Rare Category Analysis for Complex Data: A Review

DAWEI ZHOU, Computer Science, Virginia Tech,
JINGRUI HE, Information Science, University of Illinois at Urbana-Champaign,

Despite the sheer volume of data being collected, it is often the rare categories that are of the most important in many high impact domains, ranging from financial fraud detection in online transaction networks to emerging trend detection in social networks, from spam image detection in social media to rare disease diagnosis in the medical decision support system. The unique challenges of rare category analysis include: (1) the highly skewed class distribution; (2) the non-separability nature of the rare categories from the majority classes; (3) the data and task heterogeneity; (4) the time-evolving property of the input data sources. This survey aims to provide a concise review of the state-of-the-art techniques on complex rare category analysis, where the majority classes have a smooth distribution while the minority classes exhibit the compactness property in the feature space or subspace. Rare category analysis aims to identify, characterize, represent and interpret the anomalies that not only show statistical significance but also exhibit interesting patterns (e.g., compactness, high-order structures, shown in a burst, etc.). More specifically, we start with the introduction, problem definition, and unique challenges of complex rare category analysis, then present a comprehensive review of recent advances that are designed for this problem setting, from rare category exploration without any label information to the exposition step that characterizes rare examples with a compact representation, from representing rare patterns in a salient embedding space to interpreting the prediction results and providing relevant clues for the end-users' interpretation; finally, we discuss the potential problems and shed light on the future directions of complex rare category analysis.

Additional Key Words and Phrases: Rare Category Analysis, Imbalanced Data Distribution, Anomaly Detection

1 INTRODUCTION

In the era of big data, one of the main characteristics is the sheer volume of multi-modality data being collected from various platforms (e.g., graphs, images, text, etc.). In contrast, it is often the rare occurrences that are of great importance in many high-impact domains, ranging from financial fraud detection in online transaction networks to emerging trend detection in social networks, from spam image detection in social media to rare disease diagnosis in the medical decision support system. Anomaly detection refers to the problem of discovering these rare occurrences that are significantly different from other observations in the data. Extensive techniques have been proposed to characterize anomalies in various domains. Examples include financial fraud detection [104], insider threat prediction [66], novelty detection [94], gene disease discovery [17], emerging trend tracking [28], network intrusion detection [58], astronomy [135], spam image detection [95], rare disease diagnosis [119], etc.

1.1 Anomalies vs. Rare Category Examples

The branch of data mining concerned with identifying anomalies or outliers[1] has a longstanding history. Tracing back to 1980, Douglas M. Hawkins first proposed the definition of outliers [47] in Def. 1.1. Following Hawkins'

[1]In this paper, we use 'anomalies' and 'outliers' interchangeably.

Authors' addresses: Dawei Zhou, zhoud@vt.edu, Computer Science, Virginia Tech, 620 Drillfield Dr., Blacksburg, VA, 24060,; Jingrui He, jingrui@illinois.edu, Information Science, University of Illinois at Urbana-Champaign, 501 E. Daniel St, Champaign, IL, 61820-6211,

0360-0300/2023/10-ART $15.00
https://doi.org/10.1145/3626520

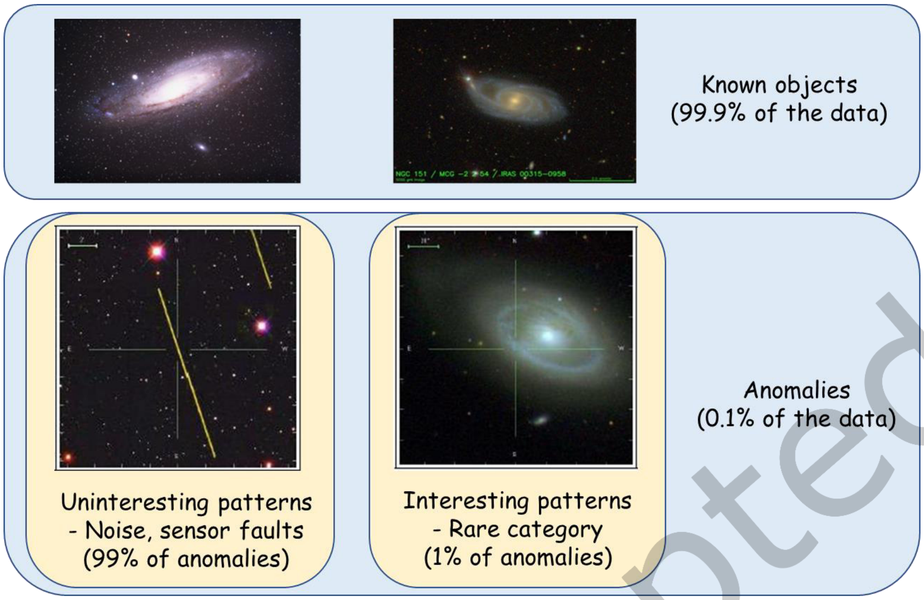


Fig. 1. Anomalies in Sloan data [111]. (top row) known objects. (bottom row) anomalies, where 99% of the anomalies are uninteresting patterns, such as diffraction spikes (bottom left), and only 1% are interesting patterns that are worthy of future research and may lead to the discovery of extraordinary objects (bottom right).

definition of outliers, the problem of anomaly detection or outlier detection has been generalized and studied in various contexts, such as high-dimensional numerical data [175], sequential data [27], time-series data [45], graph data [7], financial data [36, 115], and thus results in many domain-specific names for outliers and anomalies, such as novelty, event, surprising change, fraud, outbreak, etc.

Definition 1.1. Hawkins' Definition of Outliers [47]
An outlier is an observation that differs so much from other observations as to arouse suspicion that it was generated by a different mechanism.

Despite the tremendous success of anomaly detection methods in various domains, it is common sense that not all anomalies are necessarily useful or relevant to the actual events of interest. In fact, most anomalies are uninteresting data points drawn from the known distribution of noise or correspond to the combinations of features that are less valuable to the downstream applications [111]. For instance, in Figure 1, we present a set of sky images captured by ground-based telescopes in the program of Sloan Digital Sky Survey (SDSS)[2]. According to the scientific discovery from SDSS, 99.9% of the captured sky images (e.g., the top row of Figure 1) by SDSS can be well explained based on the known phenomena of the universe (e.g., discovered stars, comets, nebulae, etc.), and only 0.1% of the images (e.g., the bottom row of Figure 1) are anomalies. Moreover, within the anomalies, 99% of the images (e.g., bottom left image in Figure 1) are of no interest to astronomers and are caused by the diffraction spikes of satellite trails or the artifacts of the telescope. Only 1% of the instances (a minuscule 0.001% of the whole SDSS database) are useful and may lead to new scientific discoveries (e.g., the bottom right image in Figure 1). Here, we refer to the anomalies that are not only statistically significant but also interesting

[2]http://www.sdss.org/iotw/archive.html

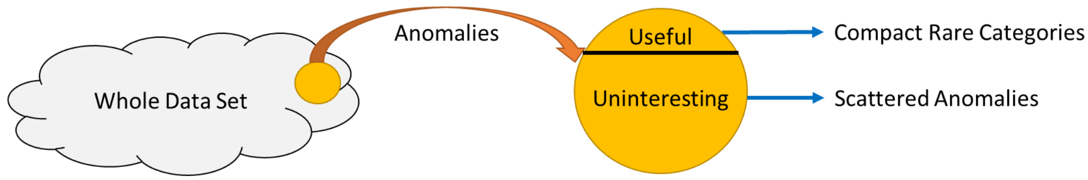


Fig. 2. Relationship between anomalies/outliers and rare category examples.

as the rare category examples. And the problem of studying the rare category examples is referred to as rare category analysis (RCA). Figure 2 illustrates the relationship between anomalies, uninteresting anomalies, and useful anomalies (i.e., rare category examples), where rare categories are subsets of anomalies that are not only statistically significant but also compact in the feature space.

Here we provide a general definition for the rare category analysis as follows. Given a dataset $\mathcal{D}$ of n examples that come from C distinct classes. Without loss of generality, we assume that $\sum_{i=1}^{n} \mathbf{x}_i = \overrightarrow{0}$ and $\frac{1}{n}\sum_{i=1}^{n}(x_i^j)^2 = 1$, where x_i^j denotes the j^{th} feature of $\mathbf{x}_i$. Besides, for the sake of simplicity, we assume the class label for the i^{th} node $y_i = 0$ corresponds to the majority class (i.e., normal examples), and $y_i \in \{1, \ldots, C-1\}$ corresponds to the rare category (i.e., suspicious examples). Note that rare category analysis is the problem of studying the patterns from the minority classes[3]. In particular, for the cases where more than one majority class is observed, we treat all the majority classes as a single class once the underlying distribution of each majority class is sufficiently smooth [49]. For the general rare category analysis, we make the following assumptions regarding the support region of the majority class and minority classes.

Assumption 1. ***Smoothness Assumption for Majority Class.*** *Given a highly imbalanced dataset, the distribution of the support region of each majority class is sufficiently smooth.*

Assumption 2. ***Compactness Assumption for Minority Class.*** *Given a highly skewed dataset, the minority class examples can be represented as a compact cluster in the feature space.*

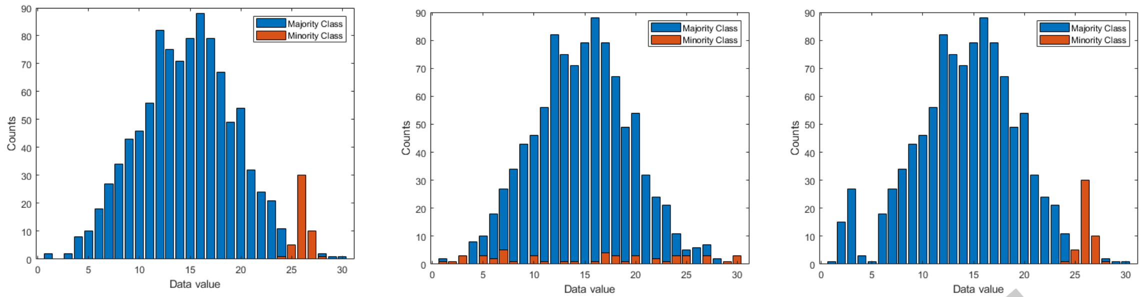
These assumptions are made for the purpose that the rare categories are identifiable and meaningful. To be more clear, let us first look at the example in Figure 3 (a), where the majority class (colored in blue) has a Gaussian distribution with large variance on the left, while the minority class (colored in orange) corresponds to a peak with small variance on the right. If the minority class is not compact and violates Assumption 2 (e.g., the minority class in Figure 3 (b) is uniformly distributed in the feature space), then no algorithm can perform better than random sampling. If the distribution of the majority class is not smooth and violates Assumption 1 (e.g., the majority class in Figure 3 (c) consists of multiple narrow and sharp peaks just as the minority class), then the minority class cannot be identified with a clear clue. Based on the assumptions concerned with rare category examples, we propose the definition of complex rare category analysis as follows.

Definition 1.2. Complex Rare Category Analysis
Rare category analysis refers to the problem of detecting, characterizing, representing, and interpreting rare examples from underrepresented minority classes in a highly imbalanced dataset.

1.2 Challenges

We first discuss the key challenges associated with rare category analysis. Different from the conventional anomalies, the rare category examples have the following unique characteristics.

[3]In this paper, we use 'rare category' and 'minority class' interchangeably.



(a) Both Assumption 1 & 2 hold (b) Only Assumption 1 holds (c) Only Assumption 2 holds

Fig. 3. The support regions of a majority class and a minority class in a one-dimensional synthetic dataset.

- **C1. Rarity.** It is often the case that the patterns of our interest are sporadic in the collected data. For example, in financial fraud detection, the vast majority of the financial transactions are legitimate, and only a small number may be related to money laundering; furthermore, money laundering activities often exhibit similar traits. Another example is network intrusion detection. New malicious network activities are hidden among huge volumes of routine network traffic, and network intrusions of the same type are often similar to one another.
- **C2. Label Scarcity**: Due to the high skewness and non-separable nature of rare categories, labeling rare category examples is extremely expensive. In various real applications, such as money laundering detection, synthetic ID detection, and network intrusion detection, we are usually required to learn from a handful of labeled examples.
- **C3. Non-separability.** The rare categories are often non-separable from the majority classes in the feature space. For example, from the alarm of grid hacking (e.g., NotPetya) to the swift growth in cyber-criminals (e.g., coin mining attacks), the past several years provide us a serious reminder of an emerging type of terrorist attack (i.e., *adversarial attacks*). Such malicious attacks are even harmful as they are designed to bypass the current detection systems by camouflaging themselves as the normal instances, i.e., the support regions of normal instances, and malicious attacks are non-separable in the given feature space.
- **C4. Data Heterogeneity.** To identify the rare categories, we often need to collect raw data from different sources, at different time stamps, using different techniques, etc. Each data source might have different types of users and attributes. Of such highly heterogeneous raw data, only a subset (e.g., data sources, attributes) may be relevant to the identification of the rare category examples (e.g., security threats).
- **C5. Dynamics.** The raw input data constantly change over time; thus, extracting and modeling useful information from such dynamic systems plays an important role in rare category analysis.

1.3 An Overview of Complex Rare Category Analysis

Complex rare category analysis can be divided into four tasks: detecting, characterizing, representing, and interpreting rare category examples. Figure 4 presents an overview of the rare category analysis, from the exploration step without any ground truth to the exposition step that aims to characterize rare examples with a compact representation, from representing rare patterns in a salient space to providing insightful interpretation over the prediction results.

- **T1: Rare Category Exploration:** In this task, we start from *de-novo* and do not have any label information of the data. The goal is to identify at least one example from each rare category with the minimum help

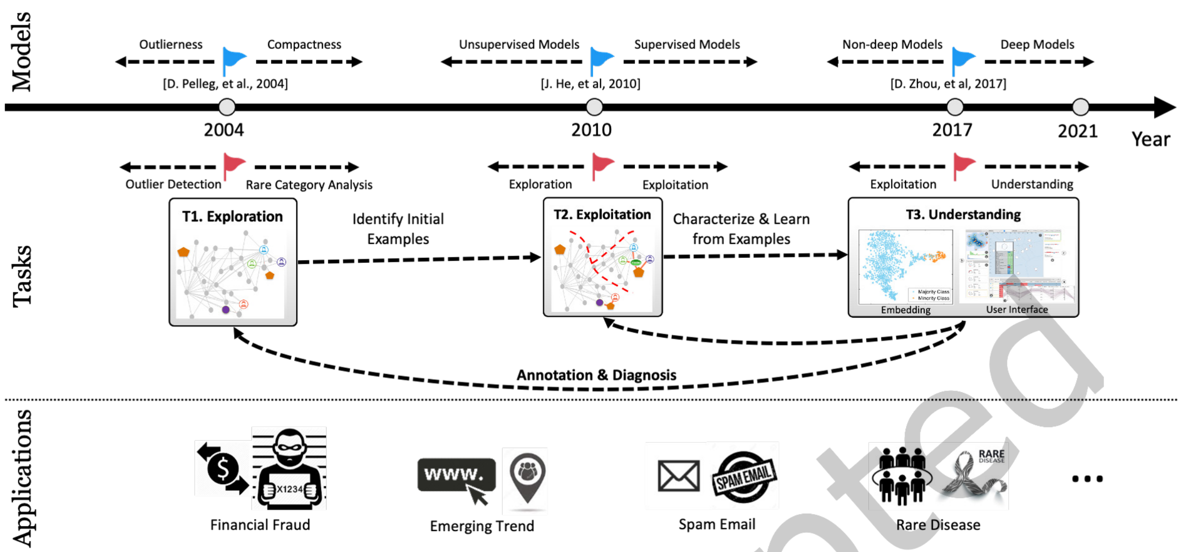


Fig. 4. Overview of complex rare category analysis.

from labeling oracles. In particular, we begin with a comprehensive review of some relevant topics, such as anomaly detection [7, 26], outlier detection [57], predictive modeling on imbalanced domains [20]; then, we elaborate on the popular approaches for rare category exploration in both static and dynamic environments.

- **T2: Rare Category Exploitation:** In this task, we have one/a few labeled examples from each class, which could be obtained from the exploratory step discussed in T1. The primary goal is to capture a compact representation of the rare categories in various data formats by learning from a few-shot annotated data points. Depending on the availability of multi-modal representation of examples, the rare category detection can be categorized into homogeneous rare category exploration and heterogeneous rare category exploration. In the former setting [52, 111], all the examples are homogeneous, as they are collected from the same data source or have identical traits; while in the latter one [41, 79, 162, 165], we aim to identify rare categories in the presence of data heterogeneity.
- **T3: Understanding Rare Categories:** This task serves as an investigation step for the end users, such as representing rare examples in a salient embedding space, visualizing the data distribution, interpreting the underlying prediction process in the previous steps (T1 and T2). We will systematically discuss the recent advances in representation learning [60, 168], interpretation [30, 86, 91], and visual analytic systems [82, 107] for rare category analysis.

1.4 Previous Work and Our Contributions

The past decade has observed extensive literature surveys on the topic of imbalanced classification, anomaly/outlier detection, and fraud detection. In the context of imbalanced classification, [48] covers the development of research in learning from imbalanced data. In the context of outlier/ anomaly detection, [57] introduces a literature review of contemporary techniques for outlier detection, [26] presents a comprehensive overview of the research on anomaly detection. Moreover, a series of surveys and special issue journal articles target more specific problems or applications in the context of outlier/ anomaly detection. In particular, [175] studies the problem of unsupervised outlier detection in high-dimensional numerical data, [27] deals with anomalies in the discrete sequences, [45] focuses on outlier detection for temporal data, [7] deals with graph-based anomaly detection algorithms, [25] provides an overview of deep-learning-based anomaly detection techniques, and [36, 115] survey the methodologies of fraud detection.

