



SUCOLA: Self-adaptive structure refinement unsupervised contrastive learning framework for food safety risk early warning

Enguang Zuo^a, Junyi Yan^b, Alimjan Aysa^{a,c}, Chen Chen^a, Cheng Chen^b, Hongbing Ma^{a,d}, Xiaoyi Lv^{b,*}, Kurban Ubul^{a,c,**}

^a College of Information Science and Engineering, Xinjiang University, Urumqi, 830046, Xinjiang, China

^b College of Software, Xinjiang University, Urumqi, 830046, Xinjiang, China

^c Xinjiang Multilingual Information Technology Key Laboratory, Urumqi, 830046, Xinjiang, China

^d Department of Electronic Engineering, Tsinghua University, Beijing, 100084, China

ARTICLE INFO

Keywords:

Food computing
Label imbalance
Graph structure learning
Unsupervised learning
Contrastive learning

ABSTRACT

Food safety is a global issue; risk prevention and control of food physical and chemical testing (PCT) data are critical. However, existing studies have not considered the risk status of qualified products near the detection limit. Therefore, this paper presents a data-driven paradigm to detect the combined risk of multiple indicators across products in real-world food safety datasets for the first time. The following challenges exist to achieve this goal. (1) In real-world scenarios, neighboring samples are somewhat correlated; however, capturing and quantifying the actual testing sample correlation network remains an open problem in food computing. (2) The low-frequency nature of the unqualified samples leads to label imbalance in the data. (3) Labeling dependency limits the application scenarios. This article proposes a novel self-supervised framework (SUCOLA) to address the abovementioned challenges. That jointly learns the optimal graph topology and also inferences the connections between every product node and its well-designed subgraph, helping to resist label imbalance. We applied it to a real-world sterilized PCT dataset from a province in China. Extensive experiments show that SUCOLA performance is outstandingly better than several baselines of the current graph self-supervision field. Meanwhile, we conducted generalizability experiments on five well-known benchmark datasets. The proposed SUCOLA surpasses the current state-of-the-art by a considerable margin (0.39%–8.7% AUC) on public datasets. Thus, our study provides a new idea for early warning research on PCT data in food computing and fills the gap in the study of structural learning under category imbalance.

1. Introduction

Background. Food profoundly influences individual life in terms of health and nutrition and acts as an essential characteristic for differentiating our habitation, food culture, and social status (Khanna, 2009). Recently, frequently reported food safety incidents have caused a worldwide sensation. For example, the “Focus on CCTV 3.15” revealed the processing of grossly unhygienic foot-stamped sauerkraut by a well-known company, the illegal addition of propylene glycol to pure milk by McChrystal, and the aggregated cases of monophasic Salmonella Typhimurium sequence type 34 infections caused by substandard chocolate produced in Belgium, as notified by the World Health Organization (WHO).¹

Making precise assessments and timely warnings of food quality and safety risks is a hot topic for current research, seeking to limit the hazard of food safety issues to population health because diseases enter through the mouth (Tang et al., 2002). Risk assessment studies can help grasp the trends of food safety risks and support market surveillance authorities in risk prevention and control. Food physical and chemical testing (PCT) reports containing specific values for each detecting index, meaning that a wealth of knowledge can be mined and objective food safety risks can be quantified, thus providing excellent opportunities for research related to food calculations (Zuo et al., 2022b).

Motivation. Specifically, previous works only considered single-sample risk without considering the impact of contextual information

* Corresponding author.

** Corresponding author at: College of Information Science and Engineering, Xinjiang University, Urumqi, 830046, Xinjiang, China.

E-mail addresses: zeg@stu.xju.edu.cn (E. Zuo), yjy@stu.xju.edu.cn (J. Yan), alim@xju.edu.cn (A. Aysa), chen_chen@stu.xju.edu.cn (C. Chen), chenchengoptics@xju.edu.cn (C. Chen), hbma@tsinghua.edu.cn (H. Ma), lxiaoyi@xju.edu.cn (X. Lv), Kurbanu@xju.edu.cn (K. Ubul).

¹ www.woyaoce.cn

(which is crucial for food risk assessment). Additionally, all training phases were performed under supervised settings, which increased costs and limited the model's applicability scenarios. Therefore, the main motivation of this paper is based on the hypothesis that there should be certain relationships among food samples (e.g., samples from the same batch, same device, etc.). For the first time, we explore how to mine potential correlation networks between food inspection and detection data samples in an unsupervised manner, fully considering the context of the samples, thus accurately identifying potential risk samples and issuing warnings.

Problem. To facilitate the food risk assessment advance, this paper collects a real dataset of sterilized milk PCT reports from the Institute of Food Quality Supervision and Inspection, with which the research team collaborates, where each sample contains six detecting indicators and the total number of samples is 2158 cases. Based on this dataset, this paper further proposes and studies the validation of the adaptive detecting data risk analysis task.

Currently, the PCT method only evaluates whether a product exceeds the standard limit based on a single indicator. However, the safety of a food product involves the comprehensive evaluation of multiple indicators. We believe the testing standard for unqualified detecting indicators only indicates the minimum standard for Food reaching the market. Qualified samples close to the detection limit should also have some risk.² Furthermore, each indicator impacts human health and needs to be considered comprehensively (Geng et al., 2017). As a result, the paper attempt to develop a data-driven approach for assessing the risk of all food products using a comprehensive multi-indicator approach.

Challenges. Previous research on food risk early warning based on detecting data employs hierarchical analysis to calculate the corresponding weights for all indicators, followed by weighted summation to obtain the product's total risk value. Finally, the risk is fitted by a model to achieve prediction (Geng et al., 2017; Lin et al., 2021; ZhiQiang et al., 2021). Compared to it, this study faces some unique challenges considering the nature of detecting data and the sample distribution's characteristics.

1. Firstly, in the food testing scenario, there is some association of detection values of neighboring samples, which inspires this paper to investigate further this potential association network that can provide additional knowledge. Still, the manually predefined association network is not data-driven and suffers from noise, such as incomplete and incorrect links (Zhu et al., 2021; Liu et al., 2022). Therefore, capturing and quantifying the real-world testing sample association network is the first challenge of this paper.
2. Further, since the vast majority of the data collected are qualified samples, there are only a few dozen or fewer cases of unqualified samples. Therefore, the low frequency of non-conforming samples leads to the label imbalance problem as the second challenge.
3. Finally, expert panel manual annotation labels are essential in providing target signals to guide parameter optimization in supervised feature extraction methods. However, this reliance on labels will limit the application of the model to more general scenarios where annotations are unavailable. The third challenge of this paper is how to use unsupervised training to achieve risk warning for samples and generalize it for application in more generalized scenarios.

Solution. To address the abovementioned challenges, this paper innovatively proposes a Self-adaptive structure refinement Unsupervised COntrastive LeArning framework for food safety risk early warning,

called SUCOLA. As shown in Fig. 1, in our learning paradigm, to meet the first obstacle, we aim to identify a potential sample association network capable of capturing the sample space within the association structure. To accomplish this, SUCOLA introduces a graph message aggregation and a structure learning method to jointly learn how to modify the initial fine-grained graph's topology. SUCOLA employs a contrastive learning approach for the second challenge to overcome the severe imbalance labeling problem (Zhao et al., 2021). Specifically, SUCOLA concentrates on constructing the association between each target node and the corresponding meticulously designed local subgraph, which can expose the substandard samples in the network. Xu et al. (2021). Meanwhile, the learning objective of SUCOLA is set to differentiate the consistency among the components in the instance pairs. Therefore, its results (risk scores) can be further utilized to evaluate the risk status of the target nodes. Both of the above steps use self-supervised contrastive learning for structure refinement and risk detection to address the third challenge. As the contrast loss decreases, all node representations are given the necessary supervised signals.

Ultimately, with all the technical designs mentioned above, the model in this paper can capture the association network in the sample space with the ability to achieve food safety alerts for risky samples. In brief, this paper has the following highlights:

1. We present and study a data-driven self-supervised paradigm for risk early warning of food PCT datasets for the first time, which is more widely applied and challenging than the models in a supervised scenario.
2. We propose the self-supervised model SUCOLA aimed at collaborative learning to optimize the association network in the sample space. Based on this, the feature representations of sample nodes with their local subgraphs are modeled to achieve risk warnings under imbalance labeling conditions.
3. Extensive experiments on one real-world food PCT dataset are conducted in this paper, and case studies show that the model in this paper has a reasonable risk warning effect. Finally, this research undertakes several extended experiments using five benchmark dataset datasets. The strong generalizability of our model is illustrated by the empirical results, which reach the best results for each dataset.

2. Related works

2.1. Food safety risk early warning

Risk assessment for food safety is a popular research direction in food computing. The research analyzes heterogeneous food data from different sources through computational methods, aiming to solve the problems of food risk screening, early warning, and control (Min et al., 2019).

Risk Screening. Social media is used to screen key stores and monitor public health (Capurro et al., 2014). Several studies have used a data mining approach to analyze people's food consumption from large-scale records in social media applications. Ofli et al. (2017), Silva et al. (2014). Studies Sadilek et al. (2016) and Zuo et al. (2022a) prevent foodborne diseases by mining the end-retailer with potential food safety risks in social media. They applied artificial intelligence (AI) technology to Twitter and Meituan takeaway comment data to automatically screen the end-retailer that may pose a public health hazard.

Risk early warning and control. Many scholars have conducted related research to understand the weight share among testing indicators, explore the correlation relationship among testing samples, and obtain a more accurate and efficient risk warning effect. The effectiveness of Bayesian networks heavily depends on anthropogenic factors because the model structural design is typically dictated by expert expertise. Bouzembrak and Marvin (2016). In contrast, the

² http://www.gov.cn/zhengce/zhengceku/2020-01/14/content_5468984.htm

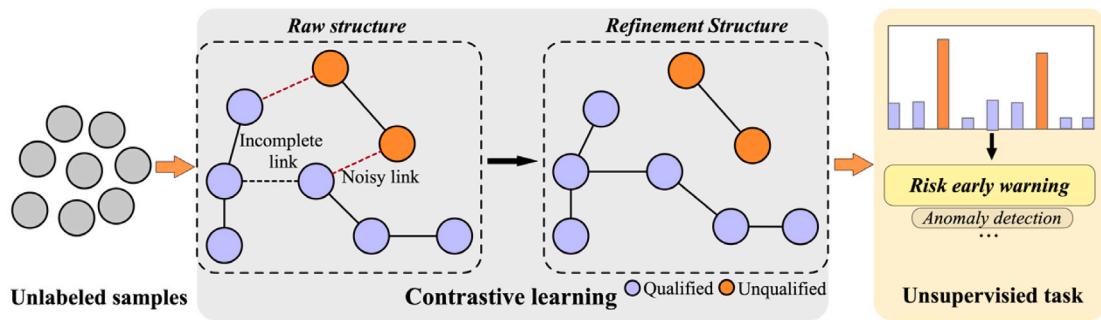


Fig. 1. A tony example to illustrate the data-driven self-supervised food safety warning paradigm.

model parameters of artificial neural networks (ANNs) are directly updated by gradient descent, which does not depend on the experience of experts. The ANN-based models are nonlinearly activated and have better robustness and data-fitting ability. Therefore, ANN techniques have been widely used in the food computing domain. To determine the overall weights of attributes based on the contributions of each indicator of the detected samples, Samuel et al. (2017), ZhiQiang et al. (2021) and Lin et al. (2021) all used analytical hierarchy process (AHP) techniques. These techniques were then combined with various ANN-based feature extractors to predict the risk status of the samples. Thus, previous research methods isolate sample associations and require pre-processing food data to calculate model expectation outputs (supervised). In contrast, the unsupervised comparative learning approach based on structure learning used in this paper efficiently enables food safety risk assessment.

2.2. Graph structure learning

Structure learning seeks to optimize the representation of the original structure, which traces back to network science prior works (Egilmez et al., 2017; Brashears, 2014; Xia et al., 2022; Jiao et al., 2020; You et al., 2020). This paper focuses on one research branch: graph neural networks (GNNs)-based graph structure learning (GSL) models. Chen et al. (2020b) forced GSL into a similarity metric learning problem and proposed a structural model describing the underlying graph optimization and an iterative framework based on Bayesian inference calculating the similarity of neighborhood information. In Wang et al. (2019), the authors create a new graph in each GNN layer based on the node embedding similarity. In Halcrow et al. (2020), several similarity metrics are fused instead of choosing a single similarity metric. Through self-supervised adjacency and the concurrent learning of GNN parameters, Fatemi et al. (2021) offers an effectively supervised method for inferring about graph structure. Kumagai et al. (2021) considers the nodes' properties, the graph's topology, and the class imbalance. Few-shot label information can be effectively propagated by this paradigm to unlabeled nodes. Liu et al. (2022) advocates an unsupervised GSL approach to cope with the label dependency problem. Nevertheless, existing GSL methods have application scenario limitations, such as not considering a graph optimization approach under extreme label imbalance. This paper adopts a more practical structure-guided approach to cope with graph structure learning under food risk analysis tasks.

