures in that data block, to obtain the final anomaly score for each sample.

Although existing methods of the social anomaly event have produced great effect in multi-data source, with the continuous expansion of the application field and the increasing complexity of the types of data sources, there remains some problem need to be solved further such as the incremental representation of data, feature fusion learning and emergency detection in multi-data source[19]. With the increasing of deep learning technology in social anomaly event detection, the model and algorithm can bring new idea to address the practical problem. Since the dynamics, diversity and sparsity of social multi-data source cause difficult representation, poor fusion learning of model and low efficiency of event detection in multi-data source, the paper proposed a novel social anomaly event detection, and the method uses spectral clustering to realized

feature extraction based on multi-data source, and uses deep graph neural network to accomplish the social anomaly event detection.

The rest of this paper is arranged as follows: Sections II mainly introduces the related work and algorithms, Sections III describes the model designed. The results of the experiment are show in Sections IV and the conclusion in Sections V.

2 RELATED WORKS AND ALGORITHMS

This section mainly reviews the related theories and introduces the useful algorithms in our paper.

2.1 The Spectral Clustering

Spectral clustering is a widely used clustering algorithm, and by comparing to the traditional k-means algorithm, spectral clustering has stronger adaptability to realize data distribution, and has a good clustering effect and low computation. Spectral clustering is a clustering method based on graph theory, by clustering for the eigenvectors of Laplace matrix of sample data to achieve the purpose of sample data clustering. The spectral clustering algorithm is simply described as

- N sample points is inputted $X = \{x_1, x_2, \ldots, x_n\}$ and the number of cluster clusters k; output clustering cluster $A_1, A_2, \ldots, A_k$.
- The similarity matrix W of $n * n$ is calculated as the following formula:

$$s_{ij} = s(x_i, x_j) = \sum_{i=1, j=1}^{n} exp \frac{-\|x_i - x_j\|^2}{2\sigma^2} \qquad (1)$$

where W is similarity matrix of s_{ij}.

- The degree matrix D is calculated as the following formula:

$$d_i = \sum_{j=1}^{n} w_{ij} \qquad (2)$$

where D is the diagonal matrix of $n * n$;

- The Laplace matrix is calculated $L = D - W$;
- To calculate the eigenvalues of L, and sort the eigenvalues from smallest to largest, and extract the first k eigenvalues, and then calculate the eigenvectors of the first k eigenvalues $u_1, u_2, \ldots, u_k$,;
- Take k of the above column vectors and form them into a matrix $U = \{u_1, u_2, \ldots, u_k\}$, $U \in R^{n*k}$;
- Assum $y_i \in R^k$ is the ith row vectors of U, where $i = 1, 2, \ldots, n$;
- Use k-means algorithm to place new sample points $Y = \{y1, y_2, \ldots, y_n$ to realize the cluster cluster $C_1, C_2, \ldots C_K$;
- Output cluster $A_1, A_2, \ldots, A_k$, where, $A_i = \{j|y_j \in C_i\}$.

2.2 Typical Algorithms

According to the partition criterion used in the spectral clustering algorithm, the algorithm can be divided into two-way spectral clustering algorithm and multi-way spectral clustering algorithm[30]. The former uses two-way division criterion and the latter uses K-way division criterion. Two-way spectral clustering algorithm includes PF algorithm, SM algorithm, SLH algorithm, KVV algorithm, and Mcut algorithm. PF algorithm proposes the eigenvectors corresponding to the maximum eigenvalue of the similarity matrix

W are used for clustering, and points out that for block diagonal similarity matrices, the points corresponding to non-zero value of eigenvectors belong to the same class, and the points corresponding to zero value belong to another class[28]. SM algorithm points out that the difference between NCUT and MNCUT lies only in the spectral mapping used, and the eigenvector corresponding to the second smallest eigenvalue, namely Fiedler vector, contains the partition information of the graph, and according to the heuristic rule, the partition point i in this vector is found to minimize the value of NCUT (A,B) obtained at this point, Finally, the value in the vector is compared with the minimum value of NCUT criterion function[22]. The points are greater than this value or equal to this value, and the points belong to one class, and points are less than this value belong to another class. The SLH algorithm reposition algorithm calculates for the first k feature vectors of the similarity matrix W, and the parameter k needs to be specified in advance. KVV algorithm, according to the heuristic rules, the division point i in the Fiedler vector is found to minimize the RCUT (A,B) value at this point, which is similar to the SM algorithm[4]. The only difference is that the SM algorithm looks for the partition point that minimizes the value of NCUT (A,B).

