Front. Comput. Sci., 2022, 16(2): 162701
https://doi.org/10.1007/s11704-020-0013-1

RESEARCH ARTICLE

iNet: visual analysis of irregular transition in multivariate dynamic networks

Dongming HAN[1], **Jiacheng PAN**[1], **Rusheng PAN**[1], **Dawei ZHOU**[2], **Nan CAO**[3], **Jingrui HE**[2], **Mingliang XU**[4], **Wei CHEN** (✉)[1]

1 State Key Lab of CAD & CG, Zhejiang University, Hangzhou 310058, China
2 Department of Computer Science and Engineering, Arizona State University, Arizona 85287, America
3 Tong Ji Intelligent Big Data Visualisation Lab (iDVx Lab), TongJi University, Shanghai 200082, China
4 Department of Computer Science and Technology, Zhengzhou University, Zhengzhou 450001, China

© Higher Education Press 2022

Abstract Multivariate dynamic networks indicate networks whose topology structure and vertex attributes are evolving along time. They are common in multimedia applications. Anomaly detection is one of the essential tasks in analyzing these networks though it is not well addressed. In this paper, we combine a rare category detection method and visualization techniques to help users to identify and analyze anomalies in multivariate dynamic networks. We conclude features of rare categories and two types of anomalies of rare categories. Then we present a novel rare category detection method, called DIRAD, to detect rare category candidates with anomalies. We develop a prototype system called iNet, which integrates two major visualization components, including a glyph-based rare category identifier, which helps users to identify rare categories among detected substructures, a major view, which assists users to analyze and interpret the anomalies of rare categories in network topology and vertex attributes. Evaluations, including an algorithm performance evaluation, a case study, and a user study, are conducted to test the effectiveness of proposed methods.

Keywords multivariate dynamic networks, rare categories, anomaly detection, visual analysis

1 Introduction

Multivariate dynamic networks refer to the time-evolving graphs with multiple attributes on each vertex [1]. Data in this type commonly exist in many real-world applications. Examples include the dynamic digital communication network and the network of multimedia among various computers. The innate capability of capturing complex relationships makes dynamic network analysis and visualization a hot research topic in the past decades [2]. However, existing techniques are mainly focused on showing the evolution trend of the dynamic network. Little research pays attention to the anomalous change of substructure, i.e., irregular transitions inside a dynamic network. These changes could be anomalous interactions among people inside a social network or a potential money laundry transaction among different accounts. Therefore, a technique for detecting and interpreting anomalous changes of substructures inside a dynamic network is desired.

The above problem lies in the domain of anomaly detection, which has been extensively studied during the past decades [3]. Especially, a recent survey [4] suggests that more and more attentions have been put on detecting anomalies inside a dynamic network. However, most techniques focused mainly on individual anomalies, i.e., how a vertex or an edge is suddenly changed or emerged, instead of capturing the changing of substructures. To combat this issue, Zhou et al. [5] introduced an algorithm, namely BIRD, for detecting the rare categories inside a time-varying network with a fixed number of vertices. It is an active learning algorithm, which finds a representative vertex in a potential minority class and requires an oracle to make a justification on the vertex. The correctness of this process highly relies on the oracle's understanding of the analysis results, where visual analytic techniques can be helpful.

However, designing a visual analysis system to support the detection and interpretation of substructures in a network that change anomalously over time is challenge [6]. First, design an algorithm to find the substructures that change anomalously based on both vertex attributes and graph topology is complicated. Second, illustrating the change of dynamic multivariate graph requires an integrated visualization for showing the change of both vertex attributes and the underlying topology, which is difficult.

In this paper, we introduced a visual analysis system, iNet, to address the above challenges for detecting irregular transitions in a multivariate dynamic network. In particular, we developed a novel rare category detection algorithm that detects anomalous substructures based on both topological information and vertex attributes. A novel visualization is also designed based on the matrix for illustrating the change of a

Received January 7, 2020; accepted September 1, 2020

E-mail: chenvis@zju.edu.cn

dynamic multivariate graph as well as the rare categories detected from the algorithm to facilitate interpretation. To the best of our knowledge, this is the first visual analysis system designed for detecting anomalous changing patterns in a multivariate dynamic network. In particular, this paper has the following contributions:

- **System** We introduce the first, to the best of our knowledge, a visual analysis system for detecting and interpreting the anomalous change of substructures in a multivariate dynamic network. The system supports both substructure identification as well as analysis results interpretation.
- **Algorithm** We introduce a novel rare category detection algorithm, DIRAD, for detecting minority classes (i.e., the potential irregular transition of substructures) inside a multivariate dynamic graph. The algorithm handles a fully dynamic network (i.e., both the numbers of vertices and edges vary from time to time) and takes both the change of the network topology and the vertex attributes into consideration.
- **Visualization** We propose multiple visualization views embedded into a matrix-flow based dynamic network visualization design to facilitate the interpretation and comparison of the change of the network from different perspectives.

2 Related work

In this section, we review the most relevant techniques in aspects of both algorithms and visualization.

2.1 Rare category detection

Rare category detection refers to a series of active learning methods that detect samples of minority classes and let users label these samples in un-labeled data. The first attempt in this area is from Pelleg and Moore. [7], who designed a mixture model-based algorithm to detect examples of rare categories in datasets. A series of methods are then presented based on different prior information, compactness assumption, and rare category schema [8]. Some RCD methods require some prior information about the dataset, such as proportion of each rare category, to detect minority classes [7,9–13]. For datasets with no prior information, researchers also developed a series methods [8,14]. Some RCD methods assume that rare categories distribute smoothly and compactly in the major categories [15], while other RCD methods require rare categories isolate from the major category [14]. Based on the rare category schema, RCD methods can be classified into three categories: model-based methods, which find rare categories that do not fit in statistical models [7,16], neighbor-based methods, which find rare categories with abnormal local density changes [14,17], and hierarchical-based methods, which find rare categories with clustering results [18]. Based on previous work, Pan et al. [19] proposed a visual analysis system called RCAnalyzer to detect rare changes of substructures in a dynamic network.

In this paper, we propose a method to detect rare categories in multivariate dynamic networks, which models various fine-grained dynamics of networks, e.g., vertices/edges are added/removed and incorporates both the network topology and multivariate attributes of vertices to find the vertices that are possibly belonging to target minority classes.