2.3. Graph anomaly detection

Graph anomaly detection aims to detect nodes containing erroneous information using node attribute and edge topology information. Early studies such as Radar by Li et al. (2017) and ANOMALOUS by Peng et al. (2018) built shallow models based on residual analysis to detect graph anomalies. However, the representation ability of such models is affected by network depth, making it difficult to achieve

satisfactory performance on complex datasets. With the rise of graph neural networks, the ability of models to mine deep nonlinear information has significantly improved, promoting the rapid development of graph anomaly detection. Based on reconstruction, DOMINANT proposed by Ding et al. (2019) is a GCN-based autoencoder that integrates structural information into node embedding representation and detects anomalies based on node reconstruction error. On this basis, ALARM proposed by Peng et al. (2020) improves the encoder stage to use multi-perspective GCN to process information. Based on contrastive learning, CoLA proposed by Liu et al. (2021) estimates the relationship between the target node and its positive and negative subgraphs through self-supervised contrastive learning to discriminate abnormal modes. ANEMONE proposed by Jin et al. (2021), Sub-CR by Zhang et al. (2022), and GRADATE by Duan et al. (2022) use enhanced multi-view contrastive learning for graph anomaly detection. Unlike the above models, this paper considers amplifying the discrimination between anomalies and normals through structure-guided methods.

3. Preliminaries

In this paper, we focus on how to mine food safety risk samples using potential association networks of PCT datasets. First, in line with current works on GNN, we set $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{X}) = (\mathbf{A}, \mathbf{X})$ as the PCT data attribute network (Jin et al., 2021; Ji et al., 2021), where $\mathbf{A} \in [0, 1]^{n \times n}$ refers to the adjacency matrix ($A_{i,j}$ means the weights of the edges between sample nodes v_i and v_j), and $\mathbf{X} \in \mathbb{R}^{n \times m}$ denotes the feature embedding matrix of \mathcal{G} , where n and m denote the nodes number and the indicators number in \mathcal{G} , respectively. With the notations mentioned above, the studied problem definition can be formalized as follows:

Algorithm Definition. (Unsupervised food safety risk early warning on noisy attributed networks).

Given an attribute network $\mathcal{G} = (\mathbf{A}, \mathbf{X})$ with artificial predefined graph structure (with noisy) \mathbf{A} for a given food PCT dataset. The objective is to explore a data-driven approach, $f(\cdot)$, to estimate the risk value $r_i = f(x_i)$ for each product. The risk value r_i can represent the warning level of sample x_i . By arranging the risk values of all samples in descending order, abnormal risk products can be easily determined in their early warning levels according to their positions.

This paper uses an unsupervised training approach to achieve an early warning of the food safety risk, widely used in previous data mining tasks such as graph anomaly detection, graph node classification, etc. Based on this setting, only $\mathcal{G} = (\mathbf{A}, \mathbf{X})$ is available in the training phase of the model in this paper, and all samples are not known whether they are qualified or not. In Table 1, we have summarized the high-frequency terminology used in this paper to make it easier for readers.

Task Definition. (Risk prevention and control of sterilized PCT dataset).

We aim to comprehensively consider multiple detection indicators to address the evaluation of the nutritional quality and environmental pollution level of dairy products. The evaluated indicators in this study

Table 1
Description and explanation of essential symbols in the SUCOLA framework.

Notation	Description and explanation
$G = (\mathcal{V}, \mathcal{E}, \mathbf{X})$	An attributed network of the constructed detecting product
m	The number of product-detecting indicators
n	The number of nodes in the initial graph G
\mathcal{V}, \mathcal{E}	The product and link set of the initial graph G
$\mathbf{X} \in \mathbb{R}^{n \times m}$, $\mathbf{A} \in [0, 1]^{n \times n}$	The feature embedding matrix and adjacency matrix of the initial graph G
$\mathbf{A}_a \in \mathbb{R}^{n \times n}$, $\mathbf{S} \in \mathbb{R}^{n \times n}$	The anchor and learned adjacency matrix
$\mathbf{H}_l, \mathbf{H}_a \in \mathbb{R}^{n \times d_1}$	The representation matrix encoded by GNN
$\mathbf{Z}_l, \mathbf{Z}_a \in \mathbb{R}^{n \times d_2}$	The representation matrix readout by MLP
$\mathbf{S}_r \in \mathbb{R}^{n \times n}$	The adjacency matrix for structural fusion graphs
$P_i = (v_i, \mathcal{G}_i)$	The product instance pair of the i th target node
v_i	The target node in the instance pair
\mathcal{G}_i	The local subgraph of the target node
$\mathbf{e}_i \in \mathbb{R}^{d_3}$	The representation vector of v_i
$\tilde{\mathbf{e}}_i \in \mathbb{R}^{d_4}$	The representation vector of \mathcal{G}_i
$\mathcal{P} = \{\mathcal{P}_{pos}, \mathcal{P}_{neg}\}$	The positive and negative example pair pools
$\hat{y}_i \in [0, 1]$	The prediction score of the model for v_i
$\hat{y}_i \in \{0, 1\}$	The labels of instance pairs
$f_{risk}(v_i)$	The risk score predicted by the model for v_i
$o_i = f_{level}(v_i)$	The risk warning level of the model for v_i
\mathcal{L}_s	The structural optimization loss of the model
\mathcal{L}_r	The risk analysis loss of the model
γ	The structure fine-tuning rate
θ, ϕ, θ_e and θ_d	The trainable parameters
R	The sampling rounds for the risk estimator
K	The local subgraph size

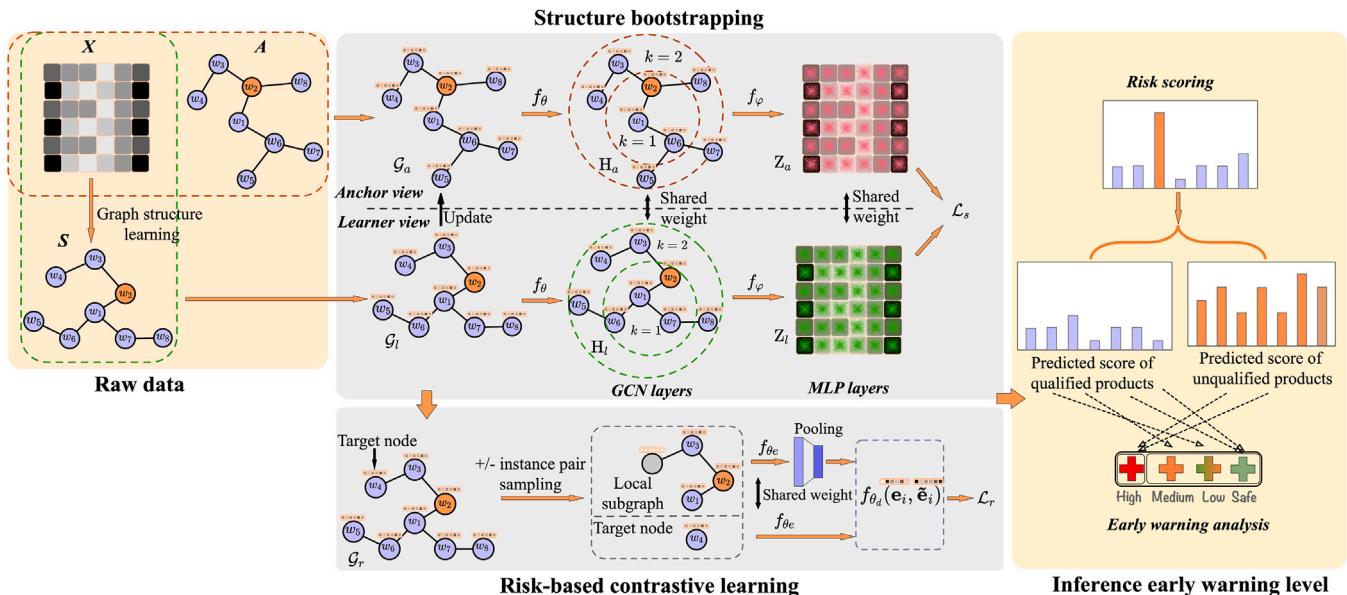


Fig. 2. The framework overview of the SUCOLA. Here, k represents the number of GCN layers, w_i denotes the product sample id, $f(\cdot)$ is the non-linear function, and θ, ϕ, θ_e , and θ_d represent the weights of shared layers. The three components of SUCOLA are graph structure generation (left), feature extraction (middle), and risk warning (right). Qualified products are classified into three categories: safe, low-risk, and medium-risk, while unqualified products are labeled as high-risk.

include nutritional indicators such as lactose, fat, protein, and non-fat milk solids, as well as environmental indicators such as acidity and aflatoxin M1.

4. Methodology

We will describe the details and principles of the SUCOLA framework in this section. As indicated in Fig. 2, SUCOLA is made up of three parts: graph structure generation (right), feature extraction (middle), and risk early warning (left). The first part, graph structure generation, the anchor graph adjacency matrix, and the learning graph adjacency matrix are obtained through the predefined and graph learning modules, respectively. The second part of feature extraction can be further

subdivided into structure-guided and graph risk detection phases. In the structure bootstrapping phase, we design a hierarchical structure bootstrapping (HSB) module to optimize the graph topology when the anchor and learning graphs are learned in comparison at the node level; In the graph risk detection phase, we sample and model the optimized graph with instance pairs (i.e., each sample node with its well-designed local subgraphs is modeled). The learning objective of SUCOLA is to differentiate the consistency among the components in the instance pairs. The risk ratings can also determine the risk status of the target nodes. In the last part, qualified items are categorized into the three warning categories: safety, low-risk, and medium-risk, while unqualified products are marked as high-risk.

In summary, each model component is designed to describe the changes in the data at three stages (graph data → feature matrix →

risk level). The specifics of each SUCOLA component will be discussed in the following subsections.

4.1. Graph structure generation

As shown in the first part of Fig. 2, this step aims to determine the adjacency matrix in the different graph views. Next, we will detail demonstrate how to construct these two different matrices.

The anchor matrix. The anchor view serves as a “teacher” by providing a reliable and accurate guide for adjusting the graph structure. For food detecting data, we define $\mathcal{G} = (\mathbf{A}_a, \mathbf{X}) = (\mathbf{A}, \mathbf{X})$ as the anchor graph. The formal representation of link weights between sample pairs $(\mathbf{X}[i, :], \mathbf{X}[j, :])$ is as follows.

$$A_{i,j} = \text{Euc}(\mathbf{X}[i, :], \mathbf{X}[j, :]) = \|\mathbf{X}[i, :] - \mathbf{X}[j, :]\|_2 \quad (1)$$

$A_{i,j}$ is the strength of the link between the product nodes $\mathbf{X}[i, :]$ and $\mathbf{X}[j, :]$, and $\text{Euc}(\cdot)$ denotes the Euclidean distance. Since there is a negative correlation between $A_{i,j}$ and $\text{Euc}(\mathbf{X}[i, :], \mathbf{X}[j, :])$, the more similar the feature vectors of samples i and j are, the smaller the Euclidean distance between them becomes, resulting in $A_{i,j}$ approaching 1. The anchor view does not keep the original structure all the time but is updated by a novel structural fine-tuning mechanism, which was introduced in Section 4.2.1.

The learned matrix. The learning view plays the role of a “student” and is a vital part of graph restructuring. As in existing work (Liu et al., 2022; Franceschi et al., 2019; Chen et al., 2020b), we first generate a sketched adjacency matrix $\tilde{\mathbf{S}}$ by a multi-layer perception (MLP). Then optimized utilizing post-processing, a high-quality adjacency matrix \mathbf{S} is produced. Both \mathbf{S} and the parameters θ used for modeling are updated directly by gradient descent to learn the better graph structure.

$$\tilde{s}_{ij}^{(sp)} = \begin{cases} \tilde{s}_{ij}, & \tilde{s}_{ij} \in \text{top } -k(\tilde{\mathbf{S}}[i, :]) \\ 0, & \tilde{s}_{ij} \notin \text{top } -k(\tilde{\mathbf{S}}[i, :]) \end{cases} \quad (2)$$

$$\tilde{\mathbf{S}}^{(sy)} = \frac{f_q(\tilde{\mathbf{S}}^{(sp)}) + f_q(\tilde{\mathbf{S}}^{(sp)})^\top}{2} \quad (3)$$

$$\mathbf{S} = \left(\tilde{\mathbf{D}}^{(sy)}\right)^{-\frac{1}{2}} \tilde{\mathbf{S}}^{(sy)} \left(\tilde{\mathbf{D}}^{(sy)}\right)^{-\frac{1}{2}} \quad (4)$$

As in Equations (2)–(4), the sparsification, non-negative, symmetrization, and normalization are implemented sequentially for the sketched adjacency matrix $\tilde{\mathbf{S}}$. Where $\text{top } -k(\tilde{\mathbf{S}}[i, :])$ denotes the highest k values of the highest similarity of the i th row of $\tilde{\mathbf{S}}$. $(\cdot)^\top$ is the transposition function, $f(\cdot)$ is an activation function and $\tilde{\mathbf{D}}^{(sy)}$ denotes the degree matrix of $\tilde{\mathbf{S}}^{(sy)}$.