However, very few works aim to present a comprehensive overview of the complex rare category analysis, where the anomalies form a compact representation. [49] is one of the first to present an end-to-end investigation of rare categories, while the introduced methodologies are largely restricted to the tabular data in the static setting. Therefore, in this survey, we aim to provide a comprehensive and structured overview of complex rare category analysis from several perspectives:

- **Data Perspective:** Different from the previous surveys on outlier detection, anomaly detection, or rare category analysis, we review the state-of-the-art state techniques on rich data types, such as texts/blogs, Electronic Health Records (EHR), weblink graphs, stream data, etc.
- **Task Perspective:** We provide a comprehensive overview of the end-to-end complex rare category analysis, from rare category exploration without any label information to the exposition step that characterizes rare examples with a compact representation, from representing rare patterns in a salient embedding space to interpreting the prediction results and providing relevant clues for the end-users' investigation.

1.5 Overview and Organization

We present our survey in five major sections. In Section 2, we focus on the *de-novo* step of rare category analysis, a.k.a. rare category exploration, which aims to find the initial example from each minority class in the unsupervised setting. In section 3, we discuss the following step - rare category exploitation that relies on the output of rare category exploration (i.e., initial labeled examples) to characterize the rare categories. In Section 4, we stress the importance of interpretation in rare category analysis in various highly regulated industries, such as finance, healthcare, etc. We systematically discuss how to understand rare category analysis from the following two aspects: (1) data diagnosis (i.e., *how is the data distributed? which piece of information is more valuable than the others for a given task?*) (2) model diagnosis (i.e., *why does the model make a certain prediction on a particular piece of information?*). At last, Section 5 enumerates the popular benchmark datasets and representative baseline models before we conclude the existing works and share our thoughts regarding the future directions in Section 6, such as how to ensure the performance in the presence of noise, missing data, adversarial examples, and how to generate task-specific rare category examples (e.g., money laundering activity) given a specific domain (e.g., transaction network)? A general outline is given as follows.

- **Section 2: Rare Category Exploration**
  - 2.1 Rare category exploration for static data
    - 2.1.1 Tabular data
    - 2.1.2 Graph-structured data
  - 2.2 Rare category exploration for dynamic data
    - 2.2.1 Time-series data
    - 2.2.2 Temporal graphs
- **Section 3: Rare Category Exploitation**
  - 3.1 Homogeneous rare category exploitation
    - 3.1.1 Global approaches
    - 3.1.2 Local approaches
  - 3.2 Heterogeneous rare category exploitation
    - 3.2.1 Data heterogeneity
    - 3.2.2 Task heterogeneity

2 RARE CATEGORY EXPLORATION

Rare category exploration, a.k.a. rare category detection, is referred to as the *de-novo* step in rare category analysis, which aims to propose initial rare category examples to the labeling oracle in the given dataset. Given an unlabeled, imbalanced dataset, the most natural way of rare category exploration is random sampling. Specifically, we randomly draw examples without replacement to be labeled by the oracle until at least one example from each minority class has been identified. However, due to the rarity of rare categories (e.g., the proportion of the rare category is 0.01%), the random sampling will require an extensive number (i.e., 10,000 times) of labeling requests to spot the rare example. Therefore, it is worthy developing effective and efficient sampling strategies to address the rare category exploration problem. In this section, we formally define the problem of rare category exploration as follows, which primarily aims to address the (C1) rarity, (C4) data heterogeneity, and (C5) dynamics in RCA.

Problem 1. ***Rare Category Exploration.***
Given: (i) an unlabeled dataset $\mathcal{D}$, (ii) a small budget $\mathcal{B}$ for querying domain experts.
Find: the set of initial examples $\mathcal{Q}$ from each rare category.

Depending on the real-world applications, researchers have customized the Problem 1 for different data formats, including tabular data, time-series data, and graph-structured data. In the following three subsections, we review the existing literature of rare category exploration from a data perspective (static data v.s dynamic data, tabular data v.s graph-structure data, etc.).

2.1 Rare category exploration for static patterns

In the static setting, existing techniques are primarily designed in an active learning scheme. The general procedure of this paradigm is presented in Figure 5. The key idea behind the sample-selection-based methods is to learn a detector that can gradually search for the potential rare examples by learning from the examples labeled by the oracle. In general, we often assume that the labels obtained from the oracles are under a fixed

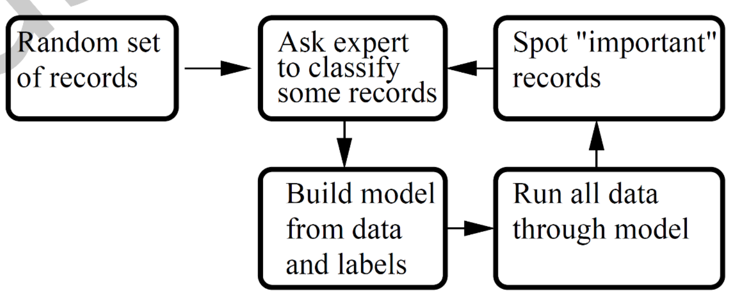


Fig. 5. A Conceptual frameworks of rare category exploration [111].

budget $\mathcal{B}$. The goal is to maximally improve the prediction accuracy with the minimum number of labeling requests to the oracle. As shown in Figure 5, rare category exploration proceeds iteratively. In each iteration, the program starts with the oracle annotating a few samples. Then the sample-selection-based methods are updated based on the newly labeled examples and further predict a small number of potential rare category examples in the sense that obtaining labels for them would further improve the prediction accuracy. The program terminates once all the target rare categories are captured. Next, we elaborate on the existing rare category exploration algorithms designed for tabular and graph-structured data.

2.1.1 Tabular data. Given an unlabeled dataset with n samples $\mathcal{D} = \{\mathbf{x}_1, \ldots, \mathbf{x}_n\}$, and each sample $\mathbf{x}_i$, $i = 1, \ldots, n$, comes with d-dimensional features $\mathbf{x}_i = \{x_i^1, \ldots, x_i^d\}$, our goal is to identify at least one example from each class $y = 1, \ldots, m$ with minimum queries. [111] is one of the first attempts at rare category exploration, which has developed a mixture model to fit the data and designed a family of hint selection methods to select the rare examples with the help of a human expert. Experimental results with different hint selection methods show the efficacy of the proposed rare category detection framework. [52] further studies the rare category exploration problem when the minority classes are non-separable from the majority classes. Specifically, the authors develop a nearest-neighbor-based rare category detection algorithm named *NNDM*, which gradually selects examples with the maximum changes in the local density on a certain scale and asks for the labeling from the oracle. The algorithm works as follows. Given the prior p of each rare category C, the algorithm first estimates the number of examples $K = n \times p$ in C. Then, the algorithm represents each example $\mathbf{x}_i$ in terms of a vector that is composed of the distances from its K nearest neighbors. By obtaining the minimum distance r among all samples, the algorithm constructs a hyper-ball with radius r for each example (as center) and computes the local density as the number of enclosed examples within the hyper-ball as follows.

$$\texttt{Hyberball}(\mathbf{x}_i, r) = \{\mathbf{x} | \mathbf{x} \in \mathcal{D}, \|\mathbf{x} - \mathbf{x}_i\| \le r\} \tag{1}$$

$$\texttt{LocalDensity}(\mathbf{x}_i) = |\texttt{Hyberball}(\mathbf{x}_i, r)| \tag{2}$$

To measure the degree of local density changes around each sample (i.e., hyper-ball), the authors propose to utilize the difference of local density between $\mathbf{x}_i$ and its neighbors.

$$\texttt{DensityChange}(\mathbf{x}_i) = \begin{cases} \max_{\mathbf{x}_k \in \texttt{Hyberball}(\mathbf{x}_i, r)} [\texttt{LocalDensity}(\mathbf{x}_i) - \texttt{LocalDensity}(\mathbf{x}_k)] & \mathbf{x}_i \text{ is labeled} \\ -\infty & \text{Otherwise} \end{cases} \tag{3}$$

In particular, for all the labeled examples, the function `DensityChange` assigns the score with $-\infty$ to avoid duplicated labeling requests. At last, the algorithm will return the examples with the largest local density changes to the oracle for labeling.

$$\texttt{Query} = \texttt{argmax}_{\mathbf{x}_i \in \mathcal{D}} \texttt{DensityChange}(\mathbf{x}_i) \tag{4}$$

Moreover, theoretical analysis shows that the methods will effectively select examples on both the boundary and in the interior of the rare categories. In particular, when the rare categories are compact, the majority class distribution is locally smooth, and the sample space is sufficiently large, NNDB is guaranteed to identify at least one example coming from the minority classes with probability $1 - \delta$ after $\lceil \frac{2\alpha}{r_2} \rceil$ iterations, where $\delta \in (0, 1)$, α, and r_2 are all positive parameters.

Despite the promising results of *NNDM* with theoretical guarantees, the performance largely relies on the prior information (e.g., the number of classes, the proportion of minority classes). To alleviate the restriction on prior knowledge, [50] proposes a prior-free rare category detection algorithm named *SEDER*. The intuition lies in the observation that the region with an abrupt local density change has a higher probability of observing the rare classes. Different from [52, 53], *SEDER* picks the potential rare examples with large neighborhood density changes for labeling, by performing semi-parametric density estimation. *SEDER* introduces a prior-free `DensityChange` function, which is defined as the norm of the gradient of the local density function in the following equation.

$$\texttt{DensityChange}(\mathbf{x}_i) = \sqrt{\sum_{l=1}^{d} \frac{(\sum_{k=1}^{n} \texttt{DensityEstimation}(\mathbf{x}_i)(x_i^l - b^l x_k^l))^2}{((\sigma^l)^2 b^l)^2}} \tag{5}$$

where x_i^l denotes the l^{th} feature of $\mathbf{x}_i$, σ^l denotes the bandwidth of the l^{th} feature, `DensityEstimation` is a kernel density estimation function with Gaussian kernel [37], and b^l is a positive parameter. In the presence of noisy data and irrelevant features, [51] formulates the rare category exploration problem as a co-selection scheme, which recovers the relevant features and the representative examples from the rare categories. To obtain the optimal sets of relevant features and rare examples, the authors propose an effective searching procedure (i.e., *PALM*) based on augmented Lagrangian to solve the optimization problem. In particular, *PALM* is designed in an alternative fashion to find the relevant features and the minority class examples.

More recently, a paradigm shift has been observed in the data mining community, from how to accurately detect rare categories to how to ensure trustworthiness in rare category exploration. To name a few, [4, 67, 85] studies the problem of how to ensure the robustness in detecting rare events, especially in the out-of-distribution setting. [122] point out that algorithmic fairness and rare category exploration are dual problems when encountering the data imbalance issue. The authors propose a fairness-aware method that jointly optimizes prediction accuracy and statistical parity. In [32], the authors study the hardness bias in the autoencoder-based rare category detection methods and propose a plug-and-play error calibration method to mitigate the harness bias issue.

2.1.2 Graph-structured data. In many fields, graphs offer a unifying data structure for modeling relational data. As a result, extensive research on rare category exploration has been conducted to spot the rare category entities on graph-structured data. Formally speaking, given an unlabeled graph $\mathcal{D} = \mathcal{G}(\mathcal{V}, \mathcal{E})$, where $\mathcal{V}$ and $\mathcal{E}$ denote the sets of nodes and edges in $\mathcal{G}$, our target is to identify initial nodes/edges from each rare categories. [53] extends the idea of [52] to the graph-structured data by proposing a graph-based rare category detection algorithm named *GRADE*. The algorithm starts computing a global similarity matrix $\mathbf{S}$ motivated from manifold ranking [171], which is used to get a compact representation for the examples from the minority classes.

$$\mathbf{S} = (\mathbf{I} - \alpha \mathbf{D}^{-\frac{1}{2}} \mathbf{A} \mathbf{D}^{-\frac{1}{2}})^{-1} \tag{6}$$

where $\mathbf{I}$ denotes the identify matrix, $\mathbf{D}$ denotes the diagonal matrix, and $\mathbf{A}$ denotes the adjacency matrix of $\mathcal{G}$. Then, a prior-guided k-nearest-neighbor matrix is computed to capture the sharp local density changes near the boundary of minority classes and thus make it easier to capture the rare patterns. On top of *GRADE*, the authors develop a variation named *GRADE-LI*, which only requires an upper bound on the proportion of each rare category. *GRADE-LI* can work with the data when detailed class-membership distribution about the data is not available to the users.

Except for the plain graph, data often exhibit node-level and edge-level heterogeneity for various critical tasks in security, finance, medicine, and so on. Each node and edge is associated with a specific type in such data (referred to as a heterogeneous graph). To accommodate the heterogeneous graphs, a collection of work has been done to detect anomalies and rare events in the unsupervised setting. For example, [129] proposes the notion of neighborhood formation for a bipartite graph, which computes the relevance scores of all nodes to a query node v and defines the neighborhood of v as the set of nodes with higher relevance scores. Based on the neighborhood formation, the authors developed an anomaly detection algorithm to spot the abnormal nodes with low "normality" scores. Based on similar intuition, [134] propose a non-negative residual matrix factorization framework named *NrMF*, which aims to detect the malicious group of entities as well as provide interpretation of prediction results for data analysts. In particular, *NrMF* is built upon the conventional matrix factorization mechanism as follows.

$$\mathbf{A} = \mathbf{F}\mathbf{G} + \mathbf{R} \tag{7}$$

where **F** and **G** are low-rank factorized matrices, and **R** is the residual matrix that indicates rare examples on the bipartite graphs. To interpret the predictions via *NrMF*, the authors impose a non-negative constraint on the residual matrix **R**, such that all entries in **R** are larger than 0. Intuitively, the higher value of the entry in **R** indicates a higher probability of existing rare and abnormal examples. To solve the optimization problem of non-negative residual matrix factorization, the authors develop a fast optimization algorithm to incrementally compute the residual matrix via the rank-1 approximation, which runs in linear w.r.t. the size of the graph.

More recently, extensive attempts have been made to predict rare category events by developing customized graph neural networks (GNNs) to depict the structure of a graph and learn representations of various graph signals (e.g., nodes, edges, subgraphs). In general, the GNNs-based approaches are designed based on the message-passing scheme, which is given as follows:

$$\mathbf{h}_v^{(l+1)} = \mathtt{UPDATE}^{(l)}(\mathbf{h}_v^{(l)}, \mathtt{AGGREGATE}^{(l)}(\mathbf{h}_u^{(l)}, \forall u \in \mathcal{N}(v))) \tag{8}$$

$$= \mathtt{UPDATE}^{(l)}(\mathbf{h}_v^{(l)}, \mathbf{m}_{\mathcal{N}(v)}^{(l)}) \tag{9}$$

where $\mathbf{h}_v^{(l+1)}$ denotes the embedding of node $v \in \mathcal{V}$ at the $(l+1)^{\text{th}}$-layer, $\mathtt{UPDATE}^{(l)}$ and $\mathtt{AGGREGATE}^{(l)}$ are both differentiable functions for the l^{th}-layer, and the $\mathbf{m}_{\mathcal{N}(v)}^{(l)}$ is the aggregated 'message' from node v's neighborhood $\mathcal{N}(v)$. To name a few, [34] proposes a GCN-based framework for predicting emerging events by capturing contextual information from the raw data. The proposed framework first extracts graph representations of the events documents then learns to predict the occurrence of future events and identify the events of interest (i.e., rare category); [164] studies the problem of rare category detection in videos, where each video is decomposed into *n* snippets (i.e., nodes in the constructed video graph), and the edges between each pair of snippets are defined based on the feature similarity, and the rare events are represented as abnormal actions. To solve this problem, a graph convolutional network (GCN) is built to simultaneously clean noisy signals in the constructed video graph and predict abnormal actions. In [42], the authors investigate the relationship between the local outlier methods[69] and message passing framework, which motivates them to develop a graph neural network-based method to identify local rare examples.

2.2 Rare category exploration for dynamic patterns

The existing literature for detecting rare patterns on temporal data is very rich and can be categorized in many ways. As a consequence, there is not a universal abstract categorization that can fully cover the developed techniques in this direction. Here, we only focus on rare category exploration for time-series data and temporal graphs.