Multi-path spectral clustering algorithm includes NJW algorithm and MS algorithm [11]. NJW algorithm, the eigenvectors corresponding to the first k maximum eigenvalues of the Laplace matrix are selected to construct a new vector space R, and the corresponding relationship with the original data is established in this new space, and then the clustering is carried out[3]. In MS algorithm, the first k non-zero eigenvalues of the random walk matrix P are used to construct the matrix, and then the rows in the matrix are treated as points in R space for clustering. The steps are similar to the NJW algorithm. The MS algorithm has achieved good results in the actual image segmentation, but the clustering effect will be poor if there is a big difference between the diagonal element values in the degree matrix D[9].

2.3 Graph Neural Network

Recent years, graph neural network(GNN)has received more and more attention in many research field, including social networking, knowledge graphs, recommendation systems, etc.[20, 26]. The powerful ability of GNN is to model the dependencies between nodes in graphs, and has made breakthroughs in the research fields related to graph analysis. A basic idea of graph neural network is that the node is embedding based on its local neighbor information[17, 24]. GNN is a kind of neural network which acts directly on graph structure. GNN has the following characteristics: the input order of nodes are ignore; in the process of calculation, the representation of a node is affected by its neighbor nodes, while the graph itself is connected unchanged; The representation of graph structure enables graph-based reasoning. Generally speaking, it is to aggregate the information of each node and its surrounding nodes through a neural network, Fig. 1 show the process of embedding:

Neural network is usually composed of two modules: propagation module and output module. Fig. 2 show the describes the basic structure of neural network:

The symbols and meanings involved in the neural network structure as follow: $G = (N, E)$, N is the set of vertices, E is the set of
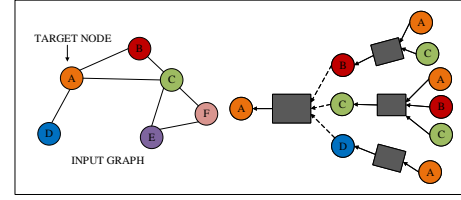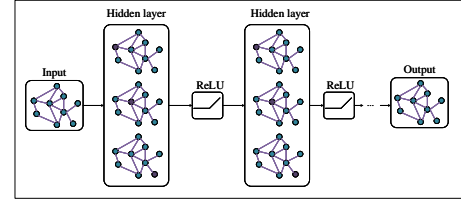

Figure 1: The process of embedding.

Figure 2: The basic structure of neural network.

edges; $ne[n]$ is the adjacency vertex of vertex n; $co[n]$ represents the edge attached to the vertex n;x_n represents the characteristics of vertex n; $x_co[n]$ represents the characteristics of the edges attached to the vertex n; $h_ne[n]$ represents the embedded representation of vertex n (state embedding); $x_ne[n]$ represents the characteristics of adjacent nodes of vertex n; h_n is the embedded representation of vertex n[2].

In propagation module of the GNN, the information is passed between nodes, and the status is updated, where there include aggregator and updater[5]. Aggregator: its purpose is to aggregate the information of the nodes around the node n, and to learn the embedding representation of node n, h_n, and h_n is represented as

$$h_n = f(x_n, x_{co[n]}, h_{ne[n]}, x_{ne[n]}) \qquad (3)$$

where f can be interpreted as a feedforward fully connected neural network. Updater: the embedded representation of node N in during the learning process the model iteratively updates is the following formula

$$H^{t+1} = F(H^t, X) \qquad (4)$$

where, t is the iteration of tth, X is all the features, and H^t is the embedded representation of all nodes in the tth iteration.