4.2. Feature extraction

As shown in the second part of Fig. 2, although we have obtained the adjacency matrices of the anchor and learned graphs, the interaction between the two matrices and the association structure of the actual data is still unknown. Therefore, this paper proposes a hierarchical structure bootstrapping (HSB) module, which aims to identify the potential sample association network that reflects the association structure in the sample space. Moreover, for the optimized association structure, we further propose a risk-based contrastive learning module to reach the risk detection and early warning of network nodes in a data-driven manner.

4.2.1. Structure bootstrapping phase

For this paper, the challenge of bootstrapping the graph structure is how to quantify the effect of one sample on another sample. We propose an HSB module to resolve this issue. First, we fine-tune the anchor structure \mathbf{A}_a with the learned structure \mathbf{S} by setting a fine-tuning structure mechanism; then, as the contrast loss decreases and after the gradient converges (i.e., the structure is fixed), we set a structure fusion

module to learn \mathbf{S} and \mathbf{A}_a for downstream risk detection jointly. Next, we introduce the hierarchical bootstrap module in this paper.

Prerequisite. Maximizing the mutual information (MI) across two perspectives through contrastive learning should be considered before doing structural bootstrapping. In line with Velickovic et al. (2019), SUCOLA adopts a standard contrastive learning structure whose formal definition is shown below.

$$\mathbf{H}_a = f(\mathcal{G}_a(\mathbf{X}, \mathbf{A}_a), \theta) \quad (5)$$

$$\mathbf{H}_l = f(\mathcal{G}_l(\mathbf{X}, \mathbf{S}), \theta) \quad (6)$$

GNN-based encoder. As shown in Equations ((5), (6)), $f(\cdot)$ denotes the nonlinear encoder GCNs (Kipf and Welling, 2017) used in this paper, and the layer is set to 2. $\mathcal{G}_a(\cdot)$ and $\mathcal{G}_l(\cdot)$ represent the anchor and learning graphs, respectively, and θ denotes the learnable parameters. \mathbf{H}_a and \mathbf{H}_l denote the anchor and learning perspective feature vector matrix, respectively.

$$\mathbf{Z}_a = g(\mathbf{H}_a, \varphi) \quad (7)$$

$$\mathbf{Z}_l = g(\mathbf{H}_l, \varphi) \quad (8)$$

Readout projector. As shown in Equations ((7), (8)), where $g(\cdot)$ denotes the nonlinear readout function MLP used in this paper, φ denotes the learnable parameters. \mathbf{Z}_a and \mathbf{Z}_l represent the feature vector matrix of the anchor and learning perspectives, respectively.

Contrastive module. Given the feature vector matrices \mathbf{Z}_a and \mathbf{Z}_l for the anchor and learning perspectives, we use the contrast loss L to force the consistency of the corresponding node representations (e.g., node representations $\mathbf{Z}_a[i, :]$ and $\mathbf{Z}_l[i, :]$) in the matrix. In line with NT-Xent (van den Oord et al., 2018; Jin et al., 2021), the loss function employed in SUCOLA is defined as follows:

$$\ell_a(\mathbf{Z}[i, :]) = -\log \frac{e^{\cos(\mathbf{Z}_a[i, :], \mathbf{Z}_l[i, :])/t}}{\sum_{k \neq i} e^{\cos(\mathbf{Z}_a[i, :], \mathbf{Z}_l[k, :])/t}} \quad (9)$$

$$\ell_l(\mathbf{Z}[i, :]) = -\log \frac{e^{\cos(\mathbf{Z}_l[i, :], \mathbf{Z}_a[i, :])/t}}{\sum_{k \neq i} e^{\cos(\mathbf{Z}_l[i, :], \mathbf{Z}_a[k, :])/t}} \quad (10)$$

$$\mathcal{L}_s = \frac{1}{2n} \sum_{i=1}^n [\ell_a(\mathbf{Z}[i, :]) + \ell_l(\mathbf{Z}[i, :])] \quad (11)$$

where $\cos(\cdot)$ represents the similarity function of cosine-based, and t shows the hyperparameter. In addition, $\ell_a(\mathbf{Z}[i, :])$ and $\ell_l(\mathbf{Z}[i, :])$ are two symmetric losses that represent comparisons from different viewpoints.

Structural fine-tuning. Previous studies have shown that the strategy of learning the graph structure \mathbf{S} through the original data or a predefined \mathbf{A} as a fixed anchor adjacency matrix \mathbf{A}_a and then by maximizing the mutual information (MI) between the two views leads to the inheritance of \mathbf{S} to \mathbf{A}_a misinformation (e.g., missing links, wrong links) (Liu et al., 2022).

We develop a structural fine-tuning module that updates the \mathbf{A}_a with a fine-tuning rate enhancement of the learning structure \mathbf{S} . Specifically, given a structure fine-tuning rate $\gamma \in [0, 1]$, the \mathbf{A}_a update detail as follows.

$$\mathbf{A}_a \leftarrow \gamma \mathbf{A}_a + (1 - \gamma) \mathbf{S} \quad (12)$$

Structural fusion. As the loss decreases, the difference between \mathbf{S} and \mathbf{A}_a also decreases. However, even if \mathbf{S} converges to a certain level and remains stable, the difference with \mathbf{A}_a still exists, and using only \mathbf{S} or \mathbf{A}_a will result in different degrees of structural loss. This loss leads to the degradation of GNN performance for downstream tasks, so we develop a structure fusion operation to enhance the representation of data association networks with the learned structure \mathbf{S} and the fused structure of \mathbf{A}_a as follows:

$$\mathbf{S}_r \leftarrow s_r = \begin{cases} \frac{s_{ij} + a_{ij}}{2}, & \mathbf{S}[i, j] > \min(\mathbf{D}_s) \\ a_{ij}, & \mathbf{S}[i, j] \leq \min(\mathbf{D}_s) \end{cases} \quad (13)$$

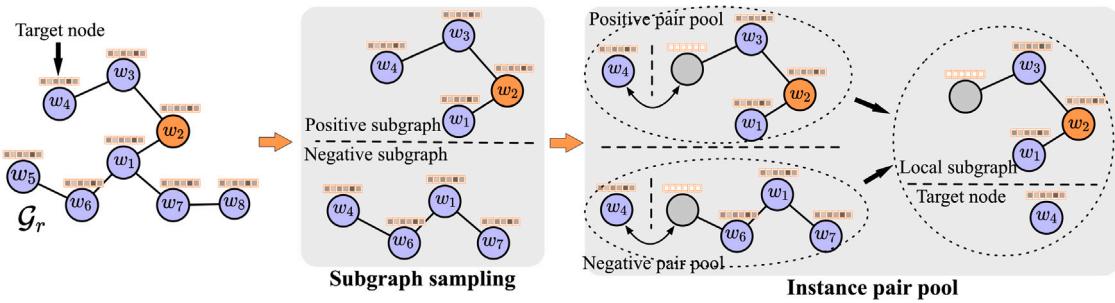


Fig. 3. Sampling process for example pairs.

where S_r represents the sample hidden structure adjacency matrix for subsequent risk analysis after structure fusion, s_r represents the elements in matrix S_r , s_{ij} and a_{ij} represent the element values in the learned matrix and anchor matrix (S and A_a), respectively, and $\min(D_s)$ represents the minimum value in the set of main diagonal elements.

Benefiting from the HSB module, SUCOLA has good properties to solve the above problem (Mining potential association networks between samples). Due to the structure fine-tuning mechanism, some edges carrying noise are gradually reduced in A_a during the updating process. Furthermore, it will effectively suppress their negative impact on downstream tasks while driving the model to continuously discover increasingly superior graph structures S graph. More importantly, due to the structure fusion mechanism, the structure S_r of the final detecting data incorporates valid information from both perspectives to guide the subsequent risk detection learning.

4.2.2. The risk-based contrastive learning phase

Due to the low frequency of the risky samples, they usually do not match well with their local substructures and globally randomly obtained substructures. In contrast, normal samples are well-matched with their local structures and globally randomly acquired substructures. Based on this observation, we design a sampling of instance sample pairs to find mismatched risky samples through learning the “node vs local subgraph” matching pattern. As shown in Fig. 3, we thoroughly introduce the sampling processing.

Target node sampling. Since we want to perform a risk assessment on each sample node in the graph, we first need to perform target node selection (i.e., the sample to be evaluated). In this section, we extract the target node in turn randomly from the given input graph G_r without put-back.

Local subgraph sampling. The motivation for performing subgraph sampling is to capture the substructure around the target node that is highly correlated with the features of the node (Liu et al., 2021) (optimized correlation structure increases the reliability of information transfer) and to provide sufficiently diverse input data for risk assessment and risk classification. Considering the above factors, we utilize random wandering with restart (RWR) (Tong et al., 2006) as a sampling strategy to avoid corrupting the semantic information learned during the structure bootstrap phase. The target node is set initially, and then the positive and negative subgraphs are sampled by random wandering. The initial node of the negative subgraph is chosen randomly, whereas the initial node of the positive subgraph is defined as the target node. The subgraph size is K .

Initial node masking. Inspired by Liu et al. (2021), the initial nodes in the local subgraph are anonymized (i.e., $X_i^{(i)}[i, :] \leftarrow \bar{0}$) such that the original feature information of the target node will have no impact on its feature reconstruction and the computation of view embedding. This mechanism can prevent target node information from contaminating other nodes and induce the model to identify risky nodes through contextual information. Finally, we combine the target node with the

corresponding positive and negative subgraphs to form the instance pair $P_i = (v_i, G_i)$ (Where i represents the target node index), and then put it into the instance pair pool $\mathcal{P} = \{P_{pos}, P_{neg}\}$.

Contrastive Learning. SUCOLA obtains high-quality potential spatially correlated structures with the help of the structure bootstrapping module, but the structure information is currently not directly exploited. We propose a pair contrastive module to use the rich structural information fully. Specifically, our contrastive module inputs the feature vector selected as the target node and its subgraph (i.e., sampled instance pair P_i). Then, we discriminate whether the target node is a risky sample by identifying the correlation between the target node and its subgraph in instance pair P_i .

As in Fig. 2, first, v_i and its local subgraph G_i are each fed into a graph-based encoder. The encoders have the same parameters as expressed in Equations (5), (6), that is, $E_l = f(G_i(X_i, S_i), \theta_e)$, $e_l = \text{ReLU}(X[i, :], \theta_e)$.

Next, a Readout projector is set up, whose purpose is to transform the feature matrix X_l generated by the GNN encoding of the local subgraph in the instance pair into the same dimension as the GNN encoded feature vector e_l of the target node. This study selects the average pooling function for the pooling strategy, which can be formulated as follows:

$$\tilde{e}_i = \frac{1}{k} \sum_{j=1}^k X_{li}[j, :] \quad (14)$$

where \tilde{e}_i is the feature representation of the local subgraph G_i of v_i , k denotes the subgraph nodes number, and $X_{li}[j, :]$ expresses the vector of node j in the subgraph embedding vector matrix.

Generally, a risky node on a graph structure should differ from its surroundings regarding properties or topology. Inspired by this, we formed our comparison objective similar to Velickovic et al. (2019), which maximizes the consistency between each sample node and its local contextual background. The formalization is shown in Equation (15).

$$\mathcal{L}_r = -\frac{1}{2n} \sum_{i=1}^n (y_i \log(\hat{y}_i) + (1 - y_i) \log(1 - \hat{y}_i)) \quad (15)$$

where, the specific meanings of \hat{y}_i and y_i are as follows:

$$\hat{y}_i = f(e_i, \tilde{e}_i, \theta_d) = \text{Sigmoid}(e_i \mathbf{W}_d \tilde{e}_i^T) \quad (16)$$

$$y_i = \begin{cases} 1, & P_i \in \mathcal{P}_{pos} \\ 0, & P_i \in \mathcal{P}_{neg} \end{cases} \quad (17)$$

where \mathcal{P}_{pos} and \mathcal{P}_{neg} denote the pool of positive and negative sample pairs, y_i is the label of the instance pair, and \hat{y}_i is the prediction score of the model for the target node (closer to 1 indicates a higher probability from the positive sample pool).