2.2.1 Time-series data. In the setting of time-series data, the rare categories can be observed at both the sequence level and the segment level, which is shown in Figure 6. For example, most of the ECG signals are normal (i.e., collected from healthy people), while only a small portion of them is abnormal (i.e., collected from patients with heart disease). Moreover, within the abnormal ECG signals, only a few segments are abnormal (e.g., heart arrhythmia), while the rest are normal. Early studies of rare category exploration for time-series data [18, 21, 22, 65, 81, 99, 102, 110, 120, 152] are closely related to the outlier detection and disorder detection methods. They largely rely on the distanced-based mechanisms [99, 103, 118, 121] that define various similarity measurements [45] of temporal sequences/segments and then identify rare patterns deviating from the normal ones. For instance, in [22], the authors propose a scalable distance-based detection algorithm for high-volume data streams, which has been demonstrated to be optimal in terms of the CPU costs; in [18], the authors study the problem of discovering rare time-series motif (i.e., repeated subsequences) from unbounded streams. To address the rarity issue of the time-series motif in a never-ending stream, the authors develop a "sticky cash" algorithm that adopts a Bloom filter to remember every incoming subsequence and efficiently detects rare motifs in the unbounded

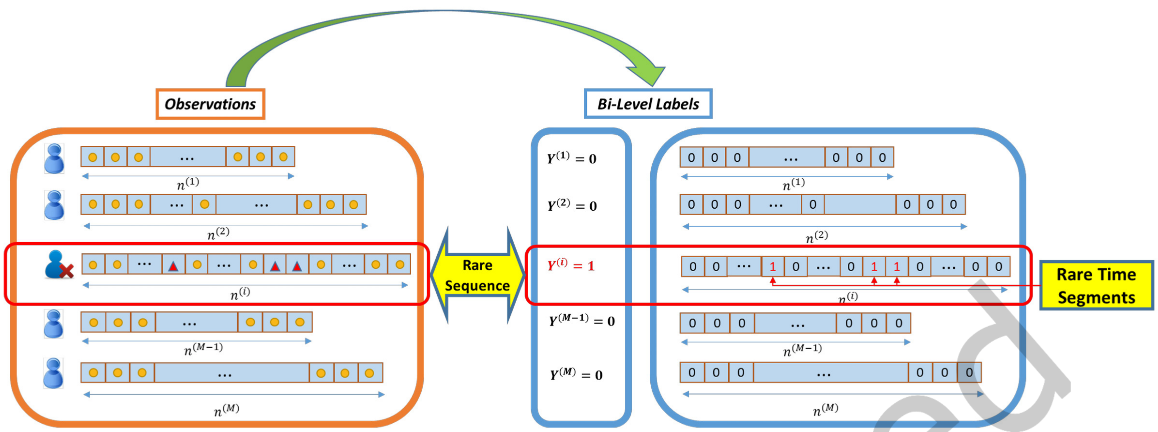


Fig. 6. An illustration of the bi-level rare categories (i.e., rare sequences and rare time segments) in time-series database [166].

real-valued time series. Moreover, to facilitate the computation of the distance-based methods, [152] introduces a fast algorithm for time-series/subsequence all-pairs-similarity-search, which shows strong implications and promising results for the task of time-series discord discovery; [43] introduces a robust random cut data structure to produce a sketch or synopsis of time-series data. With that, the authors propose a scalable anomaly detection algorithm by gradually updating the time-series sketch in a continuous data stream.

A key motivation of the above methods is that the distributions of rare categories (minorities) are deviating from the normal distribution (majorities). However, there are some obvious caveats to this idea in practice. The identified examples are often not the targets of our interest, which are drawn from noise or combinations of irrelevant features. In [166], the authors assume that the targets of our interest are rare and drawn from a compact distribution. Under such an assumption, the authors propose a bi-level generative model that aims to jointly characterize the rare temporal pattern at the time-series and subsequence levels. More recently, with the advances in deep learning architectures, a surge of research interest has been observed to develop deep models to characterize and detect rare temporal patterns in high-impact applications. In [136], the authors present a deep unsupervised framework for detecting insider threat in the online data streams, which outputs a ranked list of anomaly scores of individual user behaviors; in [77], the authors develop a Generative Adversarial Network (GAN) for unsupervised multivariate anomaly detection. Unlike conventional distance-based methods and supervised methods, the proposed framework detects rare temporal patterns by using the GAN-trained generator and discriminator to compute the Discrimination and Reconstruction Anomaly Score (DR-Score); in [123], the authors propose a recurrent network ensemble called Recurrent Autoencoder, which is designed to characterize and capture the abnormal time series segments at multiple resolutions; in [33, 161], the authors propose to model the between-sensor dependence relationship for identifying the abnormal time segments in multivariate time series.

2.2.2 Temporal graphs. Many real-world systems are intrinsically dynamic and can be represented as temporal graphs, such as social networks, communication networks, gene interaction networks, etc. In the past few years, researchers have proposed several rare category exploration models for temporal graphs [3, 13, 34, 59, 76, 91, 92, 117, 127, 142, 151, 155, 169, 170]. Depending on the way of collection data in different application domains, the existing work can be summarized as discrete temporal graphs [156] and continuous temporal graphs [87, 105, 173].

The discrete temporal graph is often referred to as time-evolving graphs, where the data $\mathcal{D}$ is presented as a sequence of snapshots $\tilde{\mathcal{G}} = \{\mathcal{G}^{(1)}, \mathcal{G}^{(2)}, \ldots, \mathcal{G}^{(T)}\}$. Each snapshot $\mathcal{G}^{(t)} = (\mathcal{V}^{(t)}, \mathcal{E}^{(t)})$, $t = 1, 2, \ldots, T$ is composed of a collection of nodes $\mathcal{V}^{(t)}$ and edges $\mathcal{E}^{(t)}$ at timestamp t. To identify rare examples on $\tilde{\mathcal{G}}$, it is natural to extend the static methods to the dynamic setting. For example, [128] proposes a parameter-free model that can monitor grouped outliers and their changes in a stream of graphs. The algorithm is designed based on Minimum Description Length (MDL), which allows the users to discover the changes in both communities as well as the points in time; [14] develops a fast incremental tensor analysis approach, which can discover both transient and periodic/repeating communities in dynamic graphs; [127] defines a commute-time distance that captures the node relationship changes and allows traditional distance-based methods to be performed on discrete temporal graphs; [76] proposes a discrete-time exponential-family random graph model to identify clusters on time-evolving graphs; [92] studies the problem anomaly detection in the streaming heterogeneous graphs by proposing a clustering-based anomaly detection approach that can simultaneously address the heterogeneity and streaming nature of the input data. In particular, the authors introduce a novel embedding mechanism that can encode the heterogeneous streaming graph into a vector representation, which will be used to perform clustering and identify anomalous patterns. [13] proposes a factorization framework that can jointly model the distribution of dynamic connections and attributes and track the evolution of evolving communities. Despite the success, the detection algorithms often suffer from the expensive computational cost, especially when extensive snapshots are given or in the online setting. To address this issue, in [169, 170], the authors propose an incremental rare category exploration scheme, which aims to gradually update the static rare category exploration models based on the local changes on a new snapshot without learning from scratch. In particular, the authors propose a closed-form solution to update the global similarity matrix $\mathbf{S}$ (defined in Eq. 6) in the dynamic setting as follows.

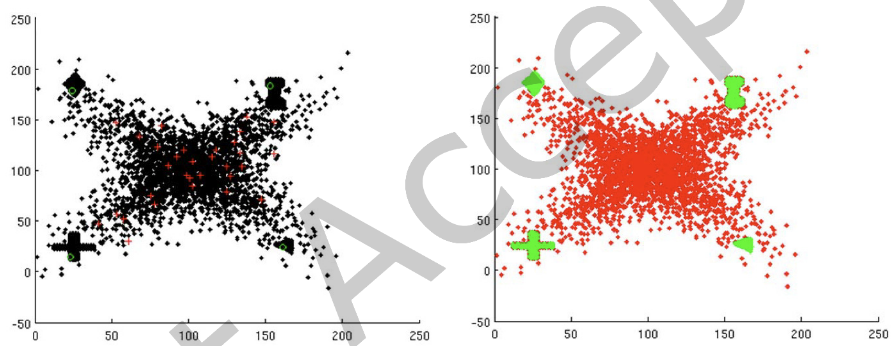
$$\mathrm{S}^{(t)} = \mathrm{S}^{(t-1)} + \alpha \frac{\mathrm{S}^{(t-1)} \mathbf{u} \mathbf{v}^T \mathrm{S}^{(t-1)}}{\mathbf{I} + \mathbf{v}^T \mathrm{S}^{(t-1)} \mathbf{u}} \tag{10}$$

where $\mathbf{u}$ and $\mathbf{v}$ are the one-hot vectors indicating the source nodes and target nodes of the updated edges, $\mathbf{I}$ denotes the identify matrix, $\mathrm{S}^{(t)}$ and $\mathrm{S}^{(t-1)}$ are the global similarity matrix at the timestamp t and $(t-1)$, respectively. Moreover, to efficiently and effectively compute the density changes $\texttt{DensityChange}(\mathbf{x}_i)$, a theoretical condition is provided to determine whether the hyper-ball $\texttt{Hyberball}(\mathbf{x}_i, r)$ are required to update from the previous timestamp. [151] studies the problem of anomaly detection in dynamic social networks, where both network structure and node attributes are observed over time. The proposed framework jointly models two processes, i.e., (1) normal modeling component and (2) anomaly detection component, to track the abnormal relationship between nodes' features and link generation in dynamic social networks; [3] introduces a novel community scoring metric named permanence and proposes an incremental algorithm to track the evolution of network communities in the dynamic setting. The theoretical analysis shows the updating procedure of the proposed algorithms leads to permanence maximization in the dynamic networks. Furthermore, to guarantee the model efficiency and algorithmic fairness, [39] introduce a fairness-aware clique-preserving spectral clustering algorithm that generalizes the static clique clustering methods to the dynamic setting via fairness-aware edge filtering and incremental eigenpair updating.

Continuous temporal graphs are also named fine-grained temporal graphs or temporal interaction graphs, where the temporal graph is presented as a sequence of timestamped edges. Different from the discrete temporal graphs, it is intractable to directly generalize the static rare category exploration approaches to the continuous temporal graphs. For this reason, in [92], the authors, for the first time, propose to represent continuous temporal graphs with a vector representation, which is easy to compute and preserves the context information of the continuous temporal graphs. With the learned continuous temporal graph representation, the authors further developed a fast and memory-efficient detection algorithm to process any incoming nodes and edges and identify anomalies in real-time. Later on, [59] studies the problem of identifying grouped anomalies in the edge streaming setting. In particular, the data is presented as a sequence of streaming edges. The authors propose a streaming algorithm with the theoretical justification that performs graph clustering with only three integers per node and does not keep any edge in memory; [10] proposes a block-structured time series model for detecting communities on time-evolving graphs, which can capture both the link persistence and community persistence over time.

3 RARE CATEGORY EXPLOITATION

In the previous section, we have introduced the existing literature for rare category exploration, which aims to identify at least one example from each class. As a follow-up step, rare category exploitation aims to learn from the identified examples and capture all the rare examples in the dataset. In Figure 7, we present an illustrative example of rare category exploitation on a synthetic dataset with one majority class and four minority classes. The key observations are as follows: 1) the distribution of the majority class is smooth, and 2) the support regions of the four minority classes are compact. That is, instances in each minority class are self-similar and form a compact representation. In the past decades, how to explore the compactness of rare categories has attracted considerable interest from the data mining and machine learning community. In this section, we will dive into the problem of rare category exploitation under two scenarios: 1) homogeneous rare category exploitation and 2) heterogeneous rare category exploitation. In the former setting, the data are collected from the same source and represented in the same format. While, in the latter setting, we aim to capture rare examples in the presence of data and task heterogeneity.



(a) Initial labels. (b) Classification results.

Fig. 7. An illustrative example of rare category exploitation with one majority class and four minority classes [54]. The labeled minority class examples are colored in green, the labeled majority class examples are colored in red, and the unlabeled examples are colored in black.

3.1 Homogeneous Rare Category Exploitation

Formally, we let $\mathcal{X}$ and $\mathcal{Y}$ denote the sample space and label space. We are given a dataset $\mathcal{D} = \{\mathcal{D}_l, \mathcal{D}_u\}$, where $\mathcal{D}_l = \{(\mathbf{x}_i)\}_{i=1}^{N_l}$ consists of N_l obtained annotated samples, $\mathcal{D}_u = \{(\mathbf{x}_j)\}_{j=1}^{N_u}$ denotes the N_u unlabeled examples, and $\mathbf{x}_i, \mathbf{x}_j \in \mathcal{X}$, $y_i \in \mathcal{Y}$. Due to the rarity (C1) and label scarcity (C2) challenges, it is often the case that the annotated data N_l is scarce. In rare category exploitation, our goal is to produce a prediction function $f(\cdot) : \mathcal{X} \rightarrow \mathcal{Y}$ by learning from $\mathcal{D}_l$ and accurately characterize unknown rare category examples in $\mathcal{D}_u$. With that, here we present the problem definition of rare category exploitation in the homogeneous setting as follows. In this problem setting, we mainly study the (C1) rarity, (C2) label scarcity, (C3) non-separability, and (C5) dynamics in RCA.

Problem 2. ***Homogeneous Rare Category Exploitation.***
Given: a dataset $\mathcal{D} = \{\mathcal{D}_l, \mathcal{D}_u\}$ with scarce labeled examples $\mathcal{D}_l = \{(\mathbf{x}_i)\}_{i=1}^{N_l}$ and extensive unlabeled examples $\mathcal{D}_u = \{(\mathbf{x}_j)\}_{j=1}^{N_u}$.
Find: (i) a prediction function $f(\cdot)$ to characterize rare category examples (2) a set of unlabeled examples which are likely coming from the rare categories.

We organize the existing approaches for homogeneous rare category exploitation into two categories: global approaches and local approaches. The first category aims to capture the global data distribution for characterizing rare categories, while the second category focuses on exploring rare categories within one or a few local regions (e.g., the neighborhood of the identified rare category examples).

3.1.1 Global approaches. This category of approaches exploits both the labeled and unlabeled examples to learn the class-membership distribution. Essentially, these methods transform the rare category exploitation problem into a classification problem with high-skewed data distribution. However, unlike the well-studied imbalanced classification problem that aims to maximize the overall classification accuracy (majority classes and minority classes), rare category exploitation emphasizes learning and characterizing the minority classes. The early rare category exploitation work addresses the rarity (C1) and non-separability (C3) challenges. One representative approach is the hyper-ball-based rare category exploitation [54, 55]. The key idea of these approaches is to formulate the rare category exploitation as an optimization problem of minimizing the radius R of the hyper-ball that well encloses the rare categories. The general objective function for the binary case (i.e., one majority class and one minority class) can be written as follows.

$$\begin{aligned} \min_{R,\mathbf{c}} \quad & R^2 \\ \text{s.t.,} & \|\mathbf{x}_i - \mathbf{c}\|^2 \le R^2, \mathbf{x}_i \in \mathcal{D}_l^{\text{maj}} \\ & \|\mathbf{x}_j - \mathbf{c}\|^2 \ge R^2, \mathbf{x}_j \in \mathcal{D}_l^{\text{min}} \\ & \|\mathbf{x}_k - \mathbf{c}\|^2 \ge R^2, \mathbf{x}_j \in \mathcal{D}_u \end{aligned} \tag{11}$$

where R denotes the radius of the hyper-ball, c denotes the center of the hyper-ball, $\mathcal{D}_l^{\text{maj}}$ denotes the set of labeled majority class examples, and $\mathcal{D}_l^{\text{min}}$ denotes the set of labeled minority class examples. Moreover, to ensure the compactness of the hyper-ball, the authors introduce three constraints towards the optimization problem: 1) the labeled examples from the majority classes should be outside of the hyper-ball; 2) the labeled examples from the minority class should be inside of the hyper-ball; 3) the hyper-ball should enclose as many unlabeled examples as possible. To tackle the non-separability challenges, RACH [54] provides a relaxed solution of Eq. 11 by introducing slack variables to allow the miss-classified instances and then converting the problem into a convex optimization problem. The relaxed problem can be easily solved in its dual form via the projected subgradient method. Following the idea of [54], [55] introduces a kernelized rare category exploitation algorithm, which generalizes RACH to model complex shapes of the support region of the target rare categories by projecting to the high-dimensional feature space induced by kernels. The kernel method provides more flexibility to characterize the complex rare categories in real applications.

More recently, a series of works have been shown in the literature to study the problem of rare category exploitation in the presence of scarce and noisy labels (C2). In [46], the authors study the problem of novel class classification on data streams. Instead of relying on redundant human-annotated labels or prior knowledge, they propose a semi-supervised framework to track the confidence changes of classifiers in order to detect the novel concept drifts and thus identify the novel classes. In [168], the authors propose a bi-level learning mechanism (shown in Figure 8), where a teacher model (i.e., curriculum learning scheme) gradually augments the training set with pseudo labels, and the student model (i.e., rare category characterization) returns the prediction results

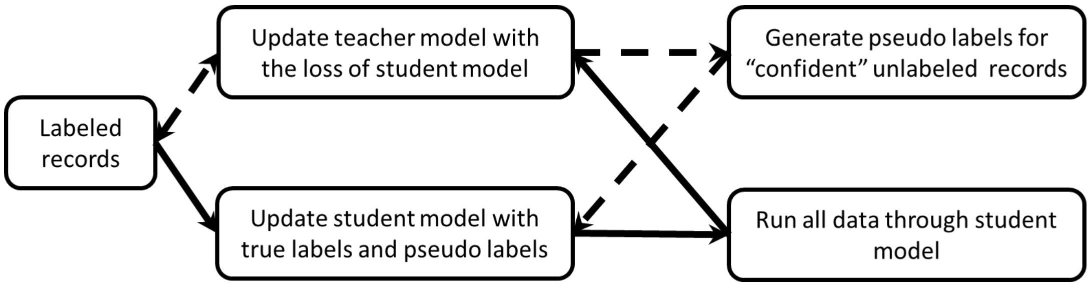


Fig. 8. A conceptual framework of bi-level learning for rare category exploitation [168], which is composed of a student learning model and a teacher learning model. The dashed arrows indicate the learning process of the teacher model that aims to generate pseudo labels, while the solid arrows indicate the learning process of the student model that is designed to predict rare category examples.

together with the prediction confidence to the teacher model. In general, the teacher model and the student model are trained in a mutually beneficial way, enabling the model to achieve better prediction accuracy in both rarity and label scarcity. Despite the label scarcity, many real-world applications (e.g., video classification [164]) come with complex data and noisy labels, which require the machine learning models to be trained with only weak supervision [174]. In [164], the authors propose a graph convolutional network (GCN) to clean the noisy labels. The GCN-based label noise cleaner is designed to provide supervisory signals from high-confidence snippets to low-confidence snippets, such that the classifier can be trained with “clean” supervision. In [146], the authors studied the problem of online transaction fraud detection. To deal with the noisy and complex user behavior information, they develop the Dual Importance-aware Factorization Machine (DIFM) that captures field value variations and field interactions simultaneously for online transaction fraud detection. In [84], the authors propose a multi-resolution rare category exploitation approach to identify online credit payment frauds at different granularities. In [145], the authors studied rare category exploitation under the scenarios of data contamination, by proposing an information-theoretic bound of performance degradation in terms of the data contamination ratio.