2.2 Visualization of anomaly

Statistical diagrams are most simple method to display the outliers [20,21]. However, they do not work well on more complex datasets. For high dimensional data, principal component analysis and multidimensional scaling are used to reduce the dimension of data and anomalies can be found in Low-dimensional space [22]. Some visualization techniques can directly visualize the high dimensional data, such as parallel coordinate plots [23] and DICON [24]. Abnormal distributions can be directly observed through these methods. For anomaly in time series data[25], ViDx [26] extends Marey’s graph to show outliers in manufacturing procedures. Machine learning methods, such as neural networks and visualizations, are combined to detect intrusions in network traffic data [27,28]. In social media data, a series of methods are presented to detect anomalies with the help of visualization techniques [29–31]. Fluxflow [30] detects the diffusion of anomalous information in social media, and TargetVue [31] utilizes glyph-based designs to show the anomalous behaviors in online communication systems based on an unsupervised learning model.

In this paper, we focus on detecting anomalies in multivariate dynamic networks with rare category detection methods, which is not studied in previous literature. Moreover, we developed a series of visualizations and interactions to help users to not only identify rare categories in data but also analyze and interpret the anomalies of these substructures.

2.3 Visualization of dynamic networks

Visualization of dynamic networks is well-studied in the past years as this data form is becoming more and more common these years. An excellent survey by Beck et al. [2] has reported the state of art of dynamic network visualizations. Beck et al. classify the visualization techniques of dynamic networks into animated diagrams [32–34] and timeline of a series of static charts, such as node-link diagram or adjacency matrix. Among these various dynamic network visualization techniques, timeline with matrix-based and flow-based representation methods are most relevant to our work.

Matrix-based techniques can be classified into two categories. The first category embeds timeline into each cell of the matrix. Visualization techniques used in the cell vary from each other, depending on the analysis tasks. Gstaltlines [35], fingerprint glyphs [36], and horizon graph [37] are used to show the evolving of dyadic relations in a matrix. The time-evolving patterns of dyadic relations are clearly shown inside each cell. However, this category of methods often does not fit well to large data sets. The second category lays a sequence of adjacency matrices in a certain order [38–41], such as Matrix Cube [39], which stacks the matrices together and visualizes them in a three-dimensional space, and MatrixWave [40], which lays a series of matrices in a zig-zag shape to visualize the transition patterns among vertices. More recently, van den Elzen et al. [42] reduce the matrices into points and lay the point by production methods.

Flow-based techniques use flow metaphors to represent the evolving of communities in networks [43−45]. Sankey diagram and Themeriver are the most common methods used. For example, Vehlow et al. [43] use sankey diagrams to show the changes in community structures. Flow-based techniques aggregate networks by group information, and thus often lack details of the local areas of the network. To reduce visual clutter in Massive Sequence View(MSV), Ying el al. [46] proposes an edge sampling method, using the edge overlapping degree (EOD) concept while preserving the time-varying features of network communication.

In this paper, we first use a glyph based design to help users to identify rare categories among the substructures detected by the RCD algorithm. Then, we design a time-line based visualization, which integrates matrix representation, node-link diagram, bar charts, and sankey diagram to help users to analyze and interpret the anomalies of identified anomalies.

3 System overview

iNet system aims to support users analyzing irregular transitions in multivariate dynamic networks. We collaborated with two domain experts in rare category detection for about ten months. Weekly meetings with experts were held to discuss requirements, problems, and possible solutions. The requirements are summarized as follows:

R1 Finding substructures instead of individual outliers that change anomalously over time The system should be able to find the anomalous connections instead of single isolated points to reveal more relational insights of the network.

R2 Supporting the analysis of a fully dynamic network The system should be able to capture the irregular changing patterns of network scales (i.e., number of vertices), topology, and the vertex attributes at the same time.

R3 Identification and interpretation in context The system should be able to put the analysis results in context to support interpretation and comparison.

With the above requirements in mind, we designed the architecture of iNet system, as shown in Fig. 1. The system consists of four modules: 1) the data storage module; 2) the analysis module; 3) the data post-processing module; 4) the visualization module. The data storage module transforms raw data to multivariate dynamic network data, stores processed data with Neo4j, and extracts topology features of vertices. The analysis module integrates a new rare category detection algorithm, namely DIRAD, which for detecting irregular transition patterns in a fully dynamic network (**R1**). In this module, DIRAD first detects representative vertices in categories possibly with irregular transition based on the topology and vertex attributes (**R2**), then a local clustering algorithm is performed to find the border of rare categories. The post-processing module first calculates the layout of the rare category based on the feature in the storage module, and then extracts and calculates contextual information of detected categories, including features of a rare category, vertex attributes in the rare category, the topology of the rare category, and similarities among a rare category and other sub-networks (**R3**). The visualization module visualizes rare categories with a rare category identifier, a topology explorer, and an attribute explorer (**R2**). A series of user interactions are provided in each view to help users make decisions (**R3**).

4 Irregular transition detection

In this section, we introduce the algorithms for irregular changing detection in a multivariate dynamic network. In particular, we introduced a novel rare category detection technique based on the state-of-the-art algorithm introduced by Zhou et al. [10], which simultaneously takes the changes of network scale, topology, and attributes into consideration. The algorithm provides a representative sample vertex for each detected rare category, based on which the boundary of the category is further determined via the checking of the change of the local density and the compactness of the structure.

4.1 Notation

In this paper, we use lowercase letters to denote scalars, boldface lowercase letters to denote vectors, and boldface uppercase letters to denote matrices. Also, we represent element-wise entries of a matrix using a convention similar to the Matlab, e.g., $M(i,j)$ is the element at the ith row and jth column of the matrix $\boldsymbol{M}$, and $M(i,:)$ is the ith row of $\boldsymbol{M}$, etc.