The output module of the GNN can be divided into two forms according to the different application types of the model:node-focused and graph-focused[25]. For node-focused classification, the label corresponding to each node after the last iteration is outputted by the model,

$$o_n = g(h_n^T, x_n) \qquad (5)$$

where t_i is the true label of the ith node, and o_i is the model output label, and p is the number of the node. For the graph-focused classification, the READOUT function is used to aggregate node features from the last iteration, and to obtain the representation vector $h(G)$ of the entire graph.

$$h(G) = READOUT(h_v^K | v \in G) \qquad (6)$$

The READOUT function can be a simple permutation invariant function

3 MODEL DESIGN

In this section, we mainly proposed an novel method of social anomaly event detection based on multi-source data; the method used spectral clustering algorithm to realize the feature space extension in multi-source data, and utilizes GNN model to achieve the anomaly event detection. Fig. 3 shows the design process of social anomaly detection based on multiple data sources.
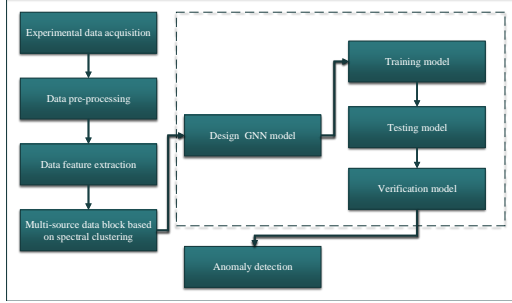

Figure 3: Design process of social anomaly detection.

The processes of abnormal event detectionin this paper include the below the steps shown:

- Data collection stage collect data samples related to the detection requirements;
- Data pre-processing includes data denoising and dimensionality reduction;
- Model design, according to the detection requirements the appropriate anomaly detection model is selected. Set the necessary parameters for the model.
- Model training is carried out and use data pre-processing to train model with supervised or unsupervised way, and this is the key step.
- In model testing, the result of testing is acquired by anomaly detection model with the test data set.
- Anomaly detection stage: the obtained model is used to detect anomalies in the target data.

3.1 Feature Space Based on Spectral Clustering

In this paper, the data generated from different data sources are analyze as a single view, and the multiple data are considered as multiple view learning. Samples from different data sources have different clustering forms, and the extended data set from multiple sources data can be considered that a description of the same problem from different angles[10]. In multiple source data sets, anomaly event is redefined as the sample is of inconsistent behavior in different data source , and the anomaly event of multiple view learning is defined as the sample belongs to the different clustering structure formed in different view spare learning[23].

We propose a consistent multi-data block anomaly detection method based on spectral clustering algorithm. In the method, the corresponding feature spaces of different data source are fused into an overall feature space, and using abnormal score to achieve anomaly judgement; in the data set, the abnormal score of the sample is defined by the each sample in different data source for the difference between the membership degree of each cluster structure[27].

In this paper, some important concepts are defined: the feature space of the different data source is called single feature space, the expression is $SF_t^{m \times d_k} = [A_t \in P^{m \times d_k}]$, where A_t represents the data of tth data source, d_k represents dimension; the feature space of each data source constitutes a full feature space, and the expression is $CSF^{m \times (\sum_{k=1}^{n} d_k)} = [CSF_1^{m \times d_1}, CSF_1^{m \times d_2}, \dots, CSF_k^{m \times d_k}]$. The difference between membership degree of anomaly data in full feature space is bigger than it in single feature space[29].