3.1.2 Local approaches. The local algorithms are also referred to seeded or targeted algorithms [56, 114, 153, 172]. This category of approaches essentially tackles the rare category exploitation as a local clustering problem. The key idea is to treat the labeled examples as seeds and return a compact cluster near the seed examples without exploring the whole dataset. In [16], the authors present a local kernel density ratio feature selection framework that seeks a salient feature subspace where the normal data points form a high-density region, while the rare examples form in a low-density region. In the learned subspace, the rare examples stand out and are easily distinguished by detection algorithms. In [126], Spielman and Teng present a design of a local clustering algorithm for massive graphs. The local algorithm is able to find a compact local cluster near a seed example, while the running time is near-linear with respect to the size of the returned cluster. In [11], Andersen et al. derive a mixing result for PageRank and propose a fast approximation method to compute the PageRank vectors. With that, the authors develop an improved local graph partitioning algorithm, which enables polylogarithmic time complexity with respect to the number of edges in the given graph. In [109], the authors propose a local and interactive rare event detection algorithm, that leverages the feedback from an anomaly-biased simulation environment and continuously updates the learned abnormality to novel rare categories.

However, the conventional local clustering algorithms are primarily designed for low-order connectivity patterns (e.g., edge in Table 1), while many applications exhibit high-order connectivity patterns (e.g., star in

Table 1. Illustrative examples of high-order connectivity patterns and local clustering algorithms.

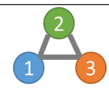
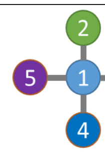
Connectivity Pattern $\mathbb{N}$	Illustration	Order of $\mathbb{N}$	Clustering Algorithms	Applications
Vertex		1^{st}-order	N/A	N/A
Edge		2^{nd}-order	1^{st}-order	Rare category detection [53]
3-node Line		3^{rd}-order	2^{nd}-order	Social community detection [71]
Triangle				
k-node Star		k^{th}-order	$(k-1)^{th}$-order	Synthetic identity detection [104]

Table 1). For example, the loop-shaped structure plays a significant role in detecting communities in user-item network [172]; a cluster of the star-shaped structures often indicates a red flag of the synthetic ID frauds in the Personal Identifiable Information (PII) networks [104]. To account for the crucial high-order connectivity patterns in real systems, Zhou et al. [172] and Yin et al. [154], for the first time, study the problem of local high-order graph clustering. In particular, the authors introduce the notion of high-order conductance and high-order diffusion kernel with the high-order Markov chain as the underlying model, which allows the end-users to model various types of high-order structures. Moreover, they generalize the well-known local clustering algorithms [11, 126] to the high-order setting and provide theoretical analysis regarding the effectiveness and scalability of the proposed algorithms. More recently, a bunch of work has been proposed to study the high-order connectivity patterns in more complicated scenarios, such as modeling evolving high-order structures on fine-grained temporal networks [167], capturing higher-order clusters in heterogeneous networks [23], and tracking high-order structures on time-evolving graphs [38].

3.2 Heterogeneous Rare Category Exploitation

In the era of big data, many application domains exhibit multiple types of heterogeneity, including data heterogeneity that originates from multiple information sources and task heterogeneity that originates from multiple application domains. In this subsection, we organize the existing work in the context of data heterogeneity and task heterogeneity.

3.2.1 Data Heterogeneity. Learning from data heterogeneity is often referred to as multi-view learning, as the data are collected from different sources, forming multiple views. In the data heterogeneity setting, rare category exploitation becomes even more challenging due to the conflicting, complementary nature among multiple views. Thus, it is crucial to leverage multiple views and identify the relevant features for distinguishing between the majority and minority classes. Here, we formally define the problem as follows, which aims to address (C1) rarity, (C2) label scarcity, and (C4) data heterogeneity in RCA.

Problem 3. ***Multi-View Rare Category Exploitation.***
Given: (i) a dataset $\mathcal{D}$ collected from V views, (ii) a small set of labeled examples $\mathcal{D}_l = \{(\mathbf{x}_i)\}_{i=1}^{N_l}$.

Find: (i) a prediction function $f(\cdot)$ to characterize rare category examples (2) a set of unlabeled examples which are likely coming from the rare categories.

To address Problem 5, a pivotal step is to leverage the view consistency and learn a unified representation to improve the performance of rare category exploitation. In the past decades, extensive multi-view learning frameworks [9, 40, 63, 64, 78–80, 93, 124, 159, 162] have been proposed for anomaly detection, which could be potential solutions for problem 5. However, these methods are mostly unsupervised and/or based on some heuristic functions, which might not well capture the characteristics of rare categories. *MUVIR* [165] is one of the principled efforts that is proposed to address Problem 5. *MUVIR* provides a generic rare category detection solution that is able to integrate the existing single-view rare category exploitation models for computing the overall posterior probability of each example. The key idea is to exploit the relationship among multiple views and estimate the overall posterior probability of examples coming from rare categories given data from multiple views. In particular, given data with multi-view features, one can train V distinct rare category exploitation models upon V views and compute view-specific posterior probability $P(y = \texttt{rare category}|\mathbf{x}^v)$ with respect to the v^{th} view. The authors propose the following theorem to effectively integrate view-specific posterior probability $P(y = \texttt{rare category}|\mathbf{x}^v)$ in a model-agnostic way.

Theorem 3.1 ([165]). *If the features from multiple views have weak dependence given the class label y_i [1], i.e., $P(\mathbf{x}|y = \texttt{rare category}) \geq \alpha \prod_{v=1}^{V} P(\mathbf{x}^v|y = \texttt{rare category})$, $\alpha > 0$, then*

$$P(y = \texttt{rare category}|\mathbf{x}) \geq C\Big(\prod_{v=1}^{V} P(y = \texttt{rare category}|\mathbf{x}^v)\Big) \times \left(\frac{\prod_{v=1}^{V} P(\mathbf{x}^v)}{P(\mathbf{x})}\right) \tag{12}$$

where $C = \frac{\alpha}{(p^2)^{V-1}}$ is a constant.

With Theorem 3.1, the overall posterior probability $P(y = \texttt{rare category}|\mathbf{x})$ can be approximated as follows.

$$P(y = \texttt{rare category}|\mathbf{x}) = \prod_{v=1}^{V} P(y = \texttt{rare category}|\mathbf{x}^v) \left(\frac{\prod_{v=1}^{V} P(\mathbf{x}^v)}{P(\mathbf{x})}\right)^d \tag{13}$$

where the marginal probabilities $P(\mathbf{x})$ and $P(\mathbf{x}^v)$ can be estimated via kernel density estimation, d is a non-negative parameter that balances the importance of the term related to the marginal probabilities. Moreover, a modified version of *MUVIR* is proposed to deal with the problems when the exact priors of minority classes are unknown. In the past few years, researchers have explored many high-impact applications that can be formulated as a multi-view rare category exploitation problem. To name a few, [75] propose an attention-based visual question-answering network to jointly model the input images and corpus and identify system errors or abnormal events; [29] studied the problem of multi-modal video anomaly detection, by proposing a bi-directional predictive network to regularize the prediction task from pixel-wise, cross-modal, and temporal-sequence levels; [125] propose an anomaly segmentation network for localizing defective areas in large-scale industrial manufacturing environments.

Another group of typical approaches for multi-view rare category exploitation is the semi-supervised classification models. For example, in [106], the authors propose a multi-view framework with adaptive neighborhood learning. The framework integrates multi-view clustering and semi-supervised classification, which allows learning attention for each view automatically. In [158], the authors develop a multi-view rare category exploitation framework for Alzheimer's disease diagnosis. They propose a multi-layer multi-view framework that automatically constructs a shared latent representation across multiple views and learns the mapping functions from the multi-view features to the prediction labels. In [140], the paper assumes the multiple views of the majority class examples are drawn from a unique distribution with different projection functions. With this assumption, the authors propose a hierarchical Bayesian model that computes the outlierness of unlabeled samples and thus identifies rare examples.

3.2.2 Task Heterogeneity. This subsection provides an overview of the rare category exploitation algorithms that have been proposed in the context of task heterogeneity. Formally speaking, in the presence of multiple tasks, we are given a dataset $\mathcal{D} = \{\mathcal{D}^1 \cup \ldots \cup \mathcal{D}^S\}$ that comes from S applications domains, and a small set of labeled examples from each task $\mathcal{D}_l^s = \{(\mathbf{x}_i^s)\}_{i=1}^{N_l^s}$, where N_l^s denotes the number of labeled examples in $\mathcal{D}^s$. Our goal is to capture the multi-modalities from different tasks/domains and learn a prediction model to accurately characterize the rare categories from different tasks/domains. The problem is commonly defined as follows, which aims to address (C1) rarity, (C2) label scarcity, (C3) non-separability, and (C5) dynamics in RCA.

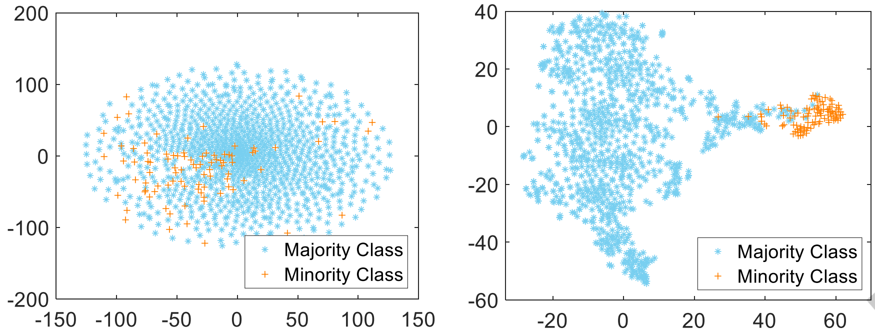
Problem 4. ***Multi-task Rare Category Exploitation.***
Given: (i) a dataset $\mathcal{D} = \{\mathcal{D}^1 \cup \ldots \cup \mathcal{D}^S\}$, where $\mathcal{D}^s$ is the dataset for the s-th task, $s = 1, \ldots, S$, (ii) a small set of labeled examples from each task $\mathcal{D}_l^s = \{(\mathbf{x}_i^s)\}_{i=1}^{N_l^s}$.
Find: (i) a prediction function $f(\cdot)$ to characterize rare category examples in each task $\mathcal{D}^s$ (2) a set of unlabeled examples which are likely coming from the rare categories in each task $\mathcal{D}^s$.

In the past decades, rich literature has been observed to study Problem 4, which can be summarized into two categories based on their objectives. The first category aims to learn a mixture of prediction functions that maximize the overall accuracy across all tasks. For example, [148] studies the problem of rare category exploitation in the presence of multiple tasks and multiple views by proposing a joint optimization framework named M^2LID. In particular, the authors start from a metric named border-degree as follows, which is used to capture the sharp changes in density near the boundary of rare categories in the feature space.

$$\texttt{BorderDegree}(\mathbf{x}) = \texttt{Hub}(\mathbf{x}) - \texttt{Authority}(\mathbf{x}) \tag{14}$$

where $\texttt{Hub}(\mathbf{x})$ and $\texttt{Authority}(\mathbf{x})$ indicate the Hub and Authority value [70] of example $\mathbf{x}$. Intuitively, the larger the border degree is, the higher probability that $\mathbf{x}$ belongs to the rare category. Then, to model the task relatedness and view consistency, they further formulate the problem by requiring: 1) task-specific learners to behave similarly on the features, 2) the view-based learners to behave similarly on the examples, and 3) the border-degree is negatively correlated with the prediction confidence score. Finally, to solve the optimization problem, they develop a block-coordinate-descent-based method to iteratively update the boundary characteristics of rare categories and multiple classification functions for different tasks. [62] propose a multi-modal learning algorithm that is based on a sparse mixture of sparse Gaussian graphical models. [163] aims to tackle the problem 4 in the presence of data scarcity and adversarial attacks. The authors propose a federated learning framework that performs multiple tasks simultaneously.

The second category is also referred to as transferable rare category exploitation. Instead of learning an optimal predictor for all tasks, it heavily emphasizes the target tasks. Some typical work include [8, 12, 72, 131, 138]. To name a few, in [12], the paper systematically evaluates the performance of anomaly detection frameworks using two types of transfer representation learning paradigms: (1) transfer learning from pre-trained networks, and (2) transfer learning from auxiliary tasks; in [72], the authors develop an autoencoder framework that learns the domain-specific latent vectors and thus improves the performance of detecting the rare examples. Unlike most transfer learning models that aim to learn the transferable representation for the target domain, [138] proposes to select a subset of labeled while relevant examples from the source domain to augment the target tasks. While the above methods all assume the target domains are related to the source domains, a more challenging scenario is so-called out-of-distribution rare category exploitation where the test data are unseen or from shifted distributions. To handle the out-of-distribution scenarios, [100] propose an analytical framework with provable guarantees to characterize and understand the out-of-distribution detection in open-world applications. In [130],



(a) Original Feature Space (b) Embedding Space

Fig. 9. Rare category oriented network representation: the majority and minority classes are not separable in the original feature space, but become well separated in the embedding space [168].

the authors propose an Adaptive In-Out-aware Learning framework to jointly model the mixed in-distribution and out-of-distribution data samples. In [2], the authors propose a graph context reasoning network for out-of-distribution object detection, by constructing a representation graph and a context graph constructed from the input image.

4 UNDERSTANDING RARE CATEGORY

Despite the tremendous success in exploring and exploiting rare category examples, many advanced methods, especially deep learning models, often remain the black box in nature. In contrast, many fundamental industries have to follow highly regulated processes - requiring prediction models to be interpretable and the output results to meet compliance. Therefore, a natural research question here is how to make our models transparent to the end-user by identifying the right context (e.g., key factors, representative entities, critical timestamps). In this section, we systematically discuss how to understand rare category patterns from the following two directions: (1) rare category representation; (2) rare category interpretation. The first direction aims to diagnose from the data perspective (i.e., *how is the data distributed? which piece of information is more valuable than the others for a given task?*), while the second direction aims to diagnose from the model perspective (i.e., *why does the model make a certain prediction on a particular piece of information?*).

4.1 Rare Category Representation

In this task, we aim to learn a rare-category-oriented embedding representation, such that the rare examples (e.g., security threats) are well separated from the majority classes (e.g., normal activities). Figure 9 presents a typical example of rare category representation learning, where the majority class and the minority class are overlapped together in the original features space while becoming separated in the learned embedding space. Formally speaking, the problem can be defined as follows.

Problem 5. ***Rare Category Representation.***
Given: (i) a dataset $\mathcal{D}$ that consists of n samples, (ii) a small set of labeled examples $\mathcal{D}_l = \{(\mathbf{x}_i)\}_{i=1}^{N_l}$, (iii) the user-defined embedding dimension d.
Find: (i) a rare category representation model $g(\cdot)$, (ii) a d-dimensional embedding representation $\mathbf{E} \in \mathbb{R}^{n \times d}$ that well preserves the underlying distribution of rare categories.

In the past decade, extensive work has been done to develop representation learning methods to learn underlying distribution for the rare category patterns. The key idea behind these methods is to track the 'footprints' of rare categories via external information (e.g., supervision from oracles) or prior knowledge (e.g., features indicating local density changes [52, 149]), and then leverage this information to regularize the representation learning models to extract low-dimensional and salient embedding for rare category analysis. To name a few, [5] proposes a set of egonet features (e.g., weighted and unweighted in- and out-degree, number of neighbors, number of triangles) from each snapshot of the graph sequence to model the "behavior" of each node, which shows the effectiveness in spotting rare category examples. [92] aims to spot anomalous patterns in a stream of heterogeneous graphs containing different types of nodes and edges by proposing a clustering-based anomaly detection approach that can simultaneously address the heterogeneity and streaming nature of the input data. In particular, the authors introduce a novel embedding mechanism that can encode the heterogeneous streaming graph into a vector representation, which will be used to perform clustering and identify anomalous patterns. [60] studies the problem of image classification when data in the vision domain exhibit highly-skewed class distribution by learning a deep representation such that rare categories are easily separable from the majority class by the contemporary classification methods. In particular, this paper first proposes to learn the data representation by maintaining both inter-cluster and inter-class margins that reduce the class imbalance inherent in the local data neighborhood. The proposed framework is built based on a CNN framework through a quintuplet sampling scheme and the associated triple-header hinge loss.