Suppose we are given a series of time-evolving graphs $\{\boldsymbol{G}^{(1)},\boldsymbol{G}^{(2)},\ldots,\boldsymbol{G}^{(T)}\}$. At each time stamp t, the graph $\boldsymbol{G}^{(t)}$ contains $n^{(t)}$ vertices, which come from m distinct classes, i.e., $y_i \in \{1,\ldots,m\}$. We use the notion aggregated adjacency matrix, denoted by $\boldsymbol{M}^{(t)}$, for the adjacency matrix of $\boldsymbol{G}^{(t)}$. Furthermore, the set of vertices in $\boldsymbol{G}^{(t)}$ is described by multivariate features $\boldsymbol{X} = [\boldsymbol{x}_1,\boldsymbol{x}_2,\ldots,\boldsymbol{x}_{n^{(t)}}]^{\mathrm{T}}$. Without loss of generality, we assume that most of the vertices belong to the majority class with prior $p_1^{(t)}$, and the remaining classes are minority classes with prior

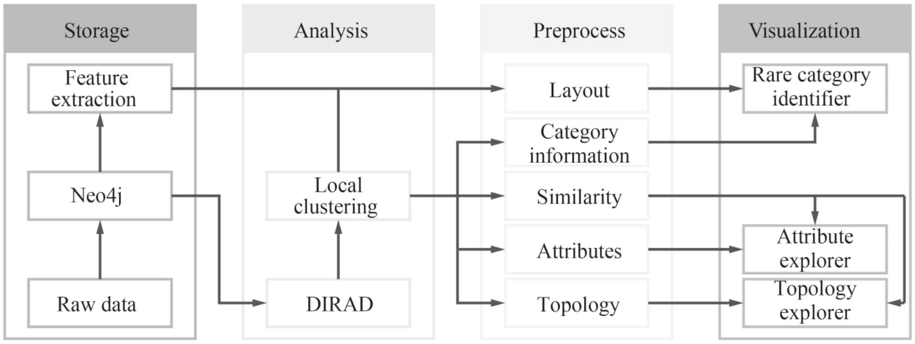


Fig. 1 System pipeline

$p_c^{(t)}, c = 2, \ldots, m$. In our studied problem, we aim to identify at least one example from each minority class (rare category), by repeatedly selecting examples to be labeled by an oracle.

4.2 Preliminaries

Now, we introduce the basics of BIRD algorithm [10], an active rare category detection method designed for time-evolving graphs. A key challenge associated with identifying rare categories in dynamic networks is the expensive computational cost. To address this issue, the BIRD algorithm incrementally updates the RCD models by incorporating the local changes (i.e., the added/deleted edges) instead of reconstructing it from scratch on the updated data at a new time stamp. In general, the BIRD algorithm can be mainly decomposed into the following steps:

S1 **Updating:** Update the global similarity matrix $\boldsymbol{A}^{(t)}$ [47] at new time stamp t from previous global similarity matrix $\boldsymbol{A}^{(t-1)}$ and updated edges at previous time stamp $(t-1)$ by the Sherman-Morrison formula:

$$\boldsymbol{A}^{(t)} = \boldsymbol{A}^{(t-1)} + \alpha \frac{\boldsymbol{A}^{(t-1)} u v^{\mathrm{T}} \boldsymbol{A}^{(t-1)}}{\boldsymbol{I} + v^{\mathrm{T}} \boldsymbol{A}^{(t-1)} u},$$

where u and v are indicator vectors to represent the updated edges, and $\alpha \in (0,1)$.

S2 **Embedding:** Compute the class-oriented neighborhood embedding $\boldsymbol{NN}^{(t)}$ by choosing the vertices with drastic evolution in their neighborhood, and locally update the corresponding rows in the neighborhood embedding $\boldsymbol{NN}^{(t-1)}$ at previous time stamp.

S3 **Query:** Calculate the score for each example and deliver the unlabeled examples with the largest score to oracle until the rare categories of the users' interest are identified.

However, the existing dynamic RCD techniques (e.g., BIRD) [10,15] have multiple limitations: (1) they assume that the number of vertices is fixed over time, while it is usually the case that the vertices may appear, vanish, or reappear in the real-world networks such as social networks, collaboration networks, and online transaction networks; (2) they cannot incorporate the rich attribute information in multivariate dynamic networks, which may be crucial for oracle to judge whether a given example belongs to certain rare categories or not. In the next subsection, we show the details of our proposed dual-view incremental rare category detection (DIRAD) framework that could better model the richness of rare category evolution and utilize the external information from multivariate dynamic networks for more accurately detecting the abnormal patterns.

4.3 Dynamic network updating

In real dynamic systems, it is usually the case that the size of the underlying networks may change over time, i.e., the networks are incremental or decremental. The existing RCD techniques [10,15] fail to model such dynamics, as the global similarity matrix $\boldsymbol{A}^{(t)}$ cannot be directly updated from the previous time stamp. Instead of recomputing the global similarity matrix $(O((n^{(t)})^3))$ at each new time stamp t, our methods locally update the global similarity matrix from the previous time stamp.

Incremental network updating Suppose that $k \ll n^{(t-1)}$ new vertices are added to the existing network $\boldsymbol{G}^{(t-1)}$ at time stamp t, the new adjacency matrix $\boldsymbol{M}^{(t)}$ can be represented as

$$\begin{bmatrix} \boldsymbol{M}_{11}^{(t)} & \boldsymbol{M}_{12}^{(t)} \\ \boldsymbol{M}_{21}^{(t)} & \boldsymbol{M}_{22}^{(t)} \end{bmatrix},$$

where $\boldsymbol{M}_{11}^{(t)} = \boldsymbol{M}^{(t-1)}$ preserves the connectivity information among the original set of vertices, $\boldsymbol{M}_{22}^{(t)}$ preserves the connectivity information among the newly added vertices, and $\boldsymbol{M}_{12}^{(t)}, \boldsymbol{M}_{21}^{(t)}$ preserves the connectivity information between the original set of vertices and the newly added set of vertices. Note that we assume the time-evolving graph is undirected, such that $\boldsymbol{M}_{12}^{(t)} = (\boldsymbol{M}_{21}^{(t)})^{\mathrm{T}}$. Here, we propose to infer the new global similarity matrix $\boldsymbol{A}^{(t)}$ from $\boldsymbol{A}^{(t-1)}$ and the changes in the adjacency matrix.