4.3. Inference early warning level phase

After the training in the previous stage, the target node in the instance pair is sufficiently similar to the subgraph. Then we use a classifier to distinguish if the target node is a risk sample. Based on the above technique, for each node v_i , SUCOLA uses the statistical-based risk estimator inference to calculate each node's risk score. The formula is defined as follows:

$$f_{\text{risk}}(v_i) = \frac{\sum_{r=1}^R \left\| \hat{y}_{i,r}^{(\text{pos})} - \hat{y}_{i,r}^{(\text{neg})} \right\|}{R} \quad (18)$$

where $f_{\text{risk}}(\cdot)$ denotes the risk score mapping function of v_i , $\hat{y}_i^{(\text{pos})}$ and $\hat{y}_i^{(\text{neg})}$ are the positive and negative pair relevance scores of prediction, respectively. Under ideal conditions, for normal nodes, $\hat{y}_i^{(\text{pos})}$ close to 1, while negative $\hat{y}_i^{(\text{neg})}$ close to 0. In contrast, the unqualified products predicted scores $\hat{y}_i^{(\text{pos})}$ and $\hat{y}_i^{(\text{neg})}$ both are around 0.5. A hyper-parameter R can efficiently prevent the error brought on by random sampling. The current food safety regulation only punishes products that are not qualified. Similarly, the qualified samples close to the detection limit should also have some risk. Due to the degree of uncertainty between the risk ratings of qualified and unqualified samples, we further classify the samples' levels of risk. As illustrated in Fig. 2, the formal representation of the risk classification process in this paper is shown in Equation (19):

$$f_{\text{level}}(v_i) = \begin{cases} 0 (\text{safe}), \min(\mathcal{V}_{\text{top}-n_0}) \geq f_{\text{risk}}(v_i) \\ 1 (\text{low risk}), \min(\mathcal{V}_{\text{top}-n_1}) > f_{\text{risk}}(v_i) > \min(\mathcal{V}_{\text{top}-n_0}) \\ 2 (\text{medium risk}), f_{\text{risk}}(v_i) \geq \min(\mathcal{V}_{\text{top}-n_1}) \\ 3 (\text{high risk}), f_{\text{risk}}(v_i) \in \mathcal{V}_{\text{unqualified}} \end{cases} \quad (19)$$

where $f_{\text{level}}(v_i)$ is the early warning level mapping function of v_i . $\mathcal{V}_{\text{top}-n_0}$ denotes the set of samples whose risk is higher than the risk of all unqualified samples, $\mathcal{V}_{\text{top}-n_1}$ represents the set of samples whose risk value is higher than the value of the prior n_1 samples (n_1 denotes the number of unqualified samples), and $\mathcal{V}_{\text{unqualified}}$ denotes the collection of unqualified samples.

Food safety issues are of great importance. All unqualified products have a risk level of 3 (high risk). These products exceed certain limit values and must be paid attention to. The risk level of qualified products with a comprehensive risk score in the top $- n_1$ range is 2 (medium risk). These products have some indicators close to the limit value, and their overall risk is comparable to most unqualified products. The risk level of qualified products with a comprehensive risk score in the top $- n_1$ to top $- n_0$ range is 1 (low risk). These products have few indicators that are close to the limit value. The risk level of qualified products with a comprehensive risk score below top $- n_0$ is 0 (low risk). These products either do not exist or pose no relevant risk.

4.4. Overall framework

Algorithm 1 summarizes the SUCOLA procedure proposed in this paper. First, the adjacency matrices of the anchor and learned graphs are obtained by the predefined and graph learning module, respectively. In the structure bootstrapping phase, to update the learned topology, we design an HSB mechanism. In the risk-based contrastive learning phase, we sample the optimized graph with instance pairs and compare the relationships between sample nodes and their local subgraphs to compute the contrast loss. In the inference early warning level phase, the risk score of each node in \mathcal{G} is repeatedly calculated R times, each time with a different local subgraph participating in the computation, which ensures that the final risk score is statistically stable. When the risk values of all nodes are calculated, the risk division of each sample is output according to the last early warning level mapping function $f_{\text{level}}(\cdot)$.

Algorithm 1 Self-adaptive structure refinement contrastive learning framework (SUCOLA)

Input: $\mathcal{G} = (\mathbf{A}, \mathbf{X})$; E_1 ; E_2 ; B ; R ; γ ; c .

Output: The risk level mapping function $f_{\text{level}}(\cdot)$.

- 1: The trainable parameters are initialized randomly θ , φ , θ_e and θ_d .
- 2: Initial state the anchor adjacency matrix definition by $\mathbf{A}_a \leftarrow \mathbf{A}_{i,j}$ via Equation (1).
- 3: // The structure bootstrapping phase.
- 4: **for** $e \in 1, 2, \dots, E_1$ **do**
- 5: Calculate the learned matrix \mathbf{S} with graph structure learner via Equations (2)–(4).
- 6: Generate two different graph views by $\mathcal{G}_a(\mathbf{X}, \mathbf{A})$, $\mathcal{G}_l(\mathbf{X}, \mathbf{S})$.
- 7: Calculate node feature vector \mathbf{H}_a , \mathbf{H}_l with encoder θ via Equations (5), (6).
- 8: Calculate representations \mathbf{Z}_l , \mathbf{Z}_a with readout φ via Equations (7), (8).
- 9: Calculate the contrastive loss \mathcal{L}_s by Equations (9)–(11).
- 10: Update parameters θ and φ by backpropagation.
- 11: **if** $e \bmod c = 0$ **then**
- 12: Fine-tuning \mathbf{A}_a with rate γ via Equation (12).
- 13: **end if**
- 14: **end for**
- 15: Update the final structure \mathbf{S}_r with \mathbf{A}_a and \mathbf{S} via Equation (13).
- 16: // The risk-based contrastive learning phase.
- 17: **for** $e \in 1, 2, \dots, E_2$ **do**
- 18: $\mathcal{B} \leftarrow$ Randomly separate \mathcal{V} into mini-batch of size B
- 19: **for** mini-batch $b = (v_1, \dots, v_B) \in \mathcal{B}$ **do**
- 20: Sample instance pairs $\mathcal{P} = \{P_{\text{pos}}, P_{\text{neg}}\}$: target nodes (v_1, \dots, v_B) , local subgraph $\mathcal{G}_1, \dots, \mathcal{G}_B$.
- 21: Calculate \mathcal{L}_r via Equations (15)–(17).
- 22: Update parameters θ_e and θ_d by backpropagation.
- 23: **end for**
- 24: **end for**
- 25: // Inference early warning level phase.
- 26: **for** $v_i \in \mathcal{V}$ **do**
- 27: Calculate the risk score mapping function $f_{\text{risk}}(v_i)$ with the number of risk estimate rounds R via Equation (18).
- 28: **end for**
- 29: Calculate the early warning level o_i with mapping function $f_{\text{level}}(v_i)$ via Equation (19).
- 30: **return** risk level $\mathcal{O} = (o_1, \dots, o_n)$

5. Experiment

We conducted comprehensive experiments on two tasks and six real-world datasets to demonstrate the effectiveness and generalizability of the suggested framework SUCOLA. The following four Research Questions will be discussed in the remaining portion of this section.

- RQ1: How effectively is SUCOLA providing risk warnings for food safety detecting data under an unsupervised setting?
- RQ2: How do the components affect the performance of SUCOLA?
- RQ3: How do key hyperparameters affect SUCOLA performance?
- RQ4: What does the graph structure look like after being learned by the SUCOLA hierarchical structure bootstrapping mechanism?

5.1. Experiment setting

5.1.1. Datasets description

In this paper, studies for assessing food risk were conducted using data on sterilized dairy products. To present this dataset comprehensively, we will elaborate on three aspects: data collection, indicator selection, and data pre-processing.

Table 2
The sample feature set.

Categories	Requirements	Inspection standard
Protein	≥ 3.1 (g/100 g)	GB 5009.5-2010
Fat	≥ 3.7 (g/100 g)	GB 5413.3-2010
Nonfat milk solids (NMS)	≥ 8.5 (g/100 g)	GB 5413.39-2010
Lactose	≤ 2.0 (g/100 g)	GB 5009.8-2016
Aflatoxin M1 (AM1)	≤ 0.5 (μg/kg)	GB 2761-2017
Acidity	11~16 (° T)	GB 5413.34-2010

Data collection. Our collaborating unit mainly completed The data collection work. The data content includes PCT data of 2158 products from a company in the Xinjiang Uygur Autonomous Region market supervision and management bureau from 2013 to 2021, with 41 unqualified samples (accounting for approximately 1.9 percent of the total product volume) included.³ Table 2 shows the specific requirements and testing methods. The sterilized milk PCT dataset follows a 3:1:6 ratio division for sequence training, validation, and detection. Table 3 introduces details of this dataset's products.

Indicator selection. We evaluated dairy products' nutritional quality and environmental pollution level using multiple indicators from the Chinese National Food Safety Standard of Sterilized Milk (GB 25190-2010).⁴ These indicators include nutritional indicators such as lactose, fat, protein, and non-fat milk solids, as well as environmental indicators such as acidity and aflatoxin M1. When the content of these nutritional or environmental indicators in food is inadequate or excessive, it may pose a risk to food safety (Zhang, 2020).

Data pre-processing. To eliminate the influence of different dimensions of risk evaluation indicators on data analysis results, we first used the min-max normalization method to transform the raw data into dimensionless data, which were then fed into downstream models for characterization. In this evaluation, the larger the value of the environmental index, the greater the risk, while the larger the value of the nutritional index, the lower the risk. The min-max normalization formalism is shown in Equations (20)–(21), after data normalization, a larger value of each indicator represents a higher level of risk.

$$x_i = \frac{x_i^{\text{env}} - \mathbf{x}^{\text{Min}}}{\mathbf{x}^{\text{Max}} - \mathbf{x}^{\text{Min}}}, i \in \mathcal{X}^{\text{env}} \quad (20)$$

$$x_j = 1 - \frac{x_j^{\text{nut}} - \mathbf{x}^{\text{Min}}}{\mathbf{x}^{\text{Max}} - \mathbf{x}^{\text{Min}}}, j \in \mathcal{X}^{\text{nut}} \quad (21)$$

Where x_i and x_j represent the dimensionless data after data normalization of sample indicators i and j , respectively. \mathcal{X}^{env} and \mathcal{X}^{nut} represent the numbers of environmental index i and nutritional index j , respectively. $\mathbf{x}^{\text{Max}} = \text{Max}\{x_1, x_2, \dots, x_n\}$, $\mathbf{x}^{\text{Min}} = \text{Min}\{x_1, x_2, \dots, x_n\}$.

Further, we verified SUCOLA on five well-known benchmark datasets for the label imbalance tasks on the attribute graph. Specifically, this work considers two social graph datasets and three citation graph datasets. Tang and Liu (2009), Sen et al. (2008). We contrasted SUCOLA with the most advanced self-supervised learning and anomaly detection techniques. For a fair comparison, we conducted experiments in the configuration of previous studies (Ding et al., 2019; Liu et al., 2021).

When injecting attribute anomalies, we specifically choose m nodes and replace their attributes with those of other randomly selected nodes. Meanwhile, we fully connect the selected p nodes as structural anomalies. We carry out this procedure q times, resulting in $m = p \times q$ nodes as the total number of structural anomalies. Table 4 lists

³ The Institute of Product Quality Supervision and Inspection, Urumqi, Xinjiang Uygur Autonomous Region, China

⁴ <http://tradechina.dairyaustralia.com.au/wp-content/uploads/2018/08/GB-25190-2010-National-Food-Safety-Standard-Sterilized-Milk-f1-.pdf>

the statistics of the datasets and the total number of outliers. It is significant to note that the anomaly injection process occurs during the “Risk-based contrastive learning phase” of SUCOLA.

5.1.2. Baseline

For the food risk warning task, our methods compared with SUCOLA include three methods with supervision GAT, SLAPS, SADAG and five models for graph anomaly detection DOMINANT, ANOMALOUS, AdONE, COLA, and ANEMONE.

1. GAT (Veličković et al., 2017) is a classical baseline model for graph tasks, which achieves adaptive adaptation of weights to different neighbors by using a self-attentive mechanism to aggregate neighbor nodes in the graph.
2. SLAPS (Fatemi et al., 2021) provides a more supervised approach to reasoning about graph structure through self-supervised adjacency and the simultaneous learning of GNN parameters.
3. SADAG (Kumagai et al., 2021) considers the nodes' attributes, graph structure, and class imbalance. This model can effectively propagate few-shot label information to unlabeled nodes.
4. DOMINANT (Ding et al., 2019) is a graph anomaly detection model in the unsupervised scenario. A graph convolution autoencoder is used to reconstruct the feature and adjacency matrix. The anomalies of each node are then noticed based on the reconstructed error.
5. ANOMALOUS (Peng et al., 2018) learns anomaly patterns by considering residual analysis and CUR decomposition. The joint learning framework chooses information attributes for the anomaly detection task.
6. AdONE (Bandyopadhyay et al., 2020) is a deep unsupervised autoencoder-based solution that generates network embeddings while minimizing the impact of abnormal nodes.
7. CoLA (Liu et al., 2021) is an anomaly detection method based on contrast self-supervised learning. It uses a contrastive learning self-supervised pattern detecting anomalous nodes by estimating the consistency between the target node and its local subgraphs.
8. ANEMONE (Jin et al., 2021) is a multi-view-based anomaly detection model for attribute graphs. It applies GNNs with multi-scale contrastive learning targets as feature encoders to capture the space feature distribution by learning the consistency between patches-level and context-level instances.