More recently, due to the notable success of network-embedding approaches, various network-embedding-based approaches are proposed for rare category analysis. For example, *SPARC* [168] is one of the first rare-category-oriented network embedding frameworks, which aims to learn a salient representation to characterize rare category examples. Inspired by the family of curriculum learning that simulates the cognitive mechanism of human beings, *SPARC* gradually selects the key network contextual information and learns the rare-category-oriented network representation by shifting from the 'easy' concept to the 'difficult' concept. The results show that (1) *SPARC* enables users to visualize the network layout in a salient embedding space, where the majority classes and minority classes are well separated, and (2) *SPARC* is able to identify valuable contextual information, which provides interpretation and guidance in the task of rare category characterization. Here we compare *SPARC* with three state-of-the-art network embedding algorithms, including two unsupervised methods, i.e., DeepWalk [113] and LINE [132], and one semi-supervised method, i.e., PLANETOID [150]. In particular, we first map the given network into a 129-dimensional space with different embedding methods. Then we employ the nonlinear dimensionality reduction method, i.e., t-SNE [90], to a 2-D space for better visualization, which is shown in Figure 10. We can clearly observe that the rare examples are better clustered using the proposed method than all the baseline methods.

Following *SPARC*, [144] proposes a biased random walk model named *VDRW* for learning imbalanced network representation. *VDRW* is designed to dynamically adjust the transition probability matrix each time a node is visited by the random particle. Via *VDRW*, the authors propose a pair of sampling strategies, i.e., context sampling and balanced-bath sampling, to learn network representation using skip-gram model [96]. NetWalk [157] is developed for rare event detection by learning the network representations, which can be updated dynamically as the network evolves. In particular, NetWalk first extracts the dynamic network context and encodes the vertices of the dynamic network to a low-dimensional representation. Then, a clustering-based scheme is employed to incrementally and dynamically track the malicious patterns in the dynamic networks. [34] also proposes a GCN-based framework for predicting future events by capturing contextual information from the raw data. The proposed framework first extracts graph representations of the events documents, then learns to predict the occurrence of future events and identify the events of interest (e.g., anomaly patterns).

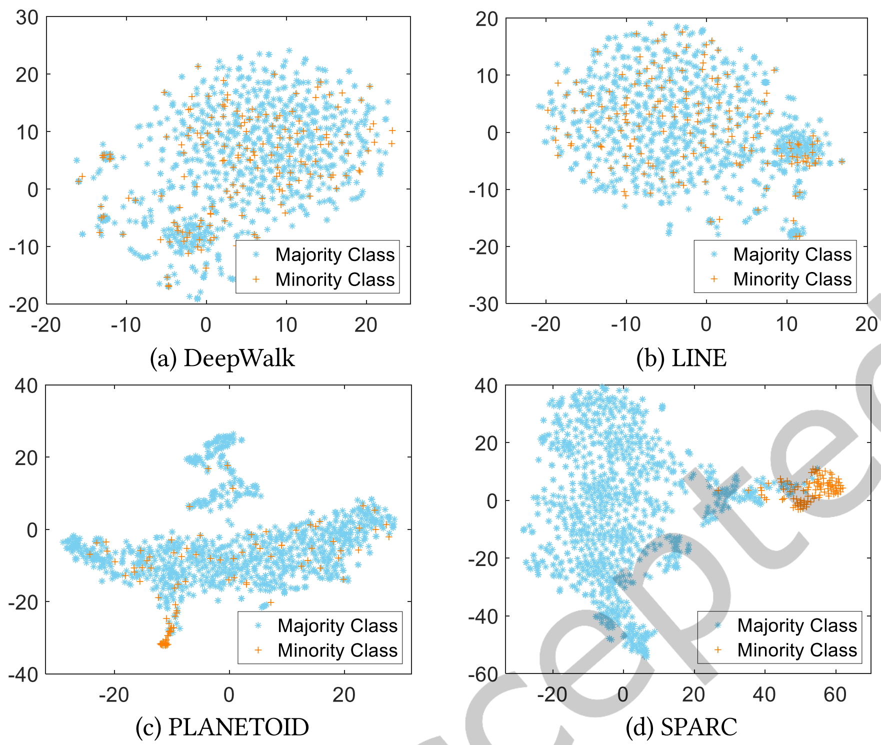


Fig. 10. Preliminary results comparing network embedding using various algorithms [168].

4.2 Rare Category Interpretation

Interpretability in machine learning models plays a crucial role in many high-impact domains and rare category analysis. In this task, we aim to characterize the rare categories with a predictive model and interpret the output from this model by providing the relevant clues, such as the relevant connectivity patterns, the relevant data sources, features, the relevant timestamps from time-series data, etc. Specifically, we define the problem setting as follows.

Problem 6. ***Rare Category Interpretation.***
Given: (i) a dataset $\mathcal{D}$ that consists of n samples, (ii) a rare category analysis model $f(\cdot)$.
Find: interpretation over the output from $f(\cdot)$ on $\mathcal{D}$.

Despite the rich literature in the context of Explainable Artificial Intelligence (XAI) [15, 31, 44, 61, 97, 98, 133], interpreting rare category analysis models is still challenging due to the following reasons: (1) rare category analysis models are naturally "biased" (i.e., focusing minority patterns instead of whole data distribution), (2) it is often the case that rare category analysis model $f(\cdot)$ is trained in the scarce of labels, (3) the learning process of rare category analysis model $f(\cdot)$ may frequently involve oracles and operate in a human-in-the-loop fashion (see Figure 5 for rare category exploration). Simply replacing $f(\cdot)$ with conventional interpretable machine learning models may deteriorate the prediction performance by providing misleading characterizations or injecting additional bias [74].

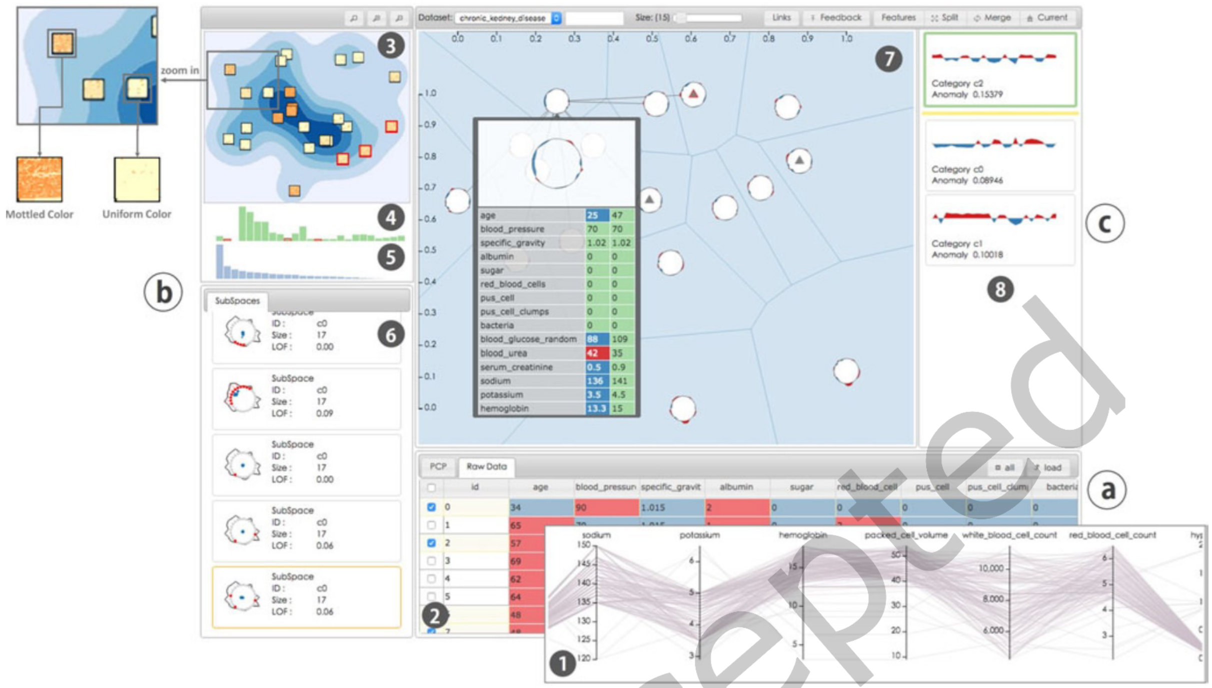


Fig. 11. User interface of *RCLens*, that is composed of three key modules: the data explorer module (shown in subfigure **a**), the feature explorer module (shown in subfigure **b**), and the category explorer module (shown in subfigure **c**).

To address Problem 6 with the above challenges, a few initial attempts have been made in the visualization domain. For example, *RCLens* [82] is an interactive visual analytic system, which is designed to explore and identify rare category examples with the guidance of end-users. Figure 12 presents the user interface of *RCLens*, which actively visualizes and interprets the three learning phases of rare category analysis models to the end-users. In particular, in the data exploration phase (shown in subfigure **a**), the system conducts personalized visualization of the data via a data querying and filtering mechanism; in the feature selection phase (shown in subfigure **a**), the system provides relevant clues and statistics (e.g., correlations of features) for the end-users to investigate the feature dimensions; in the rare category analysis model (shown in subfigure **c**), the system visualizes the identified rare categories to end-users and then refines the predictions by leveraging the users' feedback.

Later on, [107] proposes a visual analytic system named RCAnalyzer for studying rare category patterns in dynamic systems. The user interface of the *RCAnalyzer* is shown in Figure 12, which includes: (a) a timeline view showing the overview of the given dynamic networks; (b) the matrices view showing the neighborhood contextual information of each node; (c) the example view showing the feature distribution of rare patterns; (d) the label result view showing the history prediction results as well as the model diagnosis. To ensure the interpretability of rare category analysis models, the system is able to actively visualize each identified rare example in the context (e.g., neighborhood structures in the adjacent timestamps, etc.) to help oracles understand and examine the prediction results.

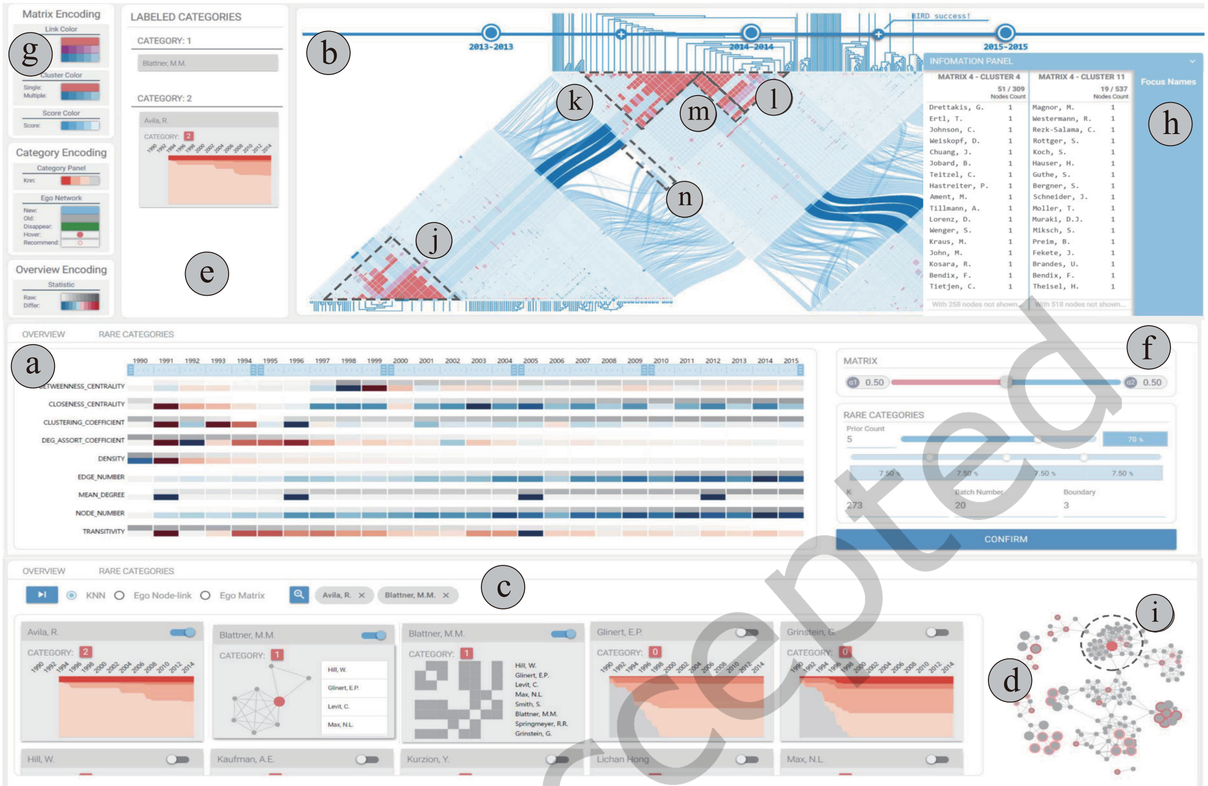


Fig. 12. User interface of RCAnalyzer, which is composed of (a) the timeline view; (b) the matrices view; (c)the instance view; (d) the sub-network view; (e) the label result view; (f) the parameter panel; (g) the encoding panel; and (h) the information panel.

5 DATASETS AND REPRESENTATIVE METHODS

This section summarizes the popular evaluation datasets and representative algorithms across different data types and downstream tasks.

5.1 Datasets

While rare categories are prevalent across various domains (e.g., finance), collecting and annotating rare examples is extremely time-consuming and labor-expensive (e.g., annotating money laundering activities). Early studies (e.g., [48, 50, 52, 53, 111]) may evaluate their models on synthetic or semi-synthetic datasets that are either generated from a pre-defined data distribution or manipulated from the public benchmark datasets. However, the evaluation on synthetic or semi-synthetic datasets may not reflect the real performance of rare category analysis models in real applications. To promote the future development for rare category analysis, Table 2 summarizes a collection of public benchmark datasets, which ranges from tabular data to graph-structured data, from static data to dynamic data.

Data Type	Dataset	Description of Rare Categories
Tabular Data	20NewsGroup [68]	Rare topics of newspapers
Tabular Data	Donors [108]	Outstanding projects proposed by K-12 teachers
Tabular Data	Census [35]	High-income people in U.S. census bureau database
Tabular Data	Fraud [116]	Fraudulent credit card transaction
Tabular Data	UNSW-NB 15 [101]	Network intrusion
Tabular Data	URL [88]	Malicious URLs
Tabular Data	campaign data [35]	Rarely successful campaigning records
Tabular Data	thyroid [35]	Rare disease
Time series Data	Vandal [73]	Wikipedia vandal warning
Time series Data	Spam [141]	Web spams
Time series Data	HDFS [147]	System errors
Graph Data	Tox21 [89]	Toxic environmental chemicals
Graph Data	ENZYMES [19]	Rare category of proteins
Graph Data	NCI1 [139]	Rare category of chemical compounds

Table 2. Publicly Accessible Real-world Datasets for Rare Category Analysis.

5.2 Representative Methods

Comparison evaluation with representative algorithms serves a pivotal role in the development of machine learning algorithms. In Table 3, we enumerate and summarize 25 representative rare category analysis algorithms w.r.t. the data types (i.e., tabular data, time-series data, graph data) and downstream tasks (i.e., exploration, exploitation, representation, and visualization). Since the methods in Table 3 are developed for diverse tasks and datasets, we are unable to provide a universal meta-analysis of their performance in a unified setting. Instead, to gain insight and in-depth understanding of the listed methods in Table 3, we summarize our major observations as follows: (i) most methods are designed in an unsupervised or semi-supervised setting; (ii) sampling-based methods [137, 165] and anomaly detection-based methods [134, 160] are still the mainstream solutions in rare category exploration; (iii) rare category visualization and interpretation are under-explored.

6 CONCLUSION AND FUTURE DIRECTIONS

Unlike previous surveys, this paper aims to present a comprehensive pipeline of the recent advances in rare category analysis. We start from the *de-novo* step without any label information and survey the rare category exploration techniques in the setting of homogeneous data and heterogeneous data. Then, we review the rare category exploitation methods that aim to characterize a compact representation of the minority classes in order to discover all the rare examples with high accuracy. At last, we discuss the problem of rare category explanation regarding how to learn a salient representation of rare categories as well as how to interpret the prediction results. Despite the significant developments in rare category analysis in the past, we identify some exciting research opportunities as follows:

- **Rare Category Generation.** With the dramatically increasing demand for machine learning systems as service providers (e.g., social networking, online advertising, data security, etc.), a massive amount of data is generated and collected from a variety of domains. However, the collected rare data are often noisy, sparse, less annotated, and evolving over time. Thus, directly training machine learning models from the raw data would introduce inevitable model bias and largely degrade the model performance in downstream applications (e.g., rare category characterization, rare category explanation). Moreover, the

Data Type	Algorithm	Task
Tabular Data	*NNDB* [52]	Rare Category Exploration
Tabular Data	*ACT* [143]	Rare Category Exploration
Tabular Data	*SEDER* [50]	Rare Category Exploration
Tabular Data	*HMS* [137]	Rare Category Exploration
Tabular Data	*MUVIR* [165]	Rare Category Exploitation
Tabular Data	*RACH* [54]	Rare Category Exploitation
Tabular Data	*MLAN* [106]	Rare Category Exploitation
Tabular Data	*RCLens* [83]	Rare Category Visualization
Time series Data	*SUITS* [81]	Rare Category Exploration
Time series Data	*CDLC* [18]	Rare Category Exploration
Time series Data	*MatrixProfile* [152]	Rare Category Exploration
Time series Data	*BIRAD* [166]	Rare Category Exploration & Exploitation
Time series Data	*RRCF* [43]	Rare Category Exploration
Time series Data	*Sand* [46]	Rare Category Exploitation
Graph Data	*Oddball* [6]	Rare Category Exploration
Graph Data	*NrMF* [134]	Rare Category Exploration
Graph Data	*HiDDen* [160]	Rare Category Exploration
Graph Data	*FocusCO* [112]	Rare Category Exploration
Graph Data	*BIRD* [170]	Rare Category Exploration
Graph Data	*StreamSpot* [92]	Rare Category Exploration
Graph Data	*AnomRank* [155]	Rare Category Exploration
Graph Data	*SPARC* [168]	Rare Category Exploitation & Representation
Graph Data	*HOSPLOC* [172]	Rare Category Exploitation
Graph Data	*L-MEGA* [38]	Rare Category Exploitation
Graph Data	*RCAnalyzer* [107]	Rare Category Visualization

Table 3. Representative Algorithms in Data Perspective.

ever-increasing size of data, together with the difficulty of releasing and sharing them, has made data generation a fundamental problem that is key in many high-impact applications, including fraud detection, recommendation, data security, and many more. Hence, it is critical to develop deep generative models that enable scalable modeling of real data to extract key contextual information, distill knowledge, and generate plausible patterns for data augmentation in rare category analysis.