Based on the definition, the global similarity matrix can be computed as follows

$$\boldsymbol{A}^{(t)} = (\boldsymbol{I}_{n^{(t)} \times n^{(t)}} - \alpha \boldsymbol{M}^{(t)})^{-1},$$

Let $\boldsymbol{M}'^{(t)} = \boldsymbol{I}_{n^{(t)} \times n^{(t)}} - \alpha \boldsymbol{M}^{(t)}$, we have

$$\begin{aligned} \boldsymbol{A}^{(t)} &= (\boldsymbol{M}'^{(t)})^{-1} \\ &= \begin{bmatrix} \boldsymbol{I}_{n^{(t-1)} \times n^{(t-1)}} - \alpha \boldsymbol{M}_{11}^{(t)} & -\alpha \boldsymbol{M}_{12}^{(t)} \\ -\alpha \boldsymbol{M}_{21}^{(t)} & \boldsymbol{I}_{k \times k} - \alpha \boldsymbol{M}_{22}^{(t)} \end{bmatrix}^{-1} \\ &= \begin{bmatrix} \boldsymbol{M}'^{(t)}_{11} & \boldsymbol{M}'^{(t)}_{12} \\ \boldsymbol{M}'^{(t)}_{21} & \boldsymbol{M}'^{(t)}_{22} \end{bmatrix}^{-1}, \end{aligned} \tag{1}$$

Since $\boldsymbol{M}_{11}^{(t)} = \boldsymbol{M}^{(t-1)}$, we have

$$(\boldsymbol{M}'^{(t)}_{11})^{-1} = (\boldsymbol{I}_{n^{(t-1)} \times n^{(t-1)}} - \alpha \boldsymbol{M}^{(t-1)})^{-1} = \boldsymbol{A}^{(t-1)},$$

In addition, by adopting block matrix inversion lemma [48], we can rewrite Eq. (1) as follows

$$\boldsymbol{A}^{(t)} = \begin{bmatrix} \boldsymbol{A}_{11}^{(t)} & \boldsymbol{A}_{12}^{(t)} \\ \boldsymbol{A}_{21}^{(t)} & \boldsymbol{A}_{22}^{(t)} \end{bmatrix}, \tag{2}$$

where

$$\begin{aligned} \boldsymbol{A}_{11}^{(t)} &= \boldsymbol{A}^{(t-1)} + \boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12} (\boldsymbol{M}'^{(t)}_{22} - \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12})^{-1} \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)}, \\ \boldsymbol{A}_{12}^{(t)} &= -\boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12} (\boldsymbol{M}'^{(t)}_{22} - \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12})^{-1}, \\ \boldsymbol{A}_{21}^{(t)} &= -(\boldsymbol{M}'^{(t)}_{22} - \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12})^{-1} \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)}, \\ \boldsymbol{A}_{22}^{(t)} &= (\boldsymbol{M}'^{(t)}_{22} - \boldsymbol{M}'^{(t)}_{21} \boldsymbol{A}^{(t-1)} \boldsymbol{M}'^{(t)}_{12})^{-1}. \end{aligned}$$

Decremental network updating Suppose that $k \ll n^{(t-1)}$ new vertices are removed from the existing network $\boldsymbol{G}^{(t-1)}$ at time stamp t, we have

$$\boldsymbol{M}^{(t-1)} = \begin{bmatrix} \boldsymbol{M}_{11}^{(t-1)} & \boldsymbol{M}_{12}^{(t-1)} \\ \boldsymbol{M}_{21}^{(t-1)} & \boldsymbol{M}_{22}^{(t-1)} \end{bmatrix}, \tag{3}$$

$$\boldsymbol{M}^{(t)} = \boldsymbol{M}_{11}^{(t-1)}, \tag{4}$$

where $\boldsymbol{M}^{(t)} = \boldsymbol{M}_{11}^{(t-1)}$ represents the adjacency matrix of the preserved $n^{(t)}$ vertices, $\boldsymbol{M}_{22}^{(t-1)}$ represents the adjacency matrix of the removed k vertices, and $\boldsymbol{M}_{12}^{(t-1)} = (\boldsymbol{M}_{21}^{(t-1)})^{\mathrm{T}}$ preserves the connectivity information between the set of preserved vertices and the set of removed vertices at time $(t-1)$. Our target is to deduce the new global similarity matrix $\boldsymbol{A}^{(t)}$ based on $\boldsymbol{A}^{(t-1)}$ and the changes happened in time t.

Similar to Eq. (1) and Eq. (2), we let $\boldsymbol{M}'^{(t-1)} = \boldsymbol{I}_{n^{(t-1)}\times n^{(t-1)}} - \alpha\boldsymbol{M}^{(t-1)}$ and have

$$\begin{aligned}\boldsymbol{A}^{(t-1)} &= (\boldsymbol{M}'^{(t-1)})^{-1}\\ &= \begin{bmatrix}\boldsymbol{I}_{n^{(t)}\times n^{(t)}} - \alpha\boldsymbol{M}_{11}^{(t-1)} & -\alpha\boldsymbol{M}_{12}^{(t-1)}\\ -\alpha\boldsymbol{M}_{21}^{(t-1)} & \boldsymbol{I}_{k\times k} - \alpha\boldsymbol{M}_{22}^{(t-1)}\end{bmatrix}^{-1} \qquad (5)\\ &= \begin{bmatrix}\boldsymbol{M}'^{(t)}_{11} & \boldsymbol{M}'^{(t)}_{12}\\ \boldsymbol{M}'^{(t)}_{21} & \boldsymbol{M}'^{(t)}_{22}\end{bmatrix}^{-1}\\ &= \begin{bmatrix}\boldsymbol{A}_{11}^{(t-1)} & \boldsymbol{A}_{12}^{(t-1)}\\ \boldsymbol{A}_{21}^{(t-1)} & \boldsymbol{A}_{22}^{(t-1)}\end{bmatrix}, \qquad (6)\end{aligned}$$

where

$$\begin{aligned}\boldsymbol{A}_{11}^{(t-1)} =& \boldsymbol{A}^{(t-1)} + \boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12}(\boldsymbol{M}'^{(t-1)}_{22}\\ &- \boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12})^{-1}\boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)},\\ \boldsymbol{A}_{12}^{(t-1)} =& -\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12}(\boldsymbol{M}'^{(t-1)}_{22} - \boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12})^{-1},\\ \boldsymbol{A}_{21}^{(t-1)} =& -(\boldsymbol{M}'^{(t-1)}_{22} - \boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12})^{-1}\boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)},\\ \boldsymbol{A}_{22}^{(t-1)} =& (\boldsymbol{M}'^{(t-1)}_{22} - \boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12})^{-1}.\end{aligned}$$