3.2 Social Anomaly Detection based on GNN

By spectral clustering, a full feature space of anomaly event based on multi data source and the characterization coefficient matrix is defined. First the model uses GNN to extracts the network element information (node, edge) of the full feature space at t, to form characterization of the feature, and then using the unsupervised representation learning algorithm DGI of the graph take the entire network at current moment to represent as a one-dimensional vector; Using RRCF algorithms to obtain the abnormal score of every moment[1]. In order to effectively use the graph information at every moment, we use long short term memory network (LSTMs) to obtain and process the change information of the global network representation at every moment. The framework of GNN algorithm is show in Fig. 4:
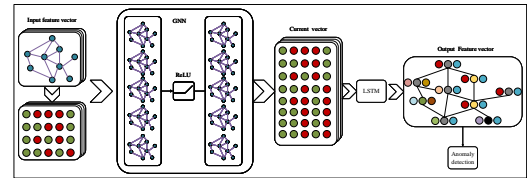

Figure 4: Framework of GNN algorithm.

4 RESULTS AND DISCUSSIONS

In this section, social network structure data and social text data collected online will be respectively used to evaluate the proposed method of this paper.

4.1 Experimental Data

In the paper, we use real multi-sources data set for experiment, and the data set is main from the micro-blog data set and sina social platform data set. Twenty thousand social message records are randomly sampled from the collected social micro-blog data set and sina social platform; the data set includes text data, picture data, and video data; In the datasets the 60% as a training sample, 20% as verification sample, and 20% as test sample; the 10,000 social micro-blog messages contain 7,880 manually marked messages of social events.

In order to eliminate the inaccurate data in the data set, and some irrelevant data are made by the network node exception, and some data collected repeatedly, and some data with wrong format or null value; these data need to be pre-processed to find out the data set; the data pre-processed is key step for mode, and feature representation provides reliable data guarantee for the algorithm[7]. In the paper, we use principal component analysis(PCA) as he

method of the data pre-processed. PCA is a multivariate statistical analysis method in which several variables are linearly transformed to select a small number of important variables. Also it is known as principal component analysis. The idea of PCA is to delete the redundant variables (closely related variables) for all the variables originally proposed, and establish as few new variables as possible, so that these new variables are paredly unrelated, and these new variables keep the original information as much as possible in reflecting the information of the subject. The main steps of principal component analysis are as follows:

- Standardization of index data;
- The correlation between the indicators;
- Determine the number of principal components m;
- Principal component expression;
- Principal component naming.

4.2 Experimental Result

To better verify the SP-GNN method we mentioned in this paper, the other methods Support Vector Machine(SVM) and convolutional neural network (CNN) are employed as a contrast. Accuracy and Macro $- F_1$ scores are selected as evaluation index of the model. In order to ensure the accuracy of the results, we repeated the experiment for three times and took the average value as the final result. The experimental results are shown in the Fig. 5.
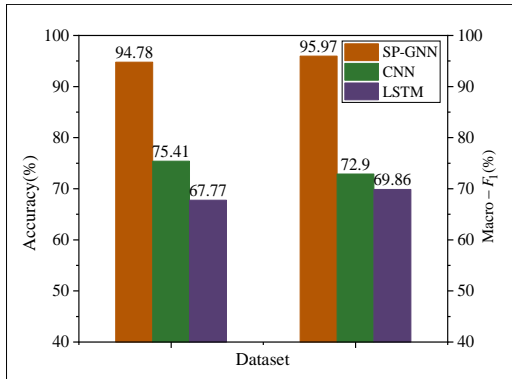

Figure 5: Comparison of experimental results.

Accuracy and Macro $- F_1$ of the SP-GNN model in social anomaly events is respectively better than the other method. The experimental results fully illustrate the proposed method is an effective means in in social anomaly events detection based on multi-data source. By drawing Receiver Operating Characteristic(ROC) curve to analyse the performance of the proposed method and baseline methds. The ROC curve is the corresponding values for each result of model anomaly. The results are demonstrated in the Fig. 6.