For detecting anomalies in attribute graphs, we compared SUCOLA with the seven latest baselines for handling label imbalance in unsupervised scenarios. The seven baseline models include the following three models: the ANOMALOUS, DOMINANT, CoLA, and ANEMONE models mentioned above.

1. AMEN (Perozzi and Akoglu, 2016) detects anomalies through ego network analysis. AMEN estimates attribute correlations between nodes in the ego network to distinguish anomalous nodes.
2. Radar (Li et al., 2017), same as ANOMALOUS, uses residual analysis to distinguish abnormal nodes. With regard to anomaly detection, it evaluates the coherence of attribute information residuals and graph information.
3. DGI (Veličković et al., 2019) is a typical contrastive learning-based self-supervised pattern. The model learns node vectors by maximizing the MI between the original and corrupt networks at the node level.

5.1.3. Evaluation metrics

The widely used metric ROC-AUC evaluates risk analysis performance. The ROC curve expresses that the qualified nodes are identified as qualified nodes versus the unqualified ones identified as qualified nodes. The AUC value represents the area under the ROC curve within the [0,1] range. Moreover, the AUC is positively related to performance, and the score more approach 1, the better the performance.

Table 3
Raw data from the food PCT dataset.

Sample id	Date of inspection	Inspection item					
		Lactose	Acidity	NMS	Fat	Protein	AM1
2013110544	05/11/2013	1.73	12.00	8.70	4.10	3.40	0.2
201805281284	28/05/2018	1.79	12.01	8.96	4.17	3.36	0.5
20200409469	09/04/2020	1.73	12.13	8.61	4.37	3.34	0.5
20210913761	13/09/2021	1.74	12.00	8.79	4.16	3.42	0.2

Table 4
Information about the five benchmark datasets. The last three datasets in the table are citation networks; the first two are social networks.

Dataset	Nodes	Edges	Attributes	Anomalies	Train/Val/Test
BlogCatalog	5196	171,743	8189	300	1558/520/3118
Flickr	7575	239,738	12,407	450	2272/758/4545
Cora	2708	5429	1433	150	812/271/1625
CiteSeer	3327	4732	3703	150	998/333/1996
Pubmed	19,717	44,338	500	600	5915/1972/11830

Table 5

Comparison of Milk dataset performance (i.e., AUC). Marker \diamond indicates the result of T-test p-value < 0.05 when compared with SADAG, respectively. During the training phase, the input data to the model are shown in the second row, where X and Y signify the feature representation matrix and label matrix, respectively. The best performance has been highlighted in **bold**.

Methods	GAT	SLAPS	SADAG	DOMINANT	ANOMALOUS	AdONE	CoLA	ANEMONE	SUCOLA
Available data	X, Y	X, Y	X, Y	X	X	X	X	X	
AUC	0.5086	0.5751	0.9310	0.5832	0.5126	0.7540	0.8162	0.8959	X 0.9419\diamond

5.1.4. Parameter configurations

For all datasets, we set the sampling map size K to 4. The structural trimming rate is fixed at 0.999 for all datasets. The interval c is fixed at 10. The number of layers of the GNN encoder is 2 for the GCN module. The number of sampling rounds R is fixed to 1024 for all data sets.

We set the sampling map size K to 4 for all datasets. The structural trimming rate is fixed at 0.999 for all datasets. The number of layers of the GNN encoder is 2 for the GCN module. The number of sampling rounds R is fixed to 1024 for all data sets. For each dataset, the B is set at 300. We set the learning rate to 0.1 on Milk, and the other five datasets are set to 0.01. For Milk, Citeseer, Flickr, Pubmed, and BlogCatalog, the epoch number is 400; for Cora and BlogCatalog, it is 420.

5.2. Main results (AQ1)

Risk early warning scenario. In this section, we validate SUCOLA's detection performance using eight baseline models. Among them, the AUC values are shown in Table 5, and the ROC curves are shown in Fig. 4. We can observe the following information with the results mentioned above.

- Comparing all models, SUCOLA obtained the best risk assessment results, proving its ability to analyze the risk of sterilized PCT data in the case of unsupervised patterns and complex structures. Compared with the optimal self-supervised baseline model ANEMONE, SUCOLA achieves a performance improvement of 5.13% (0.046) and even 1.1% (0.0109) over the optimal supervised model SADAG. The possible main reason is that SUCOLA optimizes the structure to obtain more accurate information.
- The supervised models (GAT, SLAPS) perform poorly. Due to the low differentiation between qualified and unqualified products of food data (each attribute is independent, high-dimensional and sparse in structure), resulting in their generators generating low-quality graph structures that limit the performance of the models. In contrast, the unsupervised approach achieves competitive results without label supervision, again demonstrating the effectiveness of structure optimization and contrastive learning.

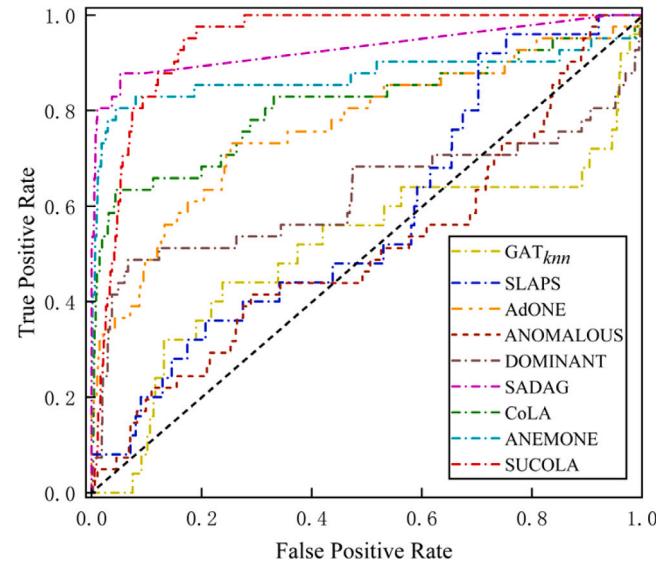


Fig. 4. ROC curves for Milk physical and chemical detecting datasets. The risk assessment performance improves with the increased area under the curve.

- SUCOLA has a better risk assessment capability compared to other deep networks. One reasonable explanation is that these baselines consider only attributes feature information propagation strategies (e.g., ANEMONE and CoLA use only view and node contrast interaction, respectively), resulting in the raw data's structural noise being ignored. In comparison, SUCOLA employs a structure-guided mechanism and a self-supervised feature vector contrastive learning strategy, reinforcing each other to obtain a higher risk assessment capability.
- The ROC curve of SUCOLA is fuller than the baseline model, indicating that our framework can accurately identify the unqualified samples from many qualified samples.

To more visually demonstrate the effect of SUCOLA on the risk classification of sterilized PCT data, we visualized the risk score calculated

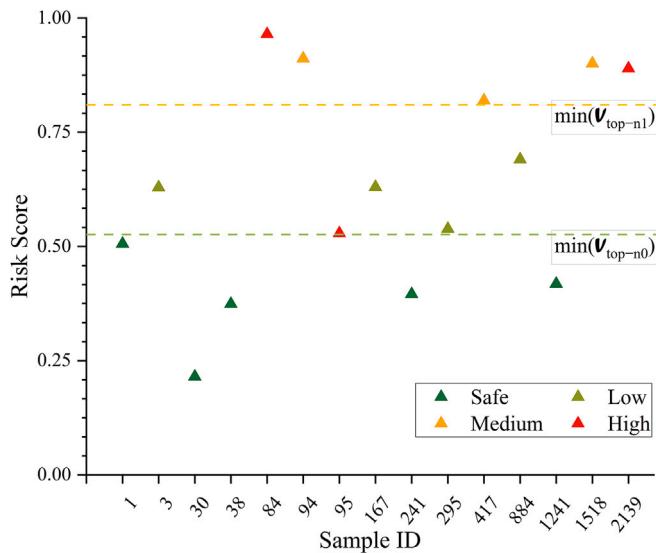


Fig. 5. Risk score distribution of sterilized PCT data. The horizontal axis represents the sample id, and the vertical axis represents the sample risk score. The higher the score, the higher the risk.

by the model based on Equation (19), as shown in Fig. 5. In summary, the threshold for setting risk levels in this case study is as follows.

1. The confusion and danger of qualified products with unqualified products are higher in the top - n1 range of comprehensive risks. Therefore, the number of unqualified products should be paid attention to.
2. Unqualified products need to be detected, so the minimum risk score (V_{top-n_1}) of unqualified samples should be paid attention to.

Graph anomaly detection scenario. We compare SUCOLA with the unsupervised anomaly detection task state-of-the-art methods. All model AUC values are shown in Table 6, and the ROC curves with the performance top three models are shown in Fig. 6. We can deduce the following information from the results mentioned above.

1. The model SUCOLA in this paper arrives at the best anomaly detection performance in all five baseline datasets. In particular, on the dataset Citeseer, SUCOLA improves 0.0468 (5.1%) over the optimal baseline model in terms of AUC, reaching 0.9657. The main reason may be that SUCOLA successfully captures more accurate relationships between nodes and their local subgraph structures using the optimized graph structure.
2. Compared to Cora, Citeseer and Pubmed, SUCOLA has a better performance advantage over the baseline on the two social network datasets. One reasonable explanation is that the average degree of the social network nodes (mean degree = 32.35) is much larger than that of the citation network nodes (mean degree = 2.51). Thus, the bootstrapped social network structure is more abundant, and the statistical-based sample estimator provides a richer representation of the local substructure of each node, enabling SUCOLA to identify the abnormal nodes more efficiently.

5.3. Ablation study (AQ2)

This section investigates the impact of different structural bootstrapping strategies and the performance of varying pooling functions for SUCOLA.

Impact of structural bootstrapping strategies. In this experiment, we investigated the choice of the SUCOLA framework based on different bootstrapping strategies. As shown in Table 7, The AUC values of our framework's three variants are as follows: the SUCOLA-SF indicates that only the fusion strategy is used as the structure bootstrapping mechanism, and the SUCOLA-SC demonstrates that only the fine-tuning strategy is used as the structure bootstrapping mechanism. The SUCOLA-HSB indicates that no structure bootstrapping module is used, and only the original structure is used as the input. We can observe that.

1. For sterilized physicochemical detecting datasets, fine-tuned and fused structure-guided strategies complement each other. The performance of SUCOLA with HSB improved by 0.0282/0.1802 over the structure bootstrapping strategy using only fusion (SUCOLA-SF)/fine tuning (SUCOLA-SC).
2. For the two social network datasets, it can be found that the fine-tuning structure strategy only works better. The strategy with HSB works better for the rest of the citation network datasets. This possible reason is that the social network nodes with high average degrees interact more frequently and carry more noise than the citation network. The original network correction is more complicated. Therefore, the representation of the social network structure using only fine-tuning is more accurate and avoids the noisy edges from the under-adjustment of the original structure.
3. The HSB strategy is not the best all the time in six datasets but the most robust option. We can achieve comparatively superior detection performance for all datasets using the HSB strategy, which indicates that the strategy has better generalization capability.

Impact of pooling strategy. In this part, we continued to explore the effect of strategies on the results of various pooling types. The empirical results are displayed in Table 8, which shows the most pronounced fluctuations in the sterilized PCT dataset compared to the other datasets. Compared with different pooling strategies, SUCOLA (average pooling) obtained the best performance on five of the six datasets and was stable overall, showing strong generalization ability. Max pooling is competitive on most datasets after average pooling, showing suboptimal results in four out of six datasets. However, using min pooling performed poorly on five of the six datasets, which indicates that using min produces a significant loss of information.

Impact of contrastive loss. We explored the effect of contrastive learning loss by comparing Info NCE loss and NT-Xent loss used in this paper, as shown in Fig. 7.

From Fig. 7, we can observe that replacing the contrastive loss with Info NCE decreases model performance across all datasets. Possible reasons for this are as follows:

1. The NT-Xent similarity function uses cosine distance for normalization, which enables the model to learn more discriminative feature representations, improving the separability of features and the model's generalization ability.
2. Info NCE loss aims to differentiate between representations of positive and negative samples (van den Oord et al., 2018), by comparing the logits of the target sample with those of the other samples (single-dimensional). By contrast, NT-Xent loss aims to maximize the consistency of representations between different views or augmented views of the same sample (Chen et al., 2020a), by comparing the embedding vectors of the target sample with those of the other samples. Therefore, the loss function used in SUCOLA can capture differences between high-dimensional representations and contains richer information for comparison.