Thus, we can easily compute $\boldsymbol{A}^{(t)}$ by

$$\boldsymbol{A}^{(t)} \approx -\boldsymbol{A}_{12}^{(t-1)}(\boldsymbol{M}'^{(t-1)}_{22} - \boldsymbol{M}'^{(t-1)}_{21}\boldsymbol{A}^{(t)}\boldsymbol{M}'^{(t-1)}_{12})(\boldsymbol{M}'^{(t-1)}_{12})^{+}, \qquad (7)$$

where $(\boldsymbol{M}'^{(t-1)}_{12})^{+}$ is denoted as a pseudoinverse of $\boldsymbol{M}'^{(t-1)}_{12}$.

4.4 Dual-view incremental rare category learning

Based on the discussion in the previous subsection, we can easily adopt the existing dynamic RCD methods to build the class-oriented neighborhood embedding $\boldsymbol{NN}^{(t)}$ using the updated global similarity matrix $\boldsymbol{A}^{(t)}$ and compute the query score for each vertex at each time stamp t [49]. However, in contrast to plain graphs where only pairwise vertices dependencies are observed, vertices in many complex systems also affiliate with a rich set of features. For example, in social networks, each user interacts and communicates with others and also posts personal profiles such as age, interest, and home location; in scientific collaboration networks, each researcher collaborates with others while also featuring his/her unique research interests. Thus, how to utilize such external information to enhance the performance of the RCD model further is crucial for our visualization system.

Recently, [50] proposed a multi-view rare category detection framework that integrates view-specific information to obtain the overall posterior probabilities of the vertices coming from rare categories. Here, we extend this idea to jointly learn the overall posterior probabilities $s_{overall}^{(t)}(v_i)$ of each vertex v_i coming from rare categories which stand on both the global topology information of the network and the local multivariate attributes of each vertex as follows.

$$s_{overall}^{(t)}(v_i) = s_t^{(t)}(v_i)s_a^{(t)}(v_i)\left(\frac{P_t(v_i)P_a(v_i)}{P(v_i)}\right)^d, \qquad (8)$$

where $s_a^{(t)}(v_i)$ denotes the feature-oriented score obtained from the multivariate attributes at time stamp t using the existing techniques [51,52]; $s_t^{(t)}(v_i)$ denotes the topology-oriented score computed from network topology which is updated from the previous time stamp $(t-1)$; $P_t(v_i)$, $P_a(v_i)$ and $P(v_i)$ denote the marginal probabilities estimated from topology domain, feature domain, and both two domains using kernel density estimation (KDE) [53] at current time stamp; and $d \geqslant 0$ is a positive parameter that controls the impact regarding the marginal probability between topology domain and feature domain.

In Algorithm 1, we describe a fast algorithm - DIRAD, that (1) models various fine-grained dynamics of temporal networks (e.g., vertices/edges are added/removed), and (2) incorporates the information of both global network topology and local multivariate attributes to estimate the overall probability of each vertex belonging to the target minority class. It works as follows. First, we update the global similarity $\boldsymbol{A}^{(t)}$ based on the $\boldsymbol{A}^{(t-1)}$ and the changes (e.g., vertex/edges are added/removed from the network) that happened at current time stamp t. Then, in Step 2, we utilize BIRD algorithm and the updated global similarity $\boldsymbol{A}^{(t)}$ to calculate the initial topology-oriented scores $\boldsymbol{s}_t^{(t)}$ for all the vertices. Next, in Step 3, we estimate the initial feature-oriented scores $\boldsymbol{s}_a^{(t)}$ of all the vertices, which can be done using any existing techniques for rare category detection. Finally, Step 4 to Step 12 aims to select the vertices that are showing at the current time stamp t with the largest posterior probabilities coming from rare categories to the oracle. In particular, for each class c, Step 6 to Step 7 iteratively update the feature-oriented scores $\boldsymbol{s}_a^{(t)}$ and topology-oriented scores $\boldsymbol{s}_t^{(t)}$. Step 8 calculates the overall probability of each vertex belonging to the minority class c and then delivers the vertex with maximum of $s^{(t)}(v_i)$ to the oracle in Step 9. In Step 10, if the labeled vertex is from minority class c, we break the inner loop; otherwise, we mark the class of this vertex as labeled.

Algorithm 1 DIRAD algorithm

```
Require: $\boldsymbol{A}^{(t-1)}$, $\boldsymbol{M}^{(t-1)}$, $\boldsymbol{M}^{(t)}$, $\boldsymbol{X}^{(t)}$, $p_1^{(1)}, \ldots, p_c^{(m)}$, $\alpha$, $\epsilon$.
Ensure: The set I of the selected examples and the set L of their labels.
1: Update the global similarity matrix $\boldsymbol{A}^{(t)}$ from $\boldsymbol{A}^{(t-1)}$ based on the changes happened at time stamp $t$ using Eq. (3) and Eq. (7).
2: Compute the topology-oriented scores $\boldsymbol{s}_t^{(t)}$ for all the vertices based on BIRD [10] algorithm and the updated global similarity matrix $\boldsymbol{A}^{(t)}$.
3: Compute the feature-oriented scores $\boldsymbol{s}_a^{(t)}$ for all the vertices at current time stamp using the existing techniques for rare category detection, such as NNDB [51] and GRADE [52].
4: for all $c$=2 : $m$ do
5:     while class $c$ is not discovered do
6:         For each vertex $v_i$ that has been labeled by the oracle, $\|v_i, v_j\|_2 \leqslant \epsilon$, then $s^{(t)}(v_j) = -\infty$.
7:         Update the $s_t^{(t)}(v_i)$, $s_a^{(t)}(v_i)$ using the existing techniques such as GRADE [52].
8:         Compute the overall score for each vertex $s^{(t)}(v_i)$ based on Eq. (8).
9:         Query the label of the vertex with the maximum of $s^{(t)}(v_i)$.
10:        If the label of $v_i$ is from class $c$, break; otherwise, mark the class of $v_i$ as labeled.
11:    end while
12: end for
```