From the result, we can see that the area under the ROC curve is more bigger, and the performance of the model is better. The area under SP-GNN curve is the most biggest, and the performance of the SP-GNN method is superior to the baseline methods in exception object detection based on multi-data source. As can be seen from Fig. 7, SP-GNN generally gives a high score to the anomaly graphs in the 800 to 1400 plots where the anomaly is most concentrated.
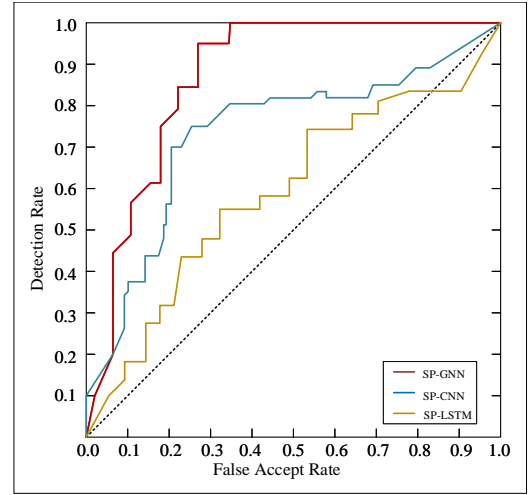

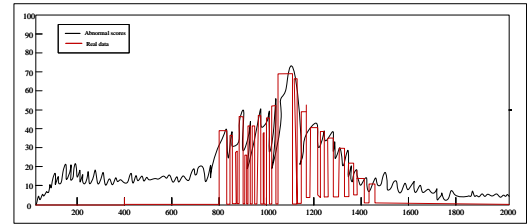**Figure 6: The ROC curve for results of anomaly detection.**

Figure 7: Anomaly score obtained by SP-GNN.

The clustering results of the proposed method are shown in Fig. 8, and the clustering results of CNN and LSTM are showed in Fig. 9, in Fig. 10 respectively. From the three figure, we can seen our method is superior to the other methods.

5 CONCLUSION

We address the problem of anomaly event detection in secure critical infrastructure domain. In the paper, we design a novel anomaly detection method based on multi-source data. We first leverage spectral clustering algorithm for feature extraction and fusion of multiple data sources. Second, we adopt deep graph neural network(Deep-GNN) to perform a fine-gained anomaly social event detection, which reveals the threatening events and guarantee the critical infrastructure security. Experimental results show that our method has good robustness and adaption. Compared with the other anomaly detection algorithms, our method has a good performance index.

ACKNOWLEDGEMENT

REFERENCES

[1] Aggarwal and C. Charu. 2013. *Outlier Detection in Categorical, Text and Mixed Attribute Data.* Springer New York. 199–223 pages.

(a) The result of SP-GNN.

(b) The result of CNN.
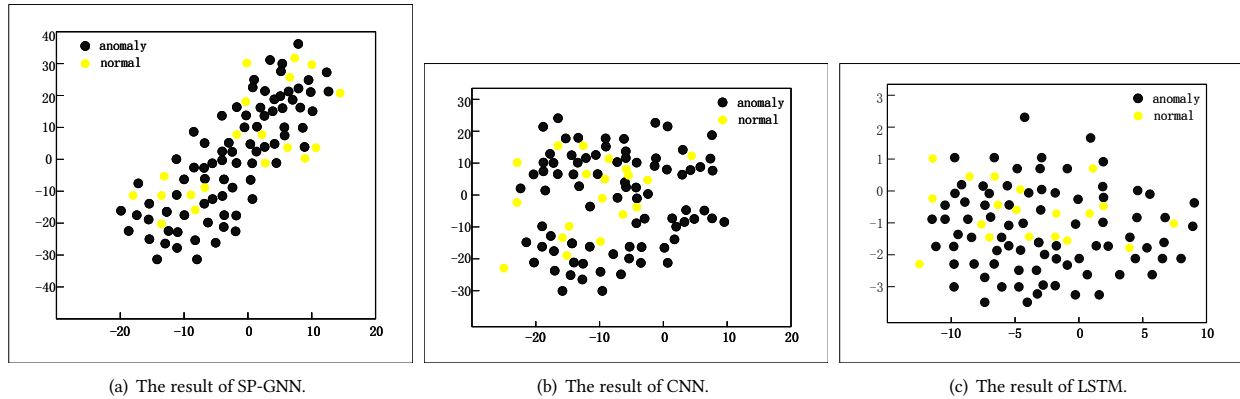

(c) The result of LSTM.