- **Long-Tail Category Analysis.** In the past decade, deep learning has achieved remarkable success in various learning tasks (e.g., image classification, speech recognition, link prediction) through training "big models" upon "big data". However, beyond these well-studied tasks (e.g., image classification over domestic cats and wild cats) with rich training data, the vast majority of real-world entities and patterns (e.g., identification of honest employees and malicious insiders in a large institution) are less-explored and lack of observational and annotated data, which often corresponds to the "long-tail" categories. Unlike the existing works on rare category analysis that focus on one or a few rare categories, here we face a massive amount of under-represented categories from a "long-tail" distribution. Moreover, the current machine learning systems are mostly tailored to specific learning scenarios, making them fail to deliver their promises in detecting the targets of interest in the presence of distribution changes (e.g., dynamic systems). Two fundamental research questions are as follows: (Q1) *How to comprehend such massive "long-tail" categories in the inherent paucity of observational and annotated data?* (Q2) *How to capture the targets of interest (e.g., rare category examples) given a novel data distribution?*

- **Robust Rare Category Analysis.** Robust rare category analysis is another fundamental while quite open research problem, which is attracting a surge of attention from many high-impact domains (e.g., spam detection, financial fraud, and system diagnosis). For example, in financial fraud detection, *how can we measure the entity sensitivity, algorithmic robustness, task hardness, and model generalization, given a prediction model? how can we achieve operational robustness and adversarial robustness in the presence of the external disturbance (e.g., noise, missing values, outliers, adversarial attacks)?* Despite the extensive work on adversarial machine learning, the vast majority of the previous works assume a balanced data distribution while neglecting realistic cases where the data is highly skewed, and the targets of interest are under-represented. Compared to conventional machine learning tools, rare category analysis models could be more sensitive and vulnerable in the presence of adversarial attacks due to the rarity, non-separability, and label scarcity of rare category examples.
- **Human-AI Collaborative Rare Category Analysis.** Many industries follow highly regulated processes, which require the RCA models to be interpretable and human experts to be involved in the prediction loop. The human-in-the-loop fashion [24, 52] enables the RCA models to leverage the human intelligence to alleviate the label scarcity issue and improve the prediction performance in high-stake applications, such as failure prediction in safety-critical systems and rare disease diagnosis in health care. However, as human intelligence is naturally a black box to AI models, a fundamental research question remains nascent: *how can we have a deeper understanding of the underlying mechanism of human intelligence, thus enabling unimpeded knowledge transfer between human intelligence and AI?* A promising while challenging research direction lies in how to build an integrated and interactive system for accurate and trustworthy human-AI collaborative rare category analysis.

ACKNOWLEDGMENTS

This work is supported by National Science Foundation under Awards No. IIS-1947203, IIS-2002540, IIS-2137468, and IIS-2117902. The views and conclusions are those of the authors and should not be interpreted as representing the official policies of the funding agencies or the government.

REFERENCES

[1] Steven P. Abney. 2002. Bootstrapping. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*.

[2] Manoj Acharya, Anirban Roy, Kaushik Koneripalli, Susmit Jha, Christopher Kanan, and Ajay Divakaran. 2022. Detecting Out-Of-Context Objects Using Graph Contextual Reasoning Network. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 629–635. https://doi.org/10.24963/ijcai.2022/89

[3] Prerna Agarwal, Richa Verma, Ayush Agarwal, and Tanmoy Chakraborty. 2018. DyPerm: Maximizing permanence for dynamic community detection. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 437–449.

[4] Amirhossein Ahmadian and Fredrik Lindsten. 2021. Likelihood-free Out-of-Distribution Detection with Invertible Generative Models. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, Zhi-Hua Zhou (Ed.). ijcai.org, 2119–2125. https://doi.org/10.24963/ijcai.2021/292

[5] Leman Akoglu and Christos Faloutsos. 2010. Event detection in time series of mobile communication graphs. In *Army science conference*, Vol. 1.

[6] Leman Akoglu, Mary McGlohon, and Christos Faloutsos. 2010. Oddball: Spotting anomalies in weighted graphs. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 410–421.

[7] Leman Akoglu, Hanghang Tong, and Danai Koutra. 2015. Graph based anomaly detection and description: a survey. *Data mining and knowledge discovery* 29, 3 (2015), 626–688.

[8] Samir Al-Stouhi and Chandan K. Reddy. 2016. Transfer learning for class imbalance problems with inadequate data. *Knowl. Inf. Syst.* 48, 1 (2016), 201–228. https://doi.org/10.1007/s10115-015-0870-3

[9] Alejandro Marcos Alvarez, Makoto Yamada, Akisato Kimura, and Tomoharu Iwata. 2013. Clustering-based anomaly detection in multi-view data. In *22nd ACM International Conference on Information and Knowledge Management, CIKM'13, San Francisco, CA, USA, October 27 - November 1, 2013*, Qi He, Arun Iyengar, Wolfgang Nejdl, Jian Pei, and Rajeev Rastogi (Eds.). ACM, 1545–1548. https://doi.org/10.1145/2505515.2507840

[10] Mehrnaz Amjadi and Theja Tulabandhula. 2018. Block-Structure Based Time-Series Models For Graph Sequences. *arXiv preprint arXiv:1804.08796* (2018).

[11] Reid Andersen, Fan R. K. Chung, and Kevin J. Lang. 2006. Local Graph Partitioning using PageRank Vectors. In *47th Annual IEEE Symposium on Foundations of Computer Science (FOCS 2006), 21-24 October 2006, Berkeley, California, USA, Proceedings*. IEEE Computer Society, 475–486. https://doi.org/10.1109/FOCS.2006.44

[12] Jerone Andrews, Thomas Tanay, Edward J Morton, and Lewis D Griffin. 2016. Transfer representation-learning for anomaly detection. JMLR.

[13] Ana Paula Appel, Renato LF Cunha, Charu C Aggarwal, and Marcela Megumi Terakado. 2018. Temporally evolving community detection and prediction in content-centric networks. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 3–18.

[14] Miguel Araujo, Spiros Papadimitriou, Stephan Günnemann, Christos Faloutsos, Prithwish Basu, Ananthram Swami, Evangelos E Papalexakis, and Danai Koutra. 2014. Com2: fast automatic discovery of temporal ('comet') communities. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining*. Springer, 271–283.

[15] Alejandro Barredo Arrieta, Natalia Díaz-Rodríguez, Javier Del Ser, Adrien Bennetot, Siham Tabik, Alberto Barbado, Salvador García, Sergio Gil-López, Daniel Molina, Richard Benjamins, et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115.

[16] Fatemeh Azmandian, Ayse Yilmazer, Jennifer G. Dy, Javed A. Aslam, and David R. Kaeli. 2012. GPU-Accelerated Feature Selection for Outlier Detection Using the Local Kernel Density Ratio. In *12th IEEE International Conference on Data Mining, ICDM 2012, Brussels, Belgium, December 10-13, 2012*, Mohammed Javeed Zaki, Arno Siebes, Jeffrey Xu Yu, Bart Goethals, Geoffrey I. Webb, and Xindong Wu (Eds.). IEEE Computer Society, 51–60. https://doi.org/10.1109/ICDM.2012.51

[17] Michael J Bamshad, Sarah B Ng, Abigail W Bigham, Holly K Tabor, Mary J Emond, Deborah A Nickerson, and Jay Shendure. 2011. Exome sequencing as a tool for Mendelian disease gene discovery. *Nature Reviews Genetics* 12, 11 (2011), 745.

[18] Nurjahan Begum and Eamonn Keogh. 2014. Rare time series motif discovery from unbounded streams. *Proceedings of the VLDB Endowment* 8, 2 (2014), 149–160.

[19] Karsten M. Borgwardt, Cheng Soon Ong, Stefan Schönauer, S. V. N. Vishwanathan, Alexander J. Smola, and Hans-Peter Kriegel. 2005. Protein function prediction via graph kernels. In *Proceedings Thirteenth International Conference on Intelligent Systems for Molecular Biology 2005, Detroit, MI, USA, 25-29 June 2005*. 47–56. https://doi.org/10.1093/bioinformatics/bti1007

[20] Paula Branco, Luís Torgo, and Rita P Ribeiro. 2016. A survey of predictive modeling on imbalanced domains. *ACM Computing Surveys (CSUR)* 49, 2 (2016), 31.

[21] Suratna Budalakoti, Ashok N Srivastava, Ram Akella, and Eugene Turkov. 2006. Anomaly detection in large sets of high-dimensional symbol sequences. (2006).

[22] Lei Cao, Di Yang, Qingyang Wang, Yanwei Yu, Jiayuan Wang, and Elke A Rundensteiner. 2014. Scalable distance-based outlier detection over high-volume data streams. In *2014 IEEE 30th International Conference on Data Engineering*. IEEE, 76–87.

[23] Aldo G. Carranza, Ryan A. Rossi, Anup Rao, and Eunyee Koh. 2020. Higher-order Clustering in Complex Heterogeneous Networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 25–35. https://doi.org/10.1145/3394486.3403045

[24] Chengliang Chai, Lei Cao, Guoliang Li, Jian Li, Yuyu Luo, and Samuel Madden. 2020. Human-in-the-loop Outlier Detection. In *Proceedings of the 2020 International Conference on Management of Data, SIGMOD Conference 2020, online conference [Portland, OR, USA], June 14-19, 2020*, David Maier, Rachel Pottinger, AnHai Doan, Wang-Chiew Tan, Abdussalam Alawini, and Hung Q. Ngo (Eds.). ACM, 19–33. https://doi.org/10.1145/3318464.3389772

[25] Raghavendra Chalapathy and Sanjay Chawla. 2019. Deep learning for anomaly detection: A survey. *arXiv preprint arXiv:1901.03407* (2019).

[26] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2009. Anomaly detection: A survey. *ACM computing surveys (CSUR)* 41, 3 (2009), 15.

[27] Varun Chandola, Arindam Banerjee, and Vipin Kumar. 2010. Anomaly detection for discrete sequences: A survey. *IEEE Transactions on Knowledge and Data Engineering* 24, 5 (2010), 823–839.

[28] Chaomei Chen. 2006. CiteSpace II: Detecting and visualizing emerging trends and transient patterns in scientific literature. *Journal of the American Society for information Science and Technology* 57, 3 (2006), 359–377.

[29] Chengwei Chen, Yuan Xie, Shaohui Lin, Angela Yao, Guannan Jiang, Wei Zhang, Yanyun Qu, Ruizhi Qiao, Bo Ren, and Lizhuang Ma. 2022. Comprehensive Regularization in a Bi-directional Predictive Network for Video Anomaly Detection. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022,*

The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022. AAAI Press, 230–238. https://ojs.aaai.org/index.php/AAAI/article/view/19898

[30] Xuan Hong Dang, Barbora Micenková, Ira Assent, and Raymond T Ng. 2013. Local outlier detection with interpretation. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 304–320.

[31] Arun Das and Paul Rad. 2020. Opportunities and challenges in explainable artificial intelligence (xai): A survey. *arXiv preprint arXiv:2006.11371* (2020).

[32] Ailin Deng, Adam Goodge, Lang Yi Ang, and Bryan Hooi. 2022. CADET: Calibrated Anomaly Detection for Mitigating Hardness Bias. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 2002–2008. https://doi.org/10.24963/ijcai.2022/278

[33] Ailin Deng and Bryan Hooi. 2021. Graph Neural Network-Based Anomaly Detection in Multivariate Time Series. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 4027–4035. https://ojs.aaai.org/index.php/AAAI/article/view/16523

[34] Songgaojun Deng, Huzefa Rangwala, and Yue Ning. 2019. Learning Dynamic Context Graphs for Predicting Social Events. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 1007–1016.

[35] Dheeru Dua and Casey Graff. 2017. UCI Machine Learning Repository. http://archive.ics.uci.edu/ml

[36] Michael Edward Edge and Pedro R Falcone Sampaio. 2009. A survey of signature based methods for financial fraud detection. *computers & security* 28, 6 (2009), 381–394.

[37] Bradley Efron and Robert Tibshirani. 1996. Using specially designed exponential families for density estimation. *The Annals of Statistics* 24, 6 (1996), 2431–2461.

[38] Dongqi Fu, Dawei Zhou, and Jingrui He. 2020. Local Motif Clustering on Time-Evolving Graphs. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 390–400. https://doi.org/10.1145/3394486.3403081

[39] Dongqi Fu, Dawei Zhou, Ross Maciejewski, Arie Croitoru, Marcus Boyd, and Jingrui He. 2023. Fairness-Aware Clique-Preserving Spectral Clustering of Temporal Graphs. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben (Eds.). ACM, 3755–3765. https://doi.org/10.1145/3543507.3583423

[40] Jing Gao, Wei Fan, Deepak Turaga, Srinivasan Parthasarathy, and Jiawei Han. 2011. A spectral framework for detecting inconsistency across multi-source object relationships. In *Data Mining (ICDM), 2011 IEEE 11th International Conference on*. IEEE, 1050–1055.

[41] Yuan Gao, Xiang Wang, Xiangnan He, Zhenguang Liu, Huamin Feng, and Yongdong Zhang. 2023. Addressing Heterophily in Graph Anomaly Detection: A Perspective of Graph Spectrum. In *Proceedings of the ACM Web Conference 2023, WWW 2023, Austin, TX, USA, 30 April 2023 - 4 May 2023*, Ying Ding, Jie Tang, Juan F. Sequeda, Lora Aroyo, Carlos Castillo, and Geert-Jan Houben (Eds.). ACM, 1528–1538. https://doi.org/10.1145/3543507.3583268

[42] Adam Goodge, Bryan Hooi, See-Kiong Ng, and Wee Siong Ng. 2022. LUNAR: Unifying Local Outlier Detection Methods via Graph Neural Networks. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 6737–6745. https://ojs.aaai.org/index.php/AAAI/article/view/20629

[43] Sudipto Guha, Nina Mishra, Gourav Roy, and Okke Schrijvers. 2016. Robust random cut forest based anomaly detection on streams. In *International conference on machine learning*. 2712–2721.

[44] David Gunning, Mark Stefik, Jaesik Choi, Timothy Miller, Simone Stumpf, and Guang-Zhong Yang. 2019. XAI—Explainable artificial intelligence. *Science Robotics* 4, 37 (2019), eaay7120.

[45] Manish Gupta, Jing Gao, Charu C Aggarwal, and Jiawei Han. 2013. Outlier detection for temporal data: A survey. *IEEE Transactions on Knowledge and Data Engineering* 26, 9 (2013), 2250–2267.

[46] Ahsanul Haque, Latifur Khan, and Michael Baron. 2016. Sand: Semi-supervised adaptive novel class detection and classification over data stream. In *THIRTIETH AAAI Conference on Artificial Intelligence*.

[47] D. M. Hawkins. 1980. *Identification of Outliers*. Springer. https://doi.org/10.1007/978-94-015-3994-4

[48] Haibo He and Edwardo A Garcia. 2009. Learning from imbalanced data. *IEEE Transactions on knowledge and data engineering* 21, 9 (2009), 1263–1284.

[49] Jingrui He. 2010. *Rare category analysis*. Ph.D. Dissertation.

[50] Jingrui He and Jaime Carbonell. 2009. Prior-free rare category detection. In *Proceedings of the 2009 SIAM International Conference on Data Mining*. SIAM, 155–163.

[51] Jingrui He and Jaime Carbonell. 2010. Coselection of features and instances for unsupervised rare category analysis. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 3, 6 (2010), 417–430.

[52] Jingrui He and Jaime G Carbonell. 2008. Nearest-neighbor-based active learning for rare category detection. In *Advances in neural information processing systems*. 633–640.

[53] Jingrui He, Yan Liu, and Richard Lawrence. 2008. Graph-based rare category detection. In *Data Mining, 2008. ICDM'08. Eighth IEEE International Conference on*. IEEE, 833–838.

[54] Jingrui He, Hanghang Tong, and Jaime Carbonell. 2010. Rare category characterization. In *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 226–235.

[55] Jingrui He, Hanghang Tong, and Jaime Carbonell. 2012. An effective framework for characterizing rare categories. *Frontiers of Computer Science* 6, 2 (2012), 154–165.