4.5 Finding substructures

The rare category should be found after its representative vertex is detected by DIRAD. In a multivariate dynamic network, a rare category has following characteristics:

- **Compact** In multivariate dynamic networks, compactness is the basic requirement for a category being rare. In topology, this character indicates that vertices in a rare category form a more compact sub-network in some local areas of the networks. In attribute, this character indicates that vertices in a rare category have consistent attribute distributions. In this paper, a group of vertices can be identified as a rare category if they are compact in the topology dimension or attribute.
- **Bordered** A compact group of vertices sometimes is not a rare category, for example, a sub-network extracted from a large clique network is compact, but the sub-network cannot be identified as a rare category. Thus, based on compact character, rare categories in the multivariate dynamic network have to be bordered at the same time, that is, vertices inside a rare category can be distinguished from the vertices outside the rare category. In topology, bordered character indicates that vertices in a rare category have more internal connections than external connections. In attribute, bordered character indicates that vertices in a rare category have different attribute distribution from vertices outside the rare category.
- **Size-sensitive** The size of a category matters when users determine the rareness of the category. Categories with the different number of vertices are not comparable even if they have the same topology structure or attributes. For example, a clique structure with 10 vertices is not rare, while a clique structure with 300 vertices is rare when the network is spliced by thousands of 10 vertices cliques and only one 300 vertices clique.

Based on these characteristics, we regard the vertex detected by DIRAD as a seed and use a local clustering algorithm [54] to detect the substructure that satisfies the characteristics around the seed. The maximum size of the cluster is a parameter requiring to be artificially set in the algorithm. We provide a user interface to enable users to set a series of maximum sizes manually to detect rare categories of different sizes.

4.6 Interpreting anomaly in context

DIRAD computes an anomaly score for each vertex in the network at a specific time. However, it is hard for analysts to interpret or trust the existence of the anomaly with a single number. Thus, iNet provides a series of contextual information of vertices with high scores to help analysts interpret and understand the irregular transition in multivariate dynamic networks.

The first contextual information that iNet should provide is the features of detected substructures. In each query, iNet returns a batch of substructures to accelerate the analysis process. With features of detected substructures, analysts are able to identify rare categories from a series of substructures and select the rare category they interested in.

Topology, vertex attributes, and time-evolving pattern of a substructure are essential contextual information for analysts to interpret anomalies of substructures. These data describe the characteristics of substructures from the relation pattern, attribute features, and temporal changes, respectively. Thus, once a substructure is detected at a specific time stamp, its topology and attributes at every time stamp are extracted from the data storage module, allowing analysts to identify anomalies potential anomalous topology structure, attribute distribution, and evolving pattern.

The other worth investigating contextual information is the similarities between detected substructures and other substructures in topology and attribute dimensions. It is hard to determine whether a detect substructure is abnormal or not at a specific time stamp without knowing how different the substructure is from other substructures in the network. Considering that sizes of detected substructures are unpredictable, finding major substructures with a certain size in the network is not feasible. In this work, we adopt an approximate strategy to describe the possibility of a substructure being abnormal: we calculate the similarity among the detected substructure and other substructures that have the same number of vertices as the detected substructure in topology and attribute at each time stamp.

With the above data, we design the visualization in iNet to represent the following information computed in the post process module: 1) substructures detected by DIRAD and the local clustering algorithm; 2) topology, attributes, and time-evolving pattern of the substructures; and 3) the distribution of the similarities among the detected substructures and other substructures in the network.

5 Rare category explorer

In this section, we first summarize the design tasks based on the requirements, the characteristics of rare categories, and anomalies in multivariate dynamic networks. Second, we introduce the user interface of iNet.

5.1 Design tasks

Before summarizing the design tasks of iNet, we first introduce two possible types of the anomaly of rare categories in multivariate dynamic networks: an anomaly in a snapshot and anomaly in network dynamics.

Anomaly in a snapshot refers to the case when the topology or attribute of a substructure is different from the most substructures at a specific time stamp. This anomaly usually means that the relations among the vertices or vertex characteristics in a substructure are special in the network, for example, a large and compact cluster is likely to be abnormal because the topology is usually sparse in a collaboration network formed by productive researchers. In this paper, we identify the existence of this anomaly by calculating the similarities among the rare category and other substructures in the network. If the rare category is similar to most of the substructures at a specific time stamp, it is not anomalous at the time stamp. Otherwise, it is anomalous. However, comparing a rare category with all substructures in a network is too complex (with 2^n possibilities, where n is the number of vertices in a network). As rare categories are size-sensitive, we only calculate the similarities among a rare category and substructures that have the same size as the rare category.

Anomaly in network dynamics refers to the case when the topology or attribute of a substructure significantly changes in a period of time. This anomaly means that the behavior of vertices in the substructure abruptly changed at some time stamps, for example, when a productive researcher, who seldom collaborates with other researchers before, begins to collaborate with others, the topology around him may significantly change. In this paper, we identify the existence of this anomaly by calculating the significance of change over time of the rare category with similarity measure. The visualization introduced in later chapters can also be used to identify this anomaly.

In the real application, the anomaly of a rare category might be a combination of above anomalies, for example, a rare category might be different from most substructures both in topology and attribute at multiple time stamps and significantly changes between some time stamps. Anomaly found by iNet strongly depends on datasets.

Based on the requirements mentioned in Section 3, the characteristics of rare categories in Section 4, and the possible anomalies in multivariate dynamic networks, we concluded a list of design tasks iNet should complete as follows.

T1 Showing the features of detected substructures iNet should provide an intuitive visualization of substructure features to help analysts to find substructures matching the characteristics of rare categories among detected substructures.

T2 Revealing the state of a substructure at each time stamp Understanding the topology and attribute pattern of substructures is the basis of identifying and interpreting anomalies of substructures. Thus the visualization should be able to present analysts with topology structure and attributes of vertices of the substructure at each time stamp.