Figure 8: Experimental Results

[2] H Asai, Péter Zoltán Baranyi, and Sz Kovács. 2001. Modelling of guiding styles based on generalized neural network (GNN). *Soft Science Workshop at Kanazawa* (2001).

[3] Yuwei Cao, Hao Peng, Jia Wu, Yingtong Dou, Jianxin Li, and Philip S Yu. 2021. Knowledge-Preserving Incremental Social Event Detection via Heterogeneous GNNs. *arXiv preprint arXiv:2101.08747* (2021).

[4] Aditya Challa, Sravan Danda, B. S. Daya Sagar, and Laurent Najman. 2020. Power Spectral Clustering. *Journal of Mathematical Imaging Vision* 62, 9 (2020), 1195–1213.

[5] Songgaojun Deng, Huzefa Rangwala, and Yue Ning. 2020. Cola-GNN: Cross-location Attention based Graph Neural Networks for Long-term ILI Prediction. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management.*

[6] Yingtong Dou, Zhiwei Liu, Li Sun, Yutong Deng, Hao Peng, and Philip S Yu. 2020. Enhancing graph neural network-based fraud detectors against camouflaged fraudsters. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management.* 315–324.

[7] Andrew F. Emmott, Shubhomoy Das, Thomas Dietterich, Alan Fern, and Weng Keen Wong. 2013. Systematic Construction of Anomaly Detection Benchmarks from Real Data. In *Acm Sigkdd Workshop on Outlier Detection Description.*

[8] Chunming Hu, Jianyong Zhu, Renyu Yang, Hao Peng, Tianyu Wo, Shiqing Xue, Xiaoqiang Yu, Jie Xu, and Rajiv Ranjan. 2020. TOPOSCH: Latency-Aware Scheduling Based on Critical Path Analysis on Shared YARN Clusters. In *2020 IEEE 13th International Conference on Cloud Computing (CLOUD).* IEEE, 619–627.

[9] Pankaj Kumar Jadwal, Sonal Jain, and Basant Agarwal. 2020. Spectral Clustering and Cost-Sensitive Deep Neural Network-Based Undersampling Approach for P2P Lending Data. *International Journal of Information Technology and Web Engineering (IJITWE)* 15 (2020).

[10] Anna Koufakou, E. G. Ortiz, Michael Georgiopoulos, G. C. Anagnostopoulos, and K. M. Reynolds. 2007. A Scalable and Efficient Outlier Detection Strategy for Categorical Data. In *IEEE International Conference on Tools with Artificial Intelligence.*

[11] Chen Li, Xutan Peng, Shanghang Zhang, Hao Peng, S Yu Philip, Min He, Linfeng Du, and Lihong Wang. 2020. Modeling relation paths for knowledge base completion via joint adversarial training. *Knowledge-Based Systems* (2020).

[12] Yaopeng Liu, Hao Peng, Jianxin Li, Yangqiu Song, and Xiong Li. 2020. Event detection and evolution in multi-lingual social streams. *Frontiers of Computer Science* 14, 5 (2020), 1–15.

[13] Zhiwei Liu, Yingtong Dou, Philip S Yu, Yutong Deng, and Hao Peng. 2020. Alleviating the inconsistency problem of applying graph neural network to fraud detection. In *Proceedings of the 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval.* 1569–1572.

[14] Zheng Liu, Xiaohan Li, Hao Peng, Lifang He, and Philip S Yu. 2021. Heterogeneous Similarity Graph Neural Network on Electronic Health Records. *arXiv preprint arXiv:2101.06800* (2021).

[15] Qianren Mao, Xi Li, Hao Peng, Jianxin Li, Dongxiao He, Shu Guo, Min He, and Lihong Wang. 2021. Event prediction based on evolutionary event ontology knowledge. *Future Generation Computer Systems* 115 (2021), 76–89.