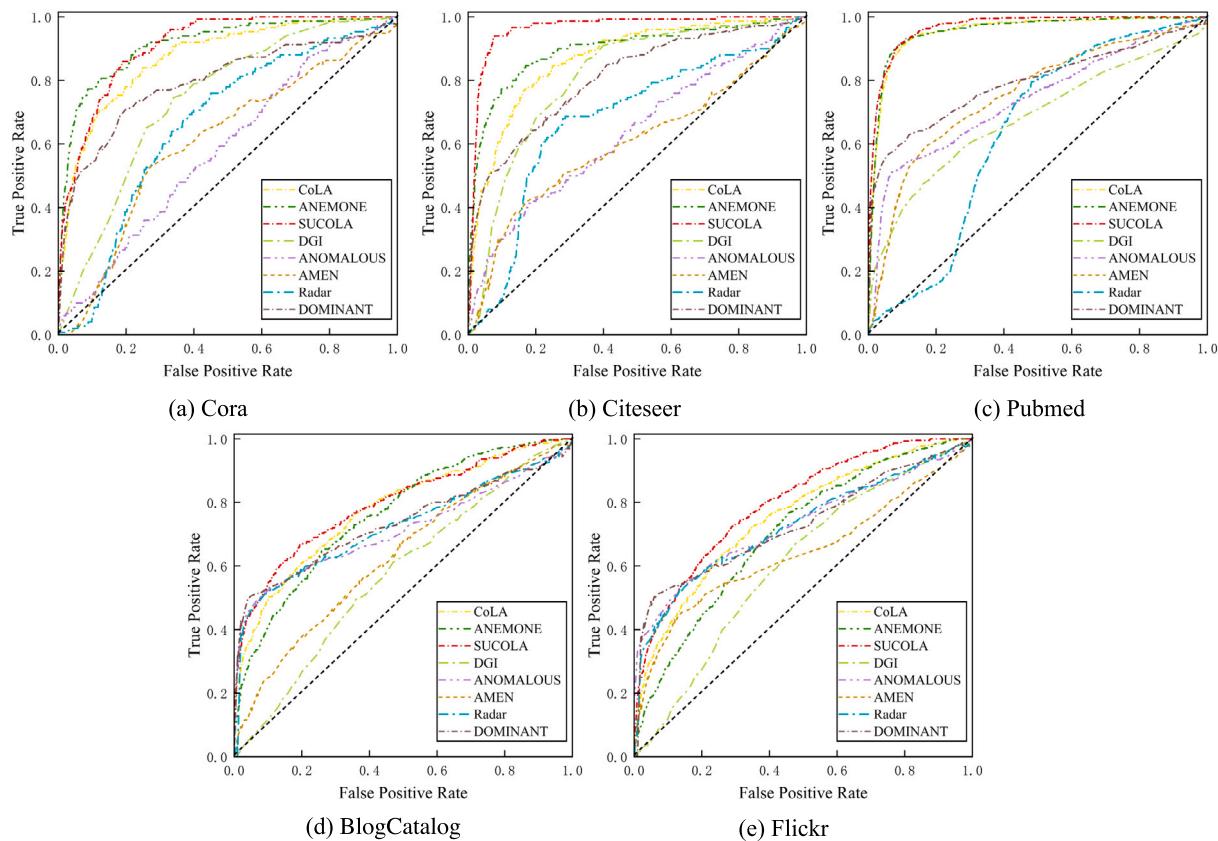


Fig. 6. ROC curves for the five benchmark datasets. The performance of the graph anomaly detection improves with the increasing area under the curve.

Table 6

Performance comparison for anomaly detection (measured by AUC) using five benchmark datasets in the unsupervised learning scenario. The results of baseline models are cited from the original paper reported. Marker # and \flat indicate the results of T-test p-value < 0.05 when compared with CoLA and ANEMONE, respectively. **Bold** values are the best results.

Methods	BlogCatalog	Flickr	Cora	Citeseer	Pubmed
AMEN	0.6392	0.6537	0.6266	0.6154	0.7713
Radar	0.7401	0.7399	0.6587	0.6709	0.6233
ANOMALOUS	0.7237	0.7434	0.5770	0.6307	0.7316
DOMINANT	0.7468	0.7442	0.8155	0.8251	0.8081
DGI	0.5827	0.6237	0.7511	0.8293	0.6962
CoLA	0.7854	0.7513	0.8779	0.8968	0.9512
ANEMONE	0.7526	0.7093	0.9057	0.9189	0.9548
SUCOLA	0.7954[#]	0.7963[#]	0.9096^b	0.9646^b	0.9682^b

Table 7

Table 7
Effect of different structure bootstrapping strategies on AUC values. Specifically, we use SUCOLA-SF, SUCOLA-SC, and SUCOLA-HSB to denote SUCOLA without the fine-tuning structure module, structural combination module, and hierarchical structure bootstrapping, respectively. The best performance has been highlighted in **bold**.

Methods	Milk	BlogCatalog	Flickr	Cora	Citeseer	Pubmed
SUCOLA	0.9419	0.7954	0.7963	0.9096	0.9646	0.9682
SUCOLA-SF	0.9137	0.7753	0.7857	0.8295	0.8617	0.9571
SUCOLA-SC	0.7617	0.8301	0.8668	0.8924	0.9600	0.9679
SUCOLA-HSB	0.8162	0.7854	0.7513	0.8779	0.8968	0.9512

Table 8

Table 3
Impact of various pooling strategies on AUC values

Pooling strategies	Milk	BlogCatalog	Flickr	Cora	Citeseer	Pubmed
Average (SUCOLA)	0.9419	0.7954	0.7963	0.9096	0.9646	0.9682
Max	0.7795	0.7919	0.7908	0.9018	0.9362	0.9575
Min	0.7627	0.7883	0.7881	0.8897	0.9407	0.9696
Weighted Average	0.7190	0.7869	0.7894	0.8999	0.9589	0.9405

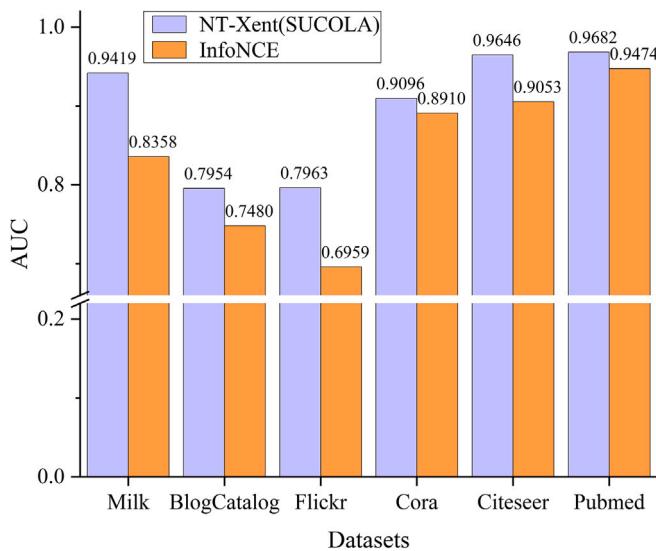


Fig. 7. Impact of different contrastive losses on model performance.

5.4. Parameter analysis (AQ3)

To answer AQ3, We conduct numerous evaluations to investigate how efficiently SUCOLA performs with various hyperparameters, including sampling rounds R , subgraph size K , and structural fine-tuning factor γ .

Sampling round R . In this part, we report the effect of the sampling round R within Equation (18). As shown in Fig. 8 (a), we adjusted R within these values of {1, 128, 256, 512, 1024, 2048, 4096}. The findings demonstrate that the AUC values positively correlate with the fetching value of R and tend to increase as R increases. When the value of R is insufficient, the performance of the AUC of each dataset is poor for effective risk warning and abnormality detection. The AUC of the sterilized physical and chemical detecting dataset is more sensitive to R . The fluctuation of AUC is greater with the increase of R . To take care of all datasets, $R = 1024$ is set in this paper to ensure reliable and efficient performance.

Subgraph size K . The impact of subgraph size K in Section 4.2.2, “Local subgraph sampling”, on experimental performance is examined in this section. As shown in Fig. 8 (b), when $K \leq 3$, the AUC values of all data sets are relatively small. The possible reason is that the subgraph contains only the target node and two neighboring nodes and lacks other structural information, which leads to poor AUC performance. When $K > 7$, the AUC performance decreases as K increases. The possible reason for this is that the rise of the subgraph leads to remote nodes entering the subgraph by sampling. However, these remote nodes generally have a low correlation with the target node and even bring noisy information. We observe that the performance of the sterilized PCT dataset is most sensitive to K , and the AUC value fluctuates most significantly as K increases. A reasonable explanation is that the structure of the sterilized dataset is inferred, and when $K \leq 3$, the nearest neighbor nodes mined by SUCOLA can effectively provide structural information, so the AUC grows significantly; when $K \geq 7$, a certain distance is reached, making the nodes newly added to the subgraph less relevant to the target nodes and carrying noise that degrades the performance. Therefore, to balance all datasets’ operational efficiency and robustness, we fix the value of K to 4.

Structural fine-tuning rate γ . In this experiment, we further report the effect of the structural fine-tuning rate γ on the performance within Equation (12). We note that as an equilibrium parameter γ controls

the impact of the learned matrix S on the A_a , the higher its value, the greater the effect of S on A_a during the structural fine-tuning process. In Fig. 9, the best point of each dataset is marked with a red pentagram. As we can see, the performance of all six datasets decreases when γ is 1. The most significant decrease is 11.94% for the Citeseer dataset. This implies that the structural fine-tuning mechanism aids in raising the level of learned graph quality. The SUCOLA performance when $\gamma = 0.999$ is not the best option on every dataset but is the most reliable. Therefore, to consider the combined efficiency on all datasets, we fix the value of γ to 0.999.

5.5. Structure visualization (AQ4)

To further understand how SUCOLA’s hierarchical structure bootstrapping mechanism works, we visualized the spatial structure, and the adjacency matrix before and after the structure guidance for 100 randomly selected samples from each dataset, where the spatial structure is shown in Figs. 10 and 11, and the adjacency matrix is shown in Figs. 12 and 13. Through the visualization, we can obtain the following information.

(1) Through intuitively capturing the spatial structure change, we find that the original graph structure of the Milk dataset is a little confusing, as shown in Fig. 10, with many connections of different communities, allowing a clearer view of the neighbors of the selected nodes. In this case, the corresponding model SUCOLA-HSB has an AUC value of only 0.8162 (86.65% of SUCOLA). In contrast, after structural bootstrapping through the hierarchy, the Milk dataset tends to converge the samples classified into the same risk level. This indicates that the structure bootstrapping module provides a high-quality association network for risk classification of downstream samples.

(2) As shown in Fig. 11, we also found a significant increase in node-to-node edges for the five benchmark datasets after the SUCOLA structural bootstrap mechanism. This indicates that the SUCOLA can learn the sample space’s potential association semantics efficiently, thus obtaining higher-quality topologies.