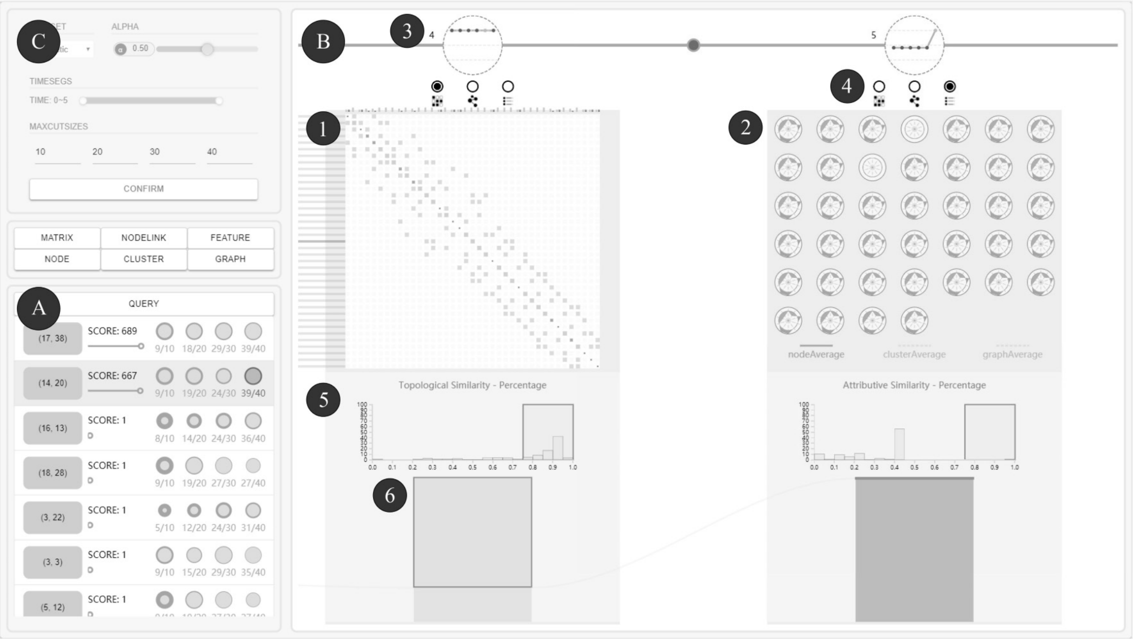
T3 Facilitating comparison among substructures This helps analysts to estimate the number of occurrence of a substructure's topology or attribute pattern in a snapshot of networks and identify anomalous topology or attribute of substructure if it is not similar to the most of substructures in the network. Hence, the differences among substructures should be revealed in iNet.

T4 Fully demonstrating the dynamics of a substructure

Identification of anomaly in network dynamics requires analysts to observe the topology and attribute of substructures and compare the differences of the substructure at different time stamps. Therefore, iNet should demonstrate the topology and attribute dynamics of substructures to analysts.

5.2 User interface

Following the guidance of design tasks, we design the user interface of iNet. iNet consists of a list of rare categories (Fig. 2(A)), a major view (Fig. 2(B)) and a parameter panel (Fig. 2(C)). The list of rare categories shows the features of rare categories (**T1**). The major view shows 1) topology structure and attribute at each time stamp with matrix representation, node-link diagram, and Z-glyphs (**T2**); 2) dynamics of topology and attribute by a time-line mapping (**T4**); 3) distribution of similarities between selected rare



ALPHA 0.50

TIMESEGS

TIME: 0~5

MAXCUTSIZES

10 20 30 40

CONFIRM

MATRIX	NODELINK	FEATURE
NODE	CLUSTER	GRAPH

QUERY

(17, 38) SCORE: 689 9/10 18/20 29/30 39/40

(14, 20) SCORE: 667 9/10 19/20 24/30 39/40

(16, 13) SCORE: 1 8/10 14/20 24/30 36/40

(18, 28) SCORE: 1 9/10 19/20 27/30 27/40

(3, 22) SCORE: 1 5/10 12/20 24/30 31/40

(3, 3) SCORE: 1 9/10 15/20 29/30 35/40

(5, 12) SCORE: 1

Fig. 2 User interface of iNet. A) The rare category identifier. B) The major view, which consists of four components: 1) the matrix representation of topology; 2) a group of Z-glyphs showing attributes of vertices in substructures; 3) a time-line with similarity glyph; 4) a switch button group; 5) similarity bar charts; 6) similarity sankey diagram. C) A parameter panel. A rare category with attribute anomaly at time stamp 5 in the synthetic dataset is shown: its topology is a grid network that is similar to the most substructures in the network and is stable over time while its attribute is different the most substructures. The similarity glyph also shows that the attribute of this category abruptly changed at time stamp 5

category and other sub-networks in topology and attribute by bar charts and a sankey diagram (**T3**).

5.2.1 The list of rare categories

The list of rare categories is designed for identifying the substructures that match the features of rare categories, i.e., compact and bordered. Analysts only need to examine the existence of the boundary of each substructure, as substructures detected by the local clustering algorithm are already compact structures around representative vertices detected by DIRAD. In this paper, a substructure has a boundary if vertices in the substructure have more internal connections than external connections.

A **substructure glyph** is used in the list to show the features of substructures, as shown in Fig. 2. It summarizes the features of a rare category candidate, including the size of the category, which is encoded by the size of the entire glyph; the ratio of internal connections, which is encoded by the size of the inner circle; and the ratio of external connections, which is encoded by the size of external ring.

The substructures found by DIRAD are ordered by the DIRAD score on the list. Each substructure is represented by an information panel. A group of candidate glyphs shows the information of clusters detected with a different upper bound of size inside each panel. According to these candidate glyphs, users can roughly determine if a candidate is potentially a rare category. Once a rare category is found in the list, users can explore the detail of the category in the major view.

Design considerations The most direct way to show the condition around a rare candidate is a node-link diagram or an adjacent matrix. However, these two representations require a rather large space. In the list of rare categories, the major task is to show the features of each rare category. Thus we use the candidate glyph design to visualize the compactness, the boundary, and the size of a rare category, which are more space-efficient than the node-link diagram and adjacent matrix.

5.2.2 The major view

The major view integrates a basic visualization of vertex sequences with matrix representation, node-link diagram, Z-glyphs, bar charts, sankey diagram, and line charts to help analysts to observe topology and attribute of the substructure and identify an anomaly in network snapshots and network dynamics.