[16] Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. 2013. Clustering-Based Anomaly Detection in Multi-View Data. In *Proceedings of the 22nd ACM international conference on Conference on information knowledge management.*

[17] Hao Peng, Jianxin Li, Qiran Gong, Yuanxin Ning, Senzhang Wang, and Lifang He. 2020. Motif-Matching Based Subgraph-Level Attentional Convolutional Network for Graph Classification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 34. 5387–5394.

[18] Hao Peng, Jianxin Li, Qiran Gong, Yangqiu Song, Yuanxing Ning, Kunfeng Lai, and PS Yu. 2019. Fine-grained Event Categorization With Heterogeneous Graph Convolutional Networks. In *IJCAI International Joint Conference on Artificial Intelligence.* 32–38.

[19] Hao Peng, Jianxin Li, Yangqiu Song, and Yaopeng Liu. 2017. Incrementally learning the hierarchical softmax function for neural language models. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 31.

[20] Hao Peng, Jianxin Li, Senzhang Wang, Lihong Wang, Qiran Gong, Renyu Yang, Bo Li, Philip Yu, and Lifang He. 2019. Hierarchical taxonomy-aware and attentional graph capsule RCNNs for large-scale multi-label text classification. *IEEE Transactions on Knowledge and Data Engineering* (2019).

[21] Hao Peng, Renyu Yang, Zheng Wang, Jianxin Li, Lifang He, PS Yu, Albert Zomaya, and Raj Ranjan. 2021. LIME: Low-Cost and Incremental Learning for Dynamic Heterogeneous Information Networks. *IEEE Trans. Comput.* (2021).

[22] By Karl Rohe, Sourav Chatterjee, and Y. U. Bin. 2012. SPECTRAL CLUSTERING AND THE HIGH-DIMENSIONAL STOCHASTIC BLOCKMODEL1. (2012).

[23] David Savage, Xiuzhen Zhang, Xinghuo Yu, Pauline Chou, and Qingmai Wang. 2014. Anomaly detection in online social networks. *Social Networks* 39 (2014), 62–70.

[24] DAVID J. SEGAL, BIRGIT DRElER, and ROGER R. BEERLI. 1999. Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5-GNN-3 DNA target sequences. *Proceedings of the National Academy of Sciences of the United States of America* 96, 6 (1999), 2758–2763.

[25] Weijing Shi, Ragunathan, and Rajkumar. 2020. Point-GNN: Graph Neural Network for 3D Object Detection in a Point Cloud. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).*

[26] Qingyun Sun, Hao Peng, Jianxin Li, Senzhang Wang, Xiangyu Dong, Liangxuan Zhao, Philip S Yu, and Lifang He. 2020. Pairwise Learning for Name Disambiguation in Large-Scale Heterogeneous Academic Networks. *arXiv preprint arXiv:2008.13099* (2020).

[27] N. N. R. Ranga Suri, M. Narasimha Murty, and G. Athithan. 2016. Detecting outliers in categorical data through rough clustering. *Natural Computing* 15, 3 (2016), 385–394.

[28] Scott White and Padhraic Smyth. 2005. A Spectral Clustering Approach To Finding Communities in Graph. In *Proceedings of the 2005 SIAM International Conference on Data Mining.*

[29] Shu Wu and Shengrui Wang. 2013. Information-Theoretic Outlier Detection for Large-Scale Categorical Data. (2013).

[30] Stella X. Yu and Jianbo Shi. 2003. Multiclass Spectral Clustering. In *IEEE International Conference on Computer Vision.*

[31] Jun Zhao, Xudong Liu, Qiben Yan, Bo Li, Minglai Shao, and Hao Peng. 2020. Multi-attributed heterogeneous graph convolutional network for bot detection. *Information Sciences* 537 (2020), 380–393.

[32] Jun Zhao, Xudong Liu, Qiben Yan, Bo Li, Minglai Shao, Hao Peng, and Lichao Sun. 2021. Automatically predicting cyber attack preference with attributed heterogeneous attention networks and transductive learning. *Computers & Security* 102 (2021), 102152.