[56] Zengyou He, Xiaofei Xu, and Shengchun Deng. 2003. Discovering cluster-based local outliers. *Pattern Recognit. Lett.* 24, 9-10 (2003), 1641–1650. https://doi.org/10.1016/S0167-8655(03)00003-5

[57] Victoria Hodge and Jim Austin. 2004. A survey of outlier detection methodologies. *Artificial intelligence review* 22, 2 (2004), 85–126.

[58] Steven A Hofmeyr, Stephanie Forrest, and Anil Somayaji. 1998. Intrusion detection using sequences of system calls. *Journal of computer security* 6, 3 (1998), 151–180.

[59] Alexandre Hollocou, Julien Maudet, Thomas Bonald, and Marc Lelarge. 2017. A Streaming Algorithm for Graph Clustering. *CoRR* abs/1712.04337 (2017). arXiv:1712.04337 http://arxiv.org/abs/1712.04337

[60] Chen Huang, Yining Li, Chen Change Loy, and Xiaoou Tang. 2016. Learning deep representation for imbalanced classification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 5375–5384.

[61] Tsuyoshi Idé, Amit Dhurandhar, Jirí Navrátil, Moninder Singh, and Naoki Abe. 2021. Anomaly Attribution with Likelihood Compensation. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 4131–4138. https://ojs.aaai.org/index.php/AAAI/article/view/16535

[62] Tsuyoshi Idé, Dzung T. Phan, and Jayant Kalagnanam. 2017. Multi-task Multi-modal Models for Collective Anomaly Detection. In *2017 IEEE International Conference on Data Mining, ICDM 2017, New Orleans, LA, USA, November 18-21, 2017*, Vijay Raghavan, Srinivas Aluru, George Karypis, Lucio Miele, and Xindong Wu (Eds.). IEEE Computer Society, 177–186. https://doi.org/10.1109/ICDM.2017.27

[63] Tomoharu Iwata and Makoto Yamada. 2016. Multi-view anomaly detection via robust probabilistic latent variable models. In *Advances In Neural Information Processing Systems*. 1136–1144.

[64] Tomoharu Iwata and Makoto Yamada. 2016. Multi-view Anomaly Detection via Robust Probabilistic Latent Variable Models. In *Advances in Neural Information Processing Systems 29: Annual Conference on Neural Information Processing Systems 2016, December 5-10, 2016, Barcelona, Spain*, Daniel D. Lee, Masashi Sugiyama, Ulrike von Luxburg, Isabelle Guyon, and Roman Garnett (Eds.). 1136–1144. https://proceedings.neurips.cc/paper/2016/hash/0f96613235062963ccde717b18f97592-Abstract.html

[65] HV Jagadish, Nick Koudas, and S Muthukrishnan. 1999. Mining Deviants in a Time Series Database.. In *VLDB*, Vol. 99. 7–10.

[66] Miltiadis Kandias, Alexios Mylonas, Nikos Virvilis, Marianthi Theoharidou, and Dimitris Gritzalis. 2010. An insider threat prediction model. In *International Conference on Trust, Privacy and Security in Digital Business*. Springer, 26–37.

[67] Ramneet Kaur, Susmit Jha, Anirban Roy, Sangdon Park, Edgar Dobriban, Oleg Sokolsky, and Insup Lee. 2022. iDECODe: In-Distribution Equivariance for Conformal Out-of-Distribution Detection. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 7104–7114. https://ojs.aaai.org/index.php/AAAI/article/view/20670

[68] Fabian Keller, Emmanuel Müller, and Klemens Böhm. 2012. HiCS: High Contrast Subspaces for Density-Based Outlier Ranking. In *IEEE 28th International Conference on Data Engineering (ICDE 2012), Washington, DC, USA (Arlington, Virginia), 1-5 April, 2012*, Anastasios Kementsietsidis and Marcos Antonio Vaz Salles (Eds.). IEEE Computer Society, 1037–1048. https://doi.org/10.1109/ICDE.2012.88

[69] Kamran Khan, Saif Ur Rehman, Kamran Aziz, Simon Fong, and Sababady Sarasvady. 2014. DBSCAN: Past, present and future. In *The fifth international conference on the applications of digital information and web technologies (ICADIWT 2014)*. IEEE, 232–238.

[70] Jon M. Kleinberg. 1998. Authoritative Sources in a Hyperlinked Environment. In *Proceedings of the Ninth Annual ACM-SIAM Symposium on Discrete Algorithms, 25-27 January 1998, San Francisco, California, USA*, Howard J. Karloff (Ed.). ACM/SIAM, 668–677. http://dl.acm.org/citation.cfm?id=314613.315045

[71] Christine Klymko, David Gleich, and Tamara G Kolda. 2014. Using triangles to improve community detection in directed networks. *arXiv preprint arXiv:1404.5874* (2014).

[72] Atsutoshi Kumagai, Tomoharu Iwata, and Yasuhiro Fujiwara. 2019. Transfer Anomaly Detection by Inferring Latent Domain Representations. In *Advances in Neural Information Processing Systems 32: Annual Conference on Neural Information Processing Systems 2019, NeurIPS 2019, December 8-14, 2019, Vancouver, BC, Canada*, Hanna M. Wallach, Hugo Larochelle, Alina Beygelzimer, Florence d'Alché-Buc, Emily B. Fox, and Roman Garnett (Eds.). 2467–2477. https://proceedings.neurips.cc/paper/2019/hash/7895fc13088ee37f511913bac71fa66f-Abstract.html

[73] Srijan Kumar, Francesca Spezzano, and V. S. Subrahmanian. 2015. VEWS: A Wikipedia Vandal Early Warning System. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Sydney, NSW, Australia, August 10-13, 2015*, Longbing Cao, Chengqi Zhang, Thorsten Joachims, Geoffrey I. Webb, Dragos D. Margineantu, and Graham Williams (Eds.). ACM, 607–616. https://doi.org/10.1145/2783258.2783367

[74] Himabindu Lakkaraju and Osbert Bastani. 2020. " How do I fool you?" Manipulating User Trust via Misleading Black Box Explanations. In *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society*. 79–85.

[75] Doyup Lee, Yeongjae Cheon, and Wook-Shin Han. 2021. Regularizing Attention Networks for Anomaly Detection in Visual Question Answering. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 1845–1853. https://ojs.aaai.org/index.php/AAAI/article/view/16279

[76] Kevin H Lee, Lingzhou Xue, and David R Hunter. 2017. Model-based clustering of time-evolving networks through temporal exponential-family random graph models. *arXiv preprint arXiv:1712.07325* (2017).

[77] Dan Li, Dacheng Chen, Lei Shi, Baihong Jin, Jonathan Goh, and See-Kiong Ng. 2019. MAD-GAN: Multivariate anomaly detection for time series data with generative adversarial networks. *arXiv preprint arXiv:1901.04997* (2019).

[78] Kai Li, Sheng Li, Zhengming Ding, Weidong Zhang, and Yun Fu. 2018. Latent Discriminant Subspace Representations for Multi-View Outlier Detection. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 3522–3529. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/17401

[79] Sheng Li, Ming Shao, and Yun Fu. 2015. Multi-view low-rank analysis for outlier detection. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 748–756.

[80] Sheng Li, Ming Shao, and Yun Fu. 2015. Multi-View Low-Rank Analysis for Outlier Detection. In *Proceedings of the 2015 SIAM International Conference on Data Mining, Vancouver, BC, Canada, April 30 - May 2, 2015*, Suresh Venkatasubramanian and Jieping Ye (Eds.). SIAM, 748–756. https://doi.org/10.1137/1.9781611974010.84

[81] Xiaolei Li and Jiawei Han. 2007. Mining approximate top-k subspace anomalies in multi-dimensional time-series data. In *Proceedings of the 33rd international conference on Very large data bases*. VLDB Endowment, 447–458.

[82] Hanfei Lin, Siyuan Gao, David Gotz, Fan Du, Jingrui He, and Nan Cao. 2017. Rclens: Interactive rare category exploration and identification. *IEEE transactions on visualization and computer graphics* 24, 7 (2017), 2223–2237.

[83] Hanfei Lin, Siyuan Gao, David Gotz, Fan Du, Jingrui He, and Nan Cao. 2018. Rclens: Interactive rare category exploration and identification. *IEEE transactions on visualization and computer graphics* 24, 7 (2018), 2223–2237.

[84] Wangli Lin, Li Sun, Qiwei Zhong, Can Liu, Jinghua Feng, Xiang Ao, and Hao Yang. 2021. Online Credit Payment Fraud Detection via Structure-Aware Hierarchical Recurrent Neural Network. In *Proceedings of the Thirtieth International Joint Conference on Artificial Intelligence, IJCAI 2021, Virtual Event / Montreal, Canada, 19-27 August 2021*, Zhi-Hua Zhou (Ed.). ijcai.org, 3670–3676. https://doi.org/10.24963/ijcai.2021/505

[85] Boyang Liu, Pang-Ning Tan, and Jiayu Zhou. 2022. Unsupervised Anomaly Detection by Robust Density Estimation. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 4101–4108. https://ojs.aaai.org/index.php/AAAI/article/view/20328

[86] Ninghao Liu, Donghwa Shin, and Xia Hu. 2017. Contextual outlier interpretation. *arXiv preprint arXiv:1711.10589* (2017).

[87] Zhining Liu, Dawei Zhou, Yada Zhu, Jinjie Gu, and Jingrui He. 2020. Towards Fine-Grained Temporal Network Representation via Time-Reinforced Random Walk. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 4973–4980. https://aaai.org/ojs/index.php/AAAI/article/view/5936

[88] Justin Ma, Lawrence K. Saul, Stefan Savage, and Geoffrey M. Voelker. 2009. Identifying suspicious URLs: an application of large-scale online learning. In *Proceedings of the 26th Annual International Conference on Machine Learning, ICML 2009, Montreal, Quebec, Canada, June 14-18, 2009 (ACM International Conference Proceeding Series)*, Andrea Pohoreckyj Danyluk, Léon Bottou, and Michael L. Littman (Eds.), Vol. 382. ACM, 681–688. https://doi.org/10.1145/1553374.1553462

[89] Rongrong Ma, Guansong Pang, Ling Chen, and Anton van den Hengel. 2021. Deep Graph-level Anomaly Detection by Glocal Knowledge Distillation. *CoRR* abs/2112.10063 (2021). arXiv:2112.10063 https://arxiv.org/abs/2112.10063

[90] Laurens van der Maaten and Geoffrey Hinton. 2008. Visualizing data using t-SNE. *Journal of machine learning research* 9, Nov (2008), 2579–2605.

[91] Meghanath Macha and Leman Akoglu. 2018. Explaining anomalies in groups with characterizing subspace rules. *Data Mining and Knowledge Discovery* 32, 5 (2018), 1444–1480.

[92] Emaad Manzoor, Sadegh M Milajerdi, and Leman Akoglu. 2016. Fast memory-efficient anomaly detection in streaming heterogeneous graphs. In *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 1035–1044.

[93] 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*. ACM, 1545–1548.

[94] Markos Markou and Sameer Singh. 2003. Novelty detection: a review—part 1: statistical approaches. *Signal processing* 83, 12 (2003), 2481–2497.

[95] Bhaskar Mehta, Saurabh Nangia, Manish Gupta, and Wolfgang Nejdl. 2008. Detecting image spam using visual features and near duplicate detection. In *Proceedings of the 17th international conference on World Wide Web*. ACM, 497–506.

[96] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. 2013. Distributed representations of words and phrases and their compositionality. In *Advances in neural information processing systems*. 3111–3119.

[97] Yifei Ming, Hang Yin, and Yixuan Li. 2022. On the Impact of Spurious Correlation for Out-of-Distribution Detection. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 10051–10059. https://ojs.aaai.org/index.php/AAAI/article/view/21244

[98] Christoph Molnar. 2020. *Interpretable machine learning*. Lulu. com.

[99] Hossein Moradi Koupaie, Suhaimi Ibrahim, and Javad Hosseinkhani. 2014. Outlier detection in stream data by clustering method. *International Journal of Advanced Computer Science and Information Technology (IJACSIT) Vol* 2 (2014), 25–34.

[100] Peyman Morteza and Yixuan Li. 2022. Provable Guarantees for Understanding Out-of-Distribution Detection. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 7831–7840. https://ojs.aaai.org/index.php/AAAI/article/view/20752

[101] Nour Moustafa and Jill Slay. 2015. UNSW-NB15: a comprehensive data set for network intrusion detection systems (UNSW-NB15 network data set). In *2015 Military Communications and Information Systems Conference, MilCIS 2015, Canberra, Australia, November 10-12, 2015*. IEEE, 1–6. https://doi.org/10.1109/MilCIS.2015.7348942

[102] Shan Muthukrishnan, Rahul Shah, and Jeffrey Scott Vitter. 2004. Mining deviants in time series data streams. In *Scientific and Statistical Database Management, 2004. Proceedings. 16th International Conference on*. IEEE, 41–50.

[103] Alexandre Nairac, Neil Townsend, Roy Carr, Steve King, Peter Cowley, and Lionel Tarassenko. 1999. A system for the analysis of jet engine vibration data. *Integrated Computer-Aided Engineering* 6, 1 (1999), 53–66.

[104] Eric WT Ngai, Yong Hu, Yiu Hing Wong, Yijun Chen, and Xin Sun. 2011. The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision support systems* 50, 3 (2011), 559–569.

[105] Giang Hoang Nguyen, John Boaz Lee, Ryan A. Rossi, Nesreen K. Ahmed, Eunyee Koh, and Sungchul Kim. 2018. Continuous-Time Dynamic Network Embeddings. In *Companion of the The Web Conference 2018 on The Web Conference 2018, WWW 2018, Lyon , France, April 23-27, 2018*, Pierre-Antoine Champin, Fabien Gandon, Mounia Lalmas, and Panagiotis G. Ipeirotis (Eds.). ACM, 969–976. https://doi.org/10.1145/3184558.3191526

[106] Feiping Nie, Guohao Cai, and Xuelong Li. 2017. Multi-View Clustering and Semi-Supervised Classification with Adaptive Neighbours. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, February 4-9, 2017, San Francisco, California, USA*, Satinder P. Singh and Shaul Markovitch (Eds.). AAAI Press, 2408–2414. http://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14833

[107] Jiacheng Pan, Dongming Han, Fangzhou Guo, Dawei Zhou, Nan Cao, Jingrui He, Mingliang Xu, and Wei Chen. 2020. RCAnalyzer: visual analytics of rare categories in dynamic networks. *Frontiers Inf. Technol. Electron. Eng.* 21, 4 (2020), 491–506. https://doi.org/10.1631/FITEE.1900310

[108] Guansong Pang, Chunhua Shen, and Anton van den Hengel. 2019. Deep anomaly detection with deviation networks. In *Proceedings of the 25th ACM SIGKDD international conference on knowledge discovery & data mining*. 353–362.

[109] Guansong Pang, Anton van den Hengel, Chunhua Shen, and Longbing Cao. 2021. Toward Deep Supervised Anomaly Detection: Reinforcement Learning from Partially Labeled Anomaly Data. In *KDD '21: The 27th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, Singapore, August 14-18, 2021*, Feida Zhu, Beng Chin Ooi, and Chunyan Miao (Eds.). ACM, 1298–1308. https://doi.org/10.1145/3447548.3467417

[110] Dhaval Patel. 2023. Time Series Anomaly Detection Tool: Hands of Lab. In *AAAI Conference on Artificial Intelligence*.

[111] Dan Pelleg and Andrew W Moore. 2005. Active learning for anomaly and rare-category detection. In *Advances in neural information processing systems*. 1073–1080.

[112] Bryan Perozzi, Leman Akoglu, Patricia Iglesias Sánchez, and Emmanuel Müller. 2014. Focused clustering and outlier detection in large attributed graphs. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 1346–1355.

[113] Bryan Perozzi, Rami Al-Rfou, and Steven Skiena. 2014. Deepwalk: Online learning of social representations. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 701–710.

[114] Geo Pertea, Xiaoqiu Huang, Feng Liang, Valentin Antonescu, Razvan Sultana, Svetlana Karamycheva, Yuandan Lee, Joseph White, Foo Cheung, Babak Parvizi, Jennifer Tsai, and John Quackenbush. 2003. TIGR Gene Indices clustering tools (TGICL): a software system for fast clustering of large EST datasets. *Bioinform.* 19, 5 (2003), 651–652. https://doi.org/10.1093/bioinformatics/btg034

[115] Clifton Phua, Vincent Lee, Kate Smith, and Ross Gayler. 2010. A comprehensive survey of data mining-based fraud detection research. *arXiv preprint arXiv:1009.6119* (2010).

[116] Andrea Dal Pozzolo, Olivier Caelen, Reid A. Johnson, and Gianluca Bontempi. 2015. Calibrating Probability with Undersampling for Unbalanced Classification. In *IEEE Symposium Series on Computational Intelligence, SSCI 2015, Cape Town, South Africa, December 7-10, 2015*. IEEE, 159–166. https://doi.org/10.1109/SSCI.2015.33

[117] Stephen Ranshous, Shitian Shen, Danai Koutra, Steve Harenberg, Christos Faloutsos, and Nagiza F Samatova. 2015. Anomaly detection in dynamic networks: a survey. *Wiley Interdisciplinary Reviews: Computational Statistics* 7, 3 (2015), 223–247.

[118] Umaa Rebbapragada, Pavlos Protopapas, Carla E Brodley, and Charles Alcock. 2009. Finding anomalous periodic time series. *Machine learning* 74, 3 (2009), 281–313.