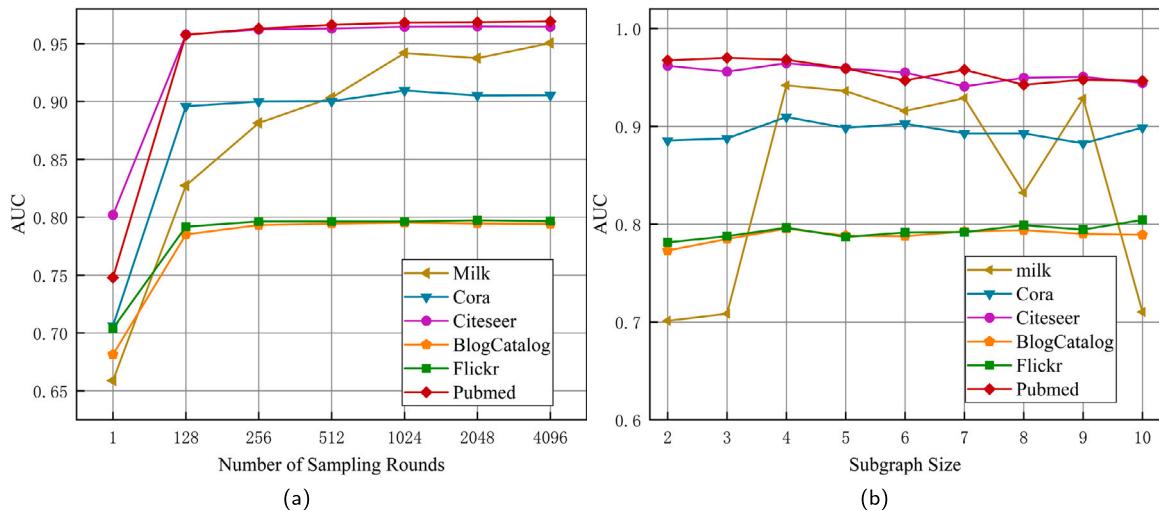
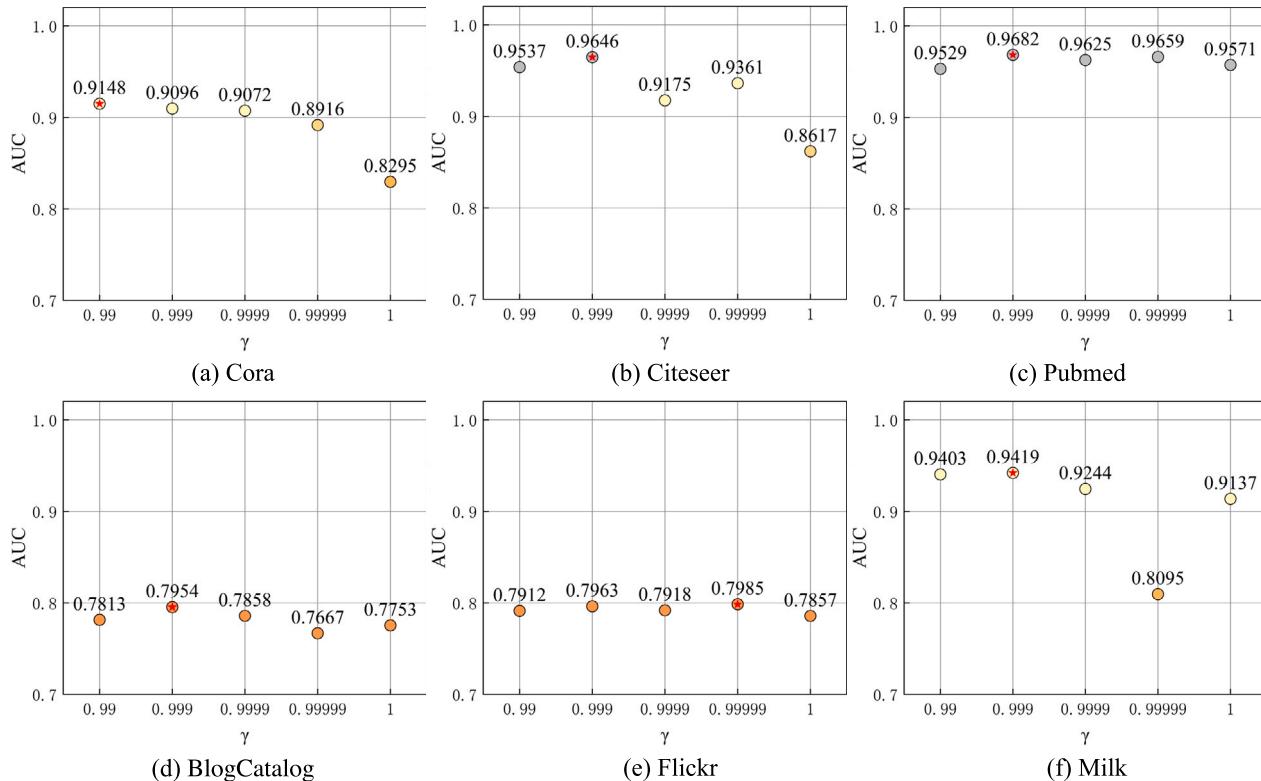
(3) In Figs. 12 and 13, by visualizing the adjacency matrix, all datasets show that the matrix after structural bootstrapping also indicates a larger weight of the edges, which leads to more attention to the essential edges, which is one reason the structural bootstrapping module can have a positive effect.

(4) Based on Figs. 10–13, compared with two dense networks (BlogCatalog and Flickr), the visualization effect of four sparse network datasets (Milk, Cora, Citeseer, and Pubmed) guided by the SUCOLA structure is more prominent. We can see that the guided sparse network structure not only enhances the information interaction of the nodes on the diagonal but also constructs many edge links between nodes of the same category, which helps these nodes receive necessary supervision signals from the contrastive objective.

Although the proposed early warning framework can enormously improve the effectiveness of current food safety monitoring and be more in line with actual needs. However, it cannot be directly employed as a guiding tool for governments to formulate food safety regulatory measures. As suggested by Zuo et al. (2022b), screened risky products need further judgment by experienced food science experts before being submitted to the authorities in charge of food regulation.

6. Conclusion and future work

To effectively carry out an early warning for products with potential food safety risks, we innovatively propose SUCOLA, a comparative learning early warning framework based on adaptive structure optimization, which distinguishes qualified and unqualified products by abnormal risk scores. Drawing on knowledge from food science, we

Fig. 8. Sensitivity of hyper-parameters R and K .Fig. 9. Effect of hyperparameter γ in each dataset.

investigate sample association networks for more accurate detection of sample risk and propose two mechanisms of structure-guided and structure-fusion to accelerate the learning of optimized structures. At the same time, we quantify the association networks of samples within real data sets. In addition, SUCOLA captures the relevant information between each product and its well-designed contextual subgraph, which can expose the network's unqualified samples and effectively help resist label imbalance. The proposed framework is used for a batch of real-world sterilized PCT data from Xinjiang province, China. The results of the exploratory trial demonstrate SUCOLA's efficacy in risk

warning and association network generation. In addition, we applied SUCOLA to similar task graph anomaly detection, and the results of five benchmark datasets demonstrate the excellent performance of SUCOLA. Our research provides a novel perspective on sample imbalance scenarios (i.e., anomaly detection and risk early warning) via structure optimization. Moreover, the proposed model can support tasks from various fields due to its outstanding generalization abilities.

In future work, we will consider more influencing factors (e.g., environmental indices, environmental quality, etc.) and a more comprehensive range of detecting samples to validate our framework further.

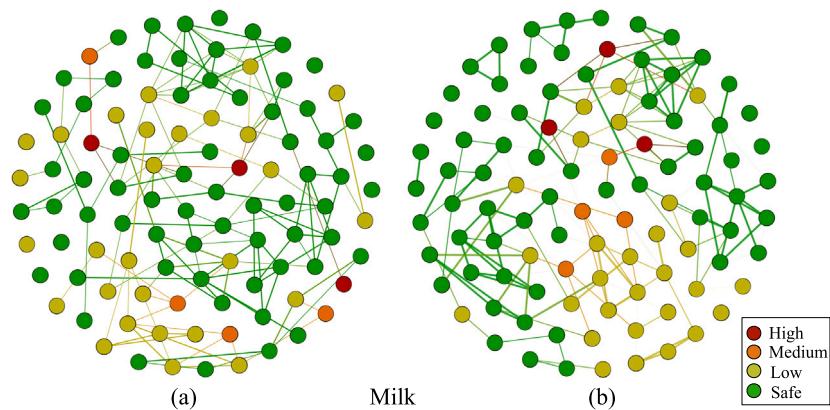


Fig. 10. Visualization of node association structure and warning level for sterilized physical and chemical detecting dataset. Where (a) is the original predefined structure and (b) is the structure after the SUCOLA bootstrapping of the model in this paper. Thicker edges indicate higher weights.

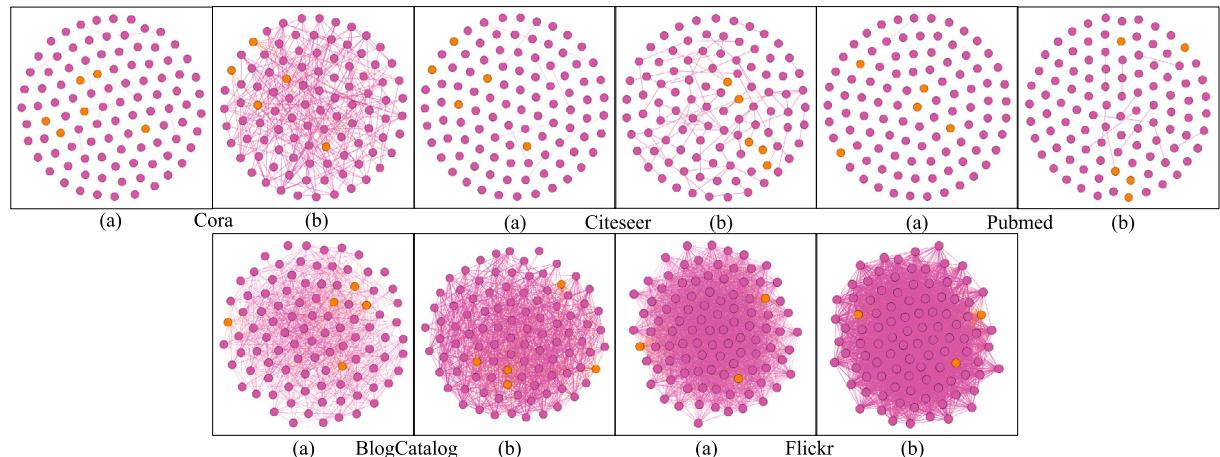


Fig. 11. Five benchmark datasets' node association structures are visualized. Where (a) is the original association structure of the dataset and (b) is the association structure after the SUCOLA bootstrapping of the model in this paper. Thicker edges indicate higher weights. Orange nodes and purple nodes indicate abnormal nodes and normal nodes, respectively.

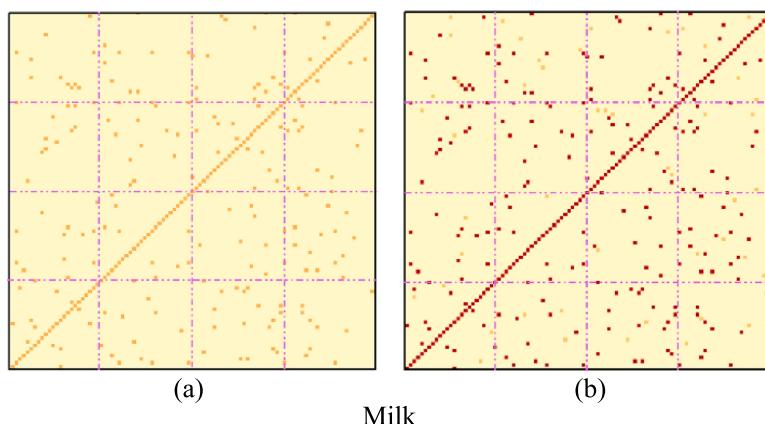


Fig. 12. Heat map of the adjacency matrix S , of the sterilized physical and chemical detecting dataset, (a) indicating the predefined adjacency matrix with self-join, and (b) indicating the adjacency matrix after the SUCOLA structure bootstrap. The darker colored blocks indicate the greater edge weights between two nodes.

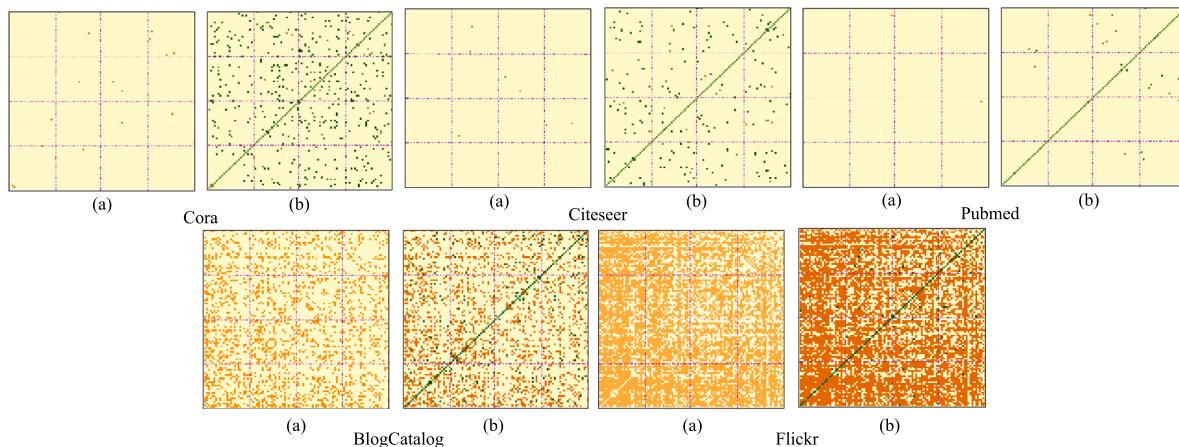


Fig. 13. Heat map of the adjacency matrix S_r of the five benchmark datasets, where (a) and (b) represent the matrix heat map visualization before and after the SUCOLA structure bootstrapping, respectively. The darker colored blocks indicate the greater edge weights between two nodes.

Exploring the differences in potential association networks obtained from different samples will be interesting.