[119] Maurizio Scarpa, Zsuzsanna Almássy, Michael Beck, Olaf Bodamer, Iain A Bruce, Linda De Meirleir, Nathalie Guffon, Encarna Guillén-Navarro, Pauline Hensman, Simon Jones, et al. 2011. Mucopolysaccharidosis type II: European recommendations for the diagnosis and multidisciplinary management of a rare disease. *Orphanet journal of rare diseases* 6, 1 (2011), 72.

[120] François Schnitzler, Thomas Liebig, Shie Mannor, Gustavo Souto, Sebastian Bothe, and Hendrik Stange. 2014. Heterogeneous Stream Processing for Disaster Detection and Alarming. In *IEEE International Conference on Big Data*. IEEE Press, 914–923.

[121] Karlton Sequeira and Mohammed Zaki. 2002. ADMIT: anomaly-based data mining for intrusions. In *Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 386–395.

[122] Shubhranshu Shekhar, Neil Shah, and Leman Akoglu. 2021. FairOD: Fairness-aware Outlier Detection. In *AIES '21: AAAI/ACM Conference on AI, Ethics, and Society, Virtual Event, USA, May 19-21, 2021*, Marion Fourcade, Benjamin Kuipers, Seth Lazar, and Deirdre K. Mulligan (Eds.). ACM, 210–220. https://doi.org/10.1145/3461702.3462517

[123] Lifeng Shen, Zhongzhong Yu, Qianli Ma, and James T. Kwok. 2021. Time Series Anomaly Detection with Multiresolution Ensemble Decoding. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 9567–9575. https://ojs.aaai.org/index.php/AAAI/article/view/17152

[124] Xiang-Rong Sheng, De-Chuan Zhan, Su Lu, and Yuan Jiang. 2019. Multi-View Anomaly Detection: Neighborhood in Locality Matters. In *The Thirty-Third AAAI Conference on Artificial Intelligence, AAAI 2019, The Thirty-First Innovative Applications of Artificial Intelligence Conference, IAAI 2019, The Ninth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2019, Honolulu, Hawaii, USA, January 27 - February 1, 2019*. AAAI Press, 4894–4901. https://doi.org/10.1609/aaai.v33i01.33014894

[125] Jou Won Song, Kyeongbo Kong, Ye In Park, Seong Gyun Kim, and Suk-Ju Kang. 2021. AnoSeg: Anomaly Segmentation Network Using Self-Supervised Learning. *CoRR* abs/2110.03396 (2021). arXiv:2110.03396 https://arxiv.org/abs/2110.03396

[126] Daniel A. Spielman and Shang-Hua Teng. 2013. A Local Clustering Algorithm for Massive Graphs and Its Application to Nearly Linear Time Graph Partitioning. *SIAM J. Comput.* 42, 1 (2013), 1–26. https://doi.org/10.1137/080744888

[127] Kumar Sricharan and Kamalika Das. 2014. Localizing anomalous changes in time-evolving graphs. In *Proceedings of the 2014 ACM SIGMOD international conference on Management of data*. ACM, 1347–1358.

[128] Jimeng Sun, Christos Faloutsos, Christos Faloutsos, Spiros Papadimitriou, and Philip S Yu. 2007. Graphscope: parameter-free mining of large time-evolving graphs. In *Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 687–696.

[129] Jimeng Sun, Huiming Qu, Deepayan Chakrabarti, and Christos Faloutsos. 2005. Neighborhood formation and anomaly detection in bipartite graphs. In *Data Mining, Fifth IEEE International Conference on*. IEEE, 8–pp.

[130] Yi-Xuan Sun and Wei Wang. 2022. Exploiting Mixed Unlabeled Data for Detecting Samples of Seen and Unseen Out-of-Distribution Classes. In *Thirty-Sixth AAAI Conference on Artificial Intelligence, AAAI 2022, Thirty-Fourth Conference on Innovative Applications of Artificial Intelligence, IAAI 2022, The Twelveth Symposium on Educational Advances in Artificial Intelligence, EAAI 2022 Virtual Event, February 22 - March 1, 2022*. AAAI Press, 8386–8394. https://ojs.aaai.org/index.php/AAAI/article/view/20814

[131] Ying Sun, Wenjun Wang, Nannan Wu, Wei Yu, and Xue Chen. 2020. Anomaly Subgraph Detection with Feature Transfer. In *CIKM '20: The 29th ACM International Conference on Information and Knowledge Management, Virtual Event, Ireland, October 19-23, 2020*, Mathieu d'Aquin, Stefan Dietze, Claudia Hauff, Edward Curry, and Philippe Cudré-Mauroux (Eds.). ACM, 1415–1424. https://doi.org/10.1145/3340531.3411968

[132] Jian Tang, Meng Qu, Mingzhe Wang, Ming Zhang, Jun Yan, and Qiaozhu Mei. 2015. LINE: Large-scale Information Network Embedding. In *Proceedings of the 24th International Conference on World Wide Web, WWW 2015, Florence, Italy, May 18-22, 2015*. 1067–1077.

[133] Erico Tjoa and Cuntai Guan. 2020. A survey on explainable artificial intelligence (xai): Toward medical xai. *IEEE transactions on neural networks and learning systems* 32, 11 (2020), 4793–4813.

[134] Hanghang Tong and Ching-Yung Lin. 2011. Non-negative residual matrix factorization with application to graph anomaly detection. In *Proceedings of the 2011 SIAM International Conference on Data Mining*. SIAM, 143–153.

[135] Jason Tumlinson. 2011. Astronomy: A new spin on the first stars. *Nature* 472, 7344 (2011), 426.

[136] Aaron Tuor, Samuel Kaplan, Brian Hutchinson, Nicole Nichols, and Sean Robinson. 2017. Deep learning for unsupervised insider threat detection in structured cybersecurity data streams. In *Workshops at the Thirty-First AAAI Conference on Artificial Intelligence*.

[137] Pavan Vatturi and Weng-Keen Wong. 2009. Category detection using hierarchical mean shift. In *Proceedings of the 15th ACM SIGKDD international conference on Knowledge discovery and data mining*. ACM, 847–856.

[138] Vincent Vercruyssen, Wannes Meert, and Jesse Davis. 2020. Transfer Learning for Anomaly Detection through Localized and Unsupervised Instance Selection. In *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 6054–6061. https://aaai.org/ojs/index.php/AAAI/article/view/6068

[139] Nikil Wale, Ian A. Watson, and George Karypis. 2008. Comparison of descriptor spaces for chemical compound retrieval and classification. *Knowl. Inf. Syst.* 14, 3 (2008), 347–375. https://doi.org/10.1007/s10115-007-0103-5

[140] Zhen Wang and Chao Lan. 2020. Towards a Hierarchical Bayesian Model of Multi-View Anomaly Detection. In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI 2020*, Christian Bessiere (Ed.). ijcai.org, 2420–2426. https://doi.org/10.24963/ijcai.2020/335

[141] Steve Webb, James Caverlee, and Calton Pu. 2006. Introducing the Webb Spam Corpus: Using Email Spam to Identify Web Spam Automatically. In *CEAS 2006 - The Third Conference on Email and Anti-Spam, July 27-28, 2006, Mountain View, California, USA*. http://www.ceas.cc/2006/listabs.html#6.pdf

[142] James D Wilson, Nathaniel T Stevens, and William H Woodall. 2016. Modeling and estimating change in temporal networks via a dynamic degree corrected stochastic block model. *arXiv preprint arXiv:1605.04049* (2016).

[143] Gang Wu and Edward Y Chang. 2003. Adaptive feature-space conformal transformation for imbalanced-data learning. In *Proceedings of the 20th International Conference on Machine Learning (ICML-03)*. 816–823.

[144] Jun Wu, Jingrui He, and Yongming Liu. 2018. ImVerde: Vertex-diminished random walk for learning imbalanced network representation. In *2018 IEEE International Conference on Big Data (Big Data)*. IEEE, 871–880.

[145] Shuang Wu, Jingyu Zhao, and Guangjian Tian. 2022. Understanding and Mitigating Data Contamination in Deep Anomaly Detection: A Kernel-based Approach. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 2319–2325. https://doi.org/10.24963/ijcai.2022/322

[146] Dongbo Xi, Bowen Song, Fuzhen Zhuang, Yongchun Zhu, Shuai Chen, Tianyi Zhang, Yuan Qi, and Qing He. 2021. Modeling the Field Value Variations and Field Interactions Simultaneously for Fraud Detection. In *Thirty-Fifth AAAI Conference on Artificial Intelligence, AAAI 2021, Thirty-Third Conference on Innovative Applications of Artificial Intelligence, IAAI 2021, The Eleventh Symposium on Educational Advances in Artificial Intelligence, EAAI 2021, Virtual Event, February 2-9, 2021*. AAAI Press, 14957–14965. https://ojs.aaai.org/index.php/AAAI/article/view/17755

[147] Wei Xu, Ling Huang, Armando Fox, David A. Patterson, and Michael I. Jordan. 2009. Online System Problem Detection by Mining Patterns of Console Logs. In *ICDM 2009, The Ninth IEEE International Conference on Data Mining, Miami, Florida, USA, 6-9 December 2009*, Wei Wang, Hillol Kargupta, Sanjay Ranka, Philip S. Yu, and Xindong Wu (Eds.). IEEE Computer Society, 588–597. https://doi.org/10.1109/ICDM.2009.19

[148] Pei Yang, Jingrui He, and Jia-Yu Pan. 2015. Learning complex rare categories with dual heterogeneity. In *Proceedings of the 2015 SIAM International Conference on Data Mining*. SIAM, 523–531.

[149] Pei Yang, Hongxia Yang, Haoda Fu, Dawei Zhou, Jieping Ye, Theodoros Lappas, and Jingrui He. 2015. Jointly Modeling Label and Feature Heterogeneity in Medical Informatics. *ACM Transactions on Knowledge Discovery from Data (TKDD-2015)* (2015).

[150] Zhilin Yang, William W. Cohen, and Ruslan Salakhutdinov. 2016. Revisiting Semi-Supervised Learning with Graph Embeddings. In *Proceedings of the 33nd International Conference on Machine Learning, ICML 2016, New York City, NY, USA, June 19-24, 2016*. 40–48.

[151] Yasser Yasami and Farshad Safaei. 2017. A statistical infinite feature cascade-based approach to anomaly detection for dynamic social networks. *Computer Communications* 100 (2017), 52–64.

[152] Chin-Chia Michael Yeh, Yan Zhu, Liudmila Ulanova, Nurjahan Begum, Yifei Ding, Hoang Anh Dau, Diego Furtado Silva, Abdullah Mueen, and Eamonn Keogh. 2016. Matrix profile I: all pairs similarity joins for time series: a unifying view that includes motifs, discords and shapelets. In *2016 IEEE 16th international conference on data mining (ICDM)*. IEEE, 1317–1322.

[153] Hao Yin, Austin R Benson, Jure Leskovec, and David F Gleich. 2017. Local higher-order graph clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 555–564.

[154] Hao Yin, Austin R. Benson, Jure Leskovec, and David F. Gleich. 2017. Local Higher-Order Graph Clustering. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, August 13 - 17, 2017*. ACM, 555–564. https://doi.org/10.1145/3097983.3098069

[155] Minji Yoon, Bryan Hooi, Kijung Shin, and Christos Faloutsos. 2019. Fast and Accurate Anomaly Detection in Dynamic Graphs with a Two-Pronged Approach. In *Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 647–657.

[156] Bing Yu, Haoteng Yin, and Zhanxing Zhu. 2018. Spatio-Temporal Graph Convolutional Networks: A Deep Learning Framework for Traffic Forecasting. In *Proceedings of the Twenty-Seventh International Joint Conference on Artificial Intelligence, IJCAI 2018, July 13-19, 2018, Stockholm, Sweden*, Jérôme Lang (Ed.). ijcai.org, 3634–3640. https://doi.org/10.24963/ijcai.2018/505

[157] Wenchao Yu, Wei Cheng, Charu C Aggarwal, Kai Zhang, Haifeng Chen, and Wei Wang. 2018. Netwalk: A flexible deep embedding approach for anomaly detection in dynamic networks. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2672–2681.

[158] Changqing Zhang, Ehsan Adeli, Tao Zhou, Xiaobo Chen, and Dinggang Shen. 2018. Multi-Layer Multi-View Classification for Alzheimer's Disease Diagnosis. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence, (AAAI-18), the 30th innovative Applications of Artificial Intelligence (IAAI-18), and the 8th AAAI Symposium on Educational Advances in Artificial Intelligence (EAAI-18), New Orleans, Louisiana, USA, February 2-7, 2018*, Sheila A. McIlraith and Kilian Q. Weinberger (Eds.). AAAI Press, 4406–4413. https://www.aaai.org/ocs/index.php/AAAI/AAAI18/paper/view/16864

[159] Guoxi Zhang, Tomoharu Iwata, and Hisashi Kashima. 2017. Robust Multi-view Topic Modeling by Incorporating Detecting Anomalies. In *Joint European Conference on Machine Learning and Knowledge Discovery in Databases*. Springer, 238–250.

[160] Si Zhang, Dawei Zhou, Mehmet Yigit Yildirim, Scott Alcorn, Jingrui He, Hasan Davulcu, and Hanghang Tong. 2017. HiDDen: hierarchical dense subgraph detection with application to financial fraud detection. In *Proceedings of the 2017 SIAM International Conference on Data Mining*. Society for Industrial and Applied Mathematics, 570–578.

[161] Weiqi Zhang, Chen Zhang, and Fugee Tsung. 2022. GRELEN: Multivariate Time Series Anomaly Detection from the Perspective of Graph Relational Learning. In *Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI 2022, Vienna, Austria, 23-29 July 2022*, Luc De Raedt (Ed.). ijcai.org, 2390–2397. https://doi.org/10.24963/ijcai.2022/332

[162] Handong Zhao and Yun Fu. 2015. Dual-Regularized Multi-View Outlier Detection.. In *IJCAI*. 4077–4083.

[163] Ying Zhao, Junjun Chen, Di Wu, Jian Teng, and Shui Yu. 2019. Multi-Task Network Anomaly Detection using Federated Learning. In *Proceedings of the Tenth International Symposium on Information and Communication Technology, Ha Noi, Ha Long Bay, Vietnam, December 4-6, 2019*. ACM, 273–279. https://doi.org/10.1145/3368926.3369705

[164] Jia-Xing Zhong, Nannan Li, Weijie Kong, Shan Liu, Thomas H Li, and Ge Li. 2019. Graph Convolutional Label Noise Cleaner: Train a Plug-and-play Action Classifier for Anomaly Detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 1237–1246.

[165] Dawei Zhou, Jingrui He, K. Seluk Candan, and Hasan Davulcu. 2015. MUVIR: Multi-View Rare Category Detection. In *The 24th International Joint Conference on Artificial Intelligence (IJCAI-15)*.

[166] Dawei Zhou, Jingrui He, Yu Cao, and Jae-sun Seo. 2016. Bi-level Rare Temporal Pattern Detection. In *IEEE International Conference on Data Mining series (ICDM2016)*.

[167] Dawei Zhou, Jingrui He, Hasan Davulcu, and Ross Maciejewski. 2018. Motif-Preserving Dynamic Local Graph Cut. In *IEEE International Conference on Big Data, Big Data 2018, Seattle, WA, USA, December 10-13, 2018*, Naoki Abe, Huan Liu, Calton Pu, Xiaohua Hu, Nesreen K. Ahmed, Mu Qiao, Yang Song, Donald Kossmann, Bing Liu, Kisung Lee, Jiliang Tang, Jingrui He, and Jeffrey S. Saltz (Eds.). IEEE, 1156–1161. https://doi.org/10.1109/BigData.2018.8622263

[168] Dawei Zhou, Jingrui He, Hongxia Yang, and Wei Fan. 2018. Sparc: Self-paced network representation for few-shot rare category characterization. In *Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining*. ACM, 2807–2816.

[169] Dawei Zhou, Arun Karthikeyan, Kangyang Wang, Nan Cao, and Jingrui He. 2017. Discovering rare categories from graph streams. *Data Mining and Knowledge Discovery* 31, 2 (2017), 400–423.

[170] Dawei Zhou, Kangyang Wang, Nan Cao, and Jingrui He. 2015. Rare Category Detection on Time-Evolving Graphs. In *IEEE International Conference on Data Mining (ICDM2015)*. IEEE.

[171] Dengyong Zhou, Jason Weston, Arthur Gretton, Olivier Bousquet, and Bernhard Schölkopf. 2004. Ranking on data manifolds. In *Advances in neural information processing systems*. 169–176.

[172] Dawei Zhou, Si Zhang, Mehmet Yigit Yildirim, Scott Alcorn, Hanghang Tong, Hasan Davulcu, and Jingrui He. 2017. A local algorithm for structure-preserving graph cut. In *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 655–664.

[173] Dawei Zhou, Lecheng Zheng, Jiawei Han, and Jingrui He. 2020. A Data-Driven Graph Generative Model for Temporal Interaction Networks. In *KDD '20: The 26th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, Virtual Event, CA, USA, August 23-27, 2020*, Rajesh Gupta, Yan Liu, Jiliang Tang, and B. Aditya Prakash (Eds.). ACM, 401–411. https://doi.org/10.1145/3394486.3403082

[174] Zhi-Hua Zhou. 2018. A brief introduction to weakly supervised learning. *National science review* 5, 1 (2018), 44–53.

[175] Arthur Zimek, Erich Schubert, and Hans-Peter Kriegel. 2012. A survey on unsupervised outlier detection in high-dimensional numerical data. *Statistical Analysis and Data Mining: The ASA Data Science Journal* 5, 5 (2012), 363–387.