CRediT authorship contribution statement

Enguang Zuo: Writing – original draft, Methodology. **Junyi Yan:** Data curation, Formal analysis. **Alimjan Aysa:** Visualization, Writing – review & editing. **Chen Chen:** Investigation. **Cheng Chen:** Validation. **Hongbing Ma:** Writing – review & editing, Project administration. **Xiaoyi Lv:** Supervision, Funding acquisition. **Kurban Ubul:** Funding acquisition, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

Data will be made available on request.

Acknowledgments

The authors of this paper would like to convey their sincere appreciation to the editors and anonymous reviewers.

Our work is supported by the Major science and technology projects of Xinjiang Uygur Autonomous Region (2020A03001, 2020A03001-3) and the National Natural Science Foundation of China (61862061, 61563052 and 61363064).

References

- Bandyopadhyay, Sambaran, Vivek, Saley Vishal, Murty, M.N., 2020. Outlier resistant unsupervised deep architectures for attributed network embedding. In: Proceedings of the 13th International Conference on Web Search and Data Mining. pp. 25–33.
- Bouzemra, Yamine, Marvin, Hans J.P., 2016. Prediction of food fraud type using data from rapid alert system for food and feed (RASFF) and Bayesian network modelling. *Food Control* 61, 180–187.
- Brashears, Matthew E., 2014. Exponential random graph models for social networks: Theory, methods, and applications. *Contemp. Sociol. J. Rev.* 43 (4), 552–553.
- Capurro, D., Cole, K., Echavarria, M.I., Joe, J., Neogi, T., Turner, A.M., 2014. The use of social networking sites for public health practice and research: a systematic review. *J. Med. Internet Res.* 16 (3), e79.
- Chen, Ting, Kornblith, Simon, Norouzi, Mohammad, Hinton, Geoffrey, 2020a. A simple framework for contrastive learning of visual representations. In: International Conference on Machine Learning. PMLR, pp. 1597–1607.
- Chen, Yu, Wu, Lingfei, Zaki, Mohammed J., 2020b. Iterative deep graph learning for graph neural networks: Better and robust node embeddings. *Adv. Neural Inf. Process. Syst.* 2020-Decem.
- Ding, Kaize, Li, Jundong, Bhanushali, Rohit, Liu, Huan, 2019. Deep anomaly detection on attributed networks. In: SIAM International Conference on Data Mining. SDM 2019, pp. 594–602.
- Duan, Jingcan, Wang, Siwei, Zhang, Pei, Zhu, En, Hu, Jingtao, Jin, Hu, Liu, Yue, Dong, Zhibin, 2022. Graph anomaly detection via multi-scale contrastive learning networks with augmented view. arXiv preprint arXiv:2212.00535.
- Egilmez, Hilmi E., Pavel, Eduardo, Ortega, Antonio, 2017. Graph learning from data under Laplacian and structural constraints. *IEEE J. Sel. Top. Sign. Proces.* 11 (6), 825–841.
- Fatemi, Bahare, El Asri, Layla, Kazemi, Seyed Mehran, 2021. SLAPS: Self-supervision improves structure learning for graph neural networks. *Adv. Neural Inf. Process. Syst.* 27, 22667–22681.
- Franceschi, Luca, Niepert, Mathias, Pontil, Massimiliano, He, Xiao, 2019. Learning discrete structures for graph neural networks. In: 36th International Conference on Machine Learning, Vol. 2019-June. ICML 2019, pp. 3481–3493, arXiv:1903.11960.
- Geng, Zhi Qiang, Zhao, Shan Shan, Tao, Guang Can, Han, Yong Ming, 2017. Early warning modeling and analysis based on analytic hierarchy process integrated extreme learning machine (AHP-ELM): Application to food safety. *Food Control* 78, 33–42.
- Halcrow, Jonathan, Mosoi, Alexandru, Ruth, Sam, Perozzi, Bryan, 2020. Grale: Designing networks for graph learning. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 2523–2532.
- Ji, Yugang, Shi, Chuan, Wang, Xiao, 2021. Prohibited item detection on heterogeneous risk graphs. In: International Conference on Information and Knowledge Management. Proceedings. pp. 3867–3877.
- Jiao, Yizhu, Xiong, Yun, Zhang, Jiawei, Zhang, Yao, Zhang, Tianqi, Zhu, Yangyong, 2020. Sub-graph contrast for scalable self-supervised graph representation learning. In: 2020 IEEE International Conference on Data Mining. ICDM, IEEE, pp. 222–231.
- Jin, Ming, Liu, Yixin, Zheng, Yu, Chi, Lianhua, Li, Yuan-Fang, Pan, Shirui, 2021. Anemone: Graph anomaly detection with multi-scale contrastive learning. In: Proceedings of the 30th ACM International Conference on Information & Knowledge Management. pp. 3122–3126.
- Khanna, Sunil K., 2009. Food and culture: A reader (2nd ed.) , by carole counihanand penny van esterik. *Ecol. Food Nutr.* 48 (2), 157–159.
- Kipf, Thomas N., Welling, Max, 2017. Semi-supervised classification with graph convolutional networks. In: 5th International Conference on Learning Representations, ICLR 2017 - Conference Track Proceedings.
- Kumagai, Atsutoshi, Iwata, Tomoharu, Fujiwara, Yasuhiro, 2021. Semi-supervised anomaly detection on attributed graphs. In: 2021 International Joint Conference on Neural Networks. IJCNN, IEEE, pp. 1–8.
- Li, Jundong, Dani, Harsh, Hu, Xia, Liu, Huan, 2017. Radar: Residual analysis for anomaly detection in attributed networks. In: IJCAI International Joint Conference on Artificial Intelligence. pp. 2152–2158.
- Lin, Xiaoyong, Li, Jiatong, Han, Yongming, Geng, Zhiqiang, Cui, Shiyong, Chu, Chong, 2021. Dynamic risk assessment of food safety based on an improved hidden Markov model integrating cuckoo search algorithm: A sterilized milk study. *J. Food Process Eng.* 44 (3).
- Liu, Yixin, Li, Zhao, Pan, Shirui, Gong, Chen, Zhou, Chuan, Karypis, George, 2021. Anomaly detection on attributed networks via contrastive self-supervised learning. *IEEE Trans. Neural Netw. Learn. Syst.* 1–15.
- Liu, Yixin, Zheng, Yu, Zhang, Daokun, Chen, Hongxu, Peng, Hao, Pan, Shirui, 2022. Towards unsupervised deep graph structure learning. In: WWW 2022 - Proceedings of the ACM Web Conference 2022. pp. 1392–1403.

- Min, Weiqing, Jiang, Shuqiang, Liu, Linhu, Rui, Yong, Jain, Ramesh, 2019. A survey on food computing. *ACM Comput. Surv.* 52 (5).
- Ofli, Ferda, Aytar, Yusuf, Weber, Ingmar, Al Hammouri, Raggi, Torralba, Antonio, 2017. Is saki #delicious? The food perception gap on instagram and its relation to health. In: 26th International World Wide Web Conference, WWW 2017. pp. 509–518.
- Peng, Zhen, Luo, Minnan, Li, Jundong, Liu, Huan, Zheng, Qinghua, 2018. Anomalous: A joint modeling approach for anomaly detection on attributed networks. In: IJCAI International Joint Conference on Artificial Intelligence, Vol. 2018-July. pp. 3513–3519.
- Peng, Zhen, Luo, Minnan, Li, Jundong, Xue, Luguo, Zheng, Qinghua, 2020. A deep multi-view framework for anomaly detection on attributed networks. *IEEE Trans. Knowl. Data Eng.* 34 (6), 2539–2552.
- Perozzi, Bryan, Akoglu, Leman, 2016. Scalable anomaly ranking of attributed neighborhoods. In: 16th SIAM International Conference on Data Mining 2016. SDM 2016, pp. 207–215.
- Sadilek, Adam, Kautz, Henry, DiPrete, Lauren, Labus, Brian, Portman, Eric, Teitel, Jack, Silenzio, Vincent, 2016. Deploying nEmesis: Preventing foodborne illness by data mining social media. In: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 30, No. 2. pp. 3982–3989.
- Samuel, Oluwarotimi Williams, Asogbon, Grace Mojisola, Sangaiah, Arun Kumar, Fang, Peng, Li, Guanglin, 2017. An integrated decision support system based on ANN and Fuzzy_AHP for heart failure risk prediction. *Expert Syst. Appl.* 68, 163–172.
- Sen, Prithviraj, Namata, Galileo Mark, Bilgic, Mustafa, Getoor, Lise, Gallagher, Brian, Eliassi-Rad, Tina, 2008. Collective classification in network data. *AI Mag.* 29 (3), 93–106.
- Silva, Thiago H., Vaz De Melo, Pedro O.S., Almeida, Jussara, Musolesi, Mirco, Loureiro, Antonio, 2014. You are what you eat (and drink): Identifying cultural boundaries by analyzing food and drink habits in foursquare. In: Proceedings of the 8th International Conference on Weblogs and Social Media, ICWSM 2014. pp. 466–475.
- Tang, Jian, Chen, Zhixiang, Fu, Ada Wai Chee, Cheung, David W., 2002. Enhancing effectiveness of Outlier detections for low Density Patterns. In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), In: Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 2336.535–548,
- Tang, Lei, Liu, Huan, 2009. Relational learning via latent social dimensions. In: Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. pp. 817–825.
- Tong, Hanghang, Faloutsos, Christos, Pan, Jia Yu, 2006. Fast random walk with restart and its applications. In: Proceedings - IEEE International Conference on Data Mining. ICDM, pp. 613–622.
- van den Oord, Aaron, Li, Yazhe, Vinyals, Oriol, 2018. Representation learning with contrastive predictive coding.
- Veličković, Petar, Cucurull, Guillem, Casanova, Arantxa, Romero, Adriana, Lio, Pietro, Bengio, Yoshua, 2017. Graph attention networks. *arXiv preprint arXiv:1710.10903*.
- Velickovic, Petar, Fedus, William, Hamilton, William L., Liò, Pietro, Bengio, Yoshua, Hjelm, R. Devon, 2019. Deep graph infomax.. In: ICLR (Poster), Vol. 2, No. 3. p. 4.
- Wang, Yue, Sun, Yongbin, Liu, Ziwei, Sarma, Sanjay E., Bronstein, Michael M., Solomon, Justin M., 2019. Dynamic graph cnn for learning on point clouds. *ACM Trans. Graph.* 38 (5).
- Xia, Jun, Wu, Lirong, Chen, Jintao, Hu, Bozhen, Li, Stan Z, 2022. Simgrace: A simple framework for graph contrastive learning without data augmentation. In: Proceedings of the ACM Web Conference 2022. pp. 1070–1079.
- Xu, Jiehui, Wu, Haixu, Wang, Jianmin, Long, Mingsheng, 2021. Anomaly Transformer: Time Series Anomaly Detection with Association Discrepancy.
- You, Yuning, Chen, Tianlong, Sui, Yongduo, Chen, Ting, Wang, Zhangyang, Shen, Yang, 2020. Graph contrastive learning with augmentations. *Adv. Neural Inf. Process. Syst.* 33, 5812–5823.
- Zhang, Yu, 2020. Food safety risk intelligence early warning based on support vector machine. *J. Intell. Fuzzy Systems* 38 (6), 6957–6969.
- Zhang, Jiaqiang, Wang, Senzhang, Chen, Songcan, 2022. Reconstruction enhanced multi-view contrastive learning for anomaly detection on attributed networks. *arXiv preprint arXiv:2205.04816*.
- Zhao, Tianxiang, Zhang, Xiang, Wang, Suhang, 2021. GraphSMOTE: Imbalanced node classification on graphs with graph neural networks. In: WSDM 2021 - Proceedings of the 14th ACM International Conference on Web Search and Data Mining. pp. 833–841.
- ZhiQiang, Geng, LingLing, Liang, YongMing, Han, GuangCan, Tao, Chong, Chu, 2021. Risk early warning of food safety using novel long short-term memory neural network integrating sum product based analytic hierarchy process. *Br. Food J.* 124 (3), 898–914.
- Zhu, Yanqiao, Xu, Weizhi, Zhang, Jinghao, Liu, Qiang, Wu, Shu, Wang, Liang, 2021. Deep graph structure learning for robust representations: A survey, arxiv, 1: 8.
- Zuo, Enguang, Aysa, Alimjan, Muhammat, Mahpirat, Zhao, Yuxia, Chen, Bing, Ubul, Kurban, 2022a. A food safety prescreening method with domain-specific information using online reviews. *J. fur Verbraucherschutz und Lebensmittelsicherheit* 17 (2), 163–175.
- Zuo, Enguang, Du, Xusheng, Aysa, Alimjan, Lv, Xiaoyi, Muhammat, Mahpirat, Zhao, Yuxia, Ubul, Kurban, 2022b. Anomaly score-based risk early warning system for rapidly controlling food safety risk. *Foods* 11 (14).