The vertex sequences show the dynamics of vertices in a user-selected substructure along a horizontal time axis (in Fig. 3). At each time stamp, vertices are represented by vertex glyphs. The **vertex glyph** is a rectangle of which the size encodes the number of external connections, and the color encodes the number of internal connections of a vertex (see in Fig. 3). X-positions of glyphs are positioned by a fixed interval horizontally. Y-positions of glyphs are calculated by an energy-based layout algorithm [55]. The algorithm aims to minimize the energy function:

$$\sum_{t=0}^{T}(\alpha\sum_{i<j}w_{ij}(t)\|y_i(t)-y_j(t)\|^2+(1-\alpha)\sum_{i}\|y_i(t)-y_i(t-1)\|^2),$$

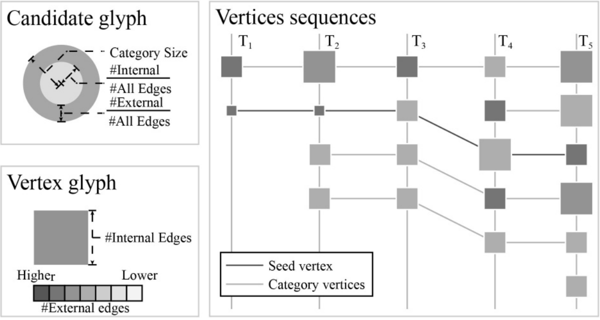


Fig. 3 Visual encoding of feature glyphs for the identification of rare categories. For candidate glyph, the size of the category is encoded by the size of the entire glyph; the ratio of internal connections is encoded by the size of the inner circle; and the ratio of external connections is encoded by the size of external ring. For vertex glyph, the size of the internal edges is encoded by the height of the block; the size of the external edges is encoded by the darkness of the color. For vertices sequences, the category vertices are linked with the lighter edges and seed vertices are linked with the darker ones

where $y_i(t)$ is the y-position of vertex i at time stamp t and $w_{ij}(t)$ is the inverse quadratic Euclidean distance of vertices i and j,

$$w_{ij}(t)=\frac{1}{d(v_i(t),v_j(t))^2},$$

where $v_i(t)$ is the feature of vertex i at time stamp t extracted in the storage module. The feature of each vertex at a time stamp is calculated by a pivot-based feature extraction method in the data storage module. A group of pivots is selected iteratively satisfying that kth has the largest sum of shortest path length to $\{1,2,...,k-1\}$ pivots. Then the feature of a vertex is the distances from the vertex to the pivots. Note that the first pivot is randomly selected.

Topology of the substructure is visualized by both matrix representation and node-link diagram (see Figs.4 (a) and (b)) are used to support the identification of the static topology pattern of rare categories [56]. In the matrix representation mode, the order of vertices is the same as in the sequence of vertices and the vertex glyphs are placed on the diagonal of the matrix to keep the mental map of users. As the layout algorithm is based on the vertex features, the layout result shows the feature of the category topology and helps users to identify the static topology pattern. The color of non-diagonal entries encodes the weight of edges. In the node-link diagram, we use vertex glyph to represent the vertices and use colored links among glyphs to represent edges to keep the mental map of users. The positions of glyphs are determined by force-directed layout.

Attribute of the substructure are visualized by a group of Z-glyphs [57], as shown in Fig. 4 (c). Each Z-glyph represents a vertex in the substructure and visualizes the divergence between attribute values and different baselines. Three-level of baselines are considered in this paper, including 1) vertex level baseline, which is the average of attributes of a single vertex and shows the consistency of the vertex attributes; 2) cluster level baseline, which is the average of each attribute in a cluster and shows the consistency of vertices in a cluster in attribute dimension; and 3) global level baseline, which is the average of each attribute in the entire network and shows the difference of the cluster and the entire network.

iNet identify anomalies in snapshots based on similarities, as previously mentioned. The procedure of calculating the similarity has two steps: first, extracting substructures around every vertex in the network at each time stamp with the same number of vertices as the selected substructure; second, calculating the topology and attribute similarities between the selected substructure and each extracted substructure. For a selected substructure with N vertices, the process of extracting a sub-network around a vertex V is: 1) find a k that satisfies number of vertices in the $k-1$ hop network is smaller than N and the k hop network is larger than N; 2) sort vertices in the k hop network with key (i,d), where i is the shortest path length between the vertex and V, and d is the degree of the vertex; 3) select top-N vertices to form the sub-substructure. We use Weisfeiler-Lehman graph kernel [58] to measure the similarity between two topology structure and the attribute similarity between the two substructures A and B is defined as:

$$s(A,B) = \frac{1}{1+d(\vec{a},\vec{b})}, \tag{9}$$

where $d(\vec{a},\vec{b})$ is the Euclidean distance between the feature vector of A and B. The feature vector is conducted by the average of each attribute and the average deviation of attributes of vertices in the substructure.

Similarities among the selected substructure and extracted substructures are visualized by a group of bar charts (Fig. 4(d)) and a sankey diagram (Fig. 4(e)). At a specific time stamp, a bar chart is used to show the distribution of the similarities. The x-axis is the similarity interval and the y-axis is the number of substructures. Substructures are aggregated into groups according to the similarity interval. The sankey diagram shows the dynamics of similarity of the substructures. In the sankey diagram, a rectangle represents all extracted substructures at a specific time stamp. The substructures are ordered descendingly according to the similarities from the top to the bottom of the rectangle. When analysts brush on a bar chart at a specific time stamp, brushed substructures will be highlighted on the rectangle at the time stamp in the sankey diagram. Simultaneously, the brushed operation will be copied to bar charts at other time stamps and brushed substructures at these time stamps will also be highlighted on the corresponding rectangles. Bands are then added between adjacent rectangles to show the change of substructure groups. Darker band indicates that the similarities of substructures changed and the lighter band indicates that the similarities of substructures are the same between the two time stamps. Bar charts enable users to identify anomalous in a snapshot of the multivariate dynamic network, and the sankey diagram enables users to track the similarity change of vertices selected in bar charts.

Similarities among different snapshots of the substructure are visualized by a line chart at each time stamp on the timeline (see in Fig. 4(f)). We use the same topology and attribute similarity measures to calculate the similarity among snapshot at a specific time stamp and snapshots at other time stamps and visualize the similarities with a line chart. The similarity at the time stamp of the snapshot equals 1 (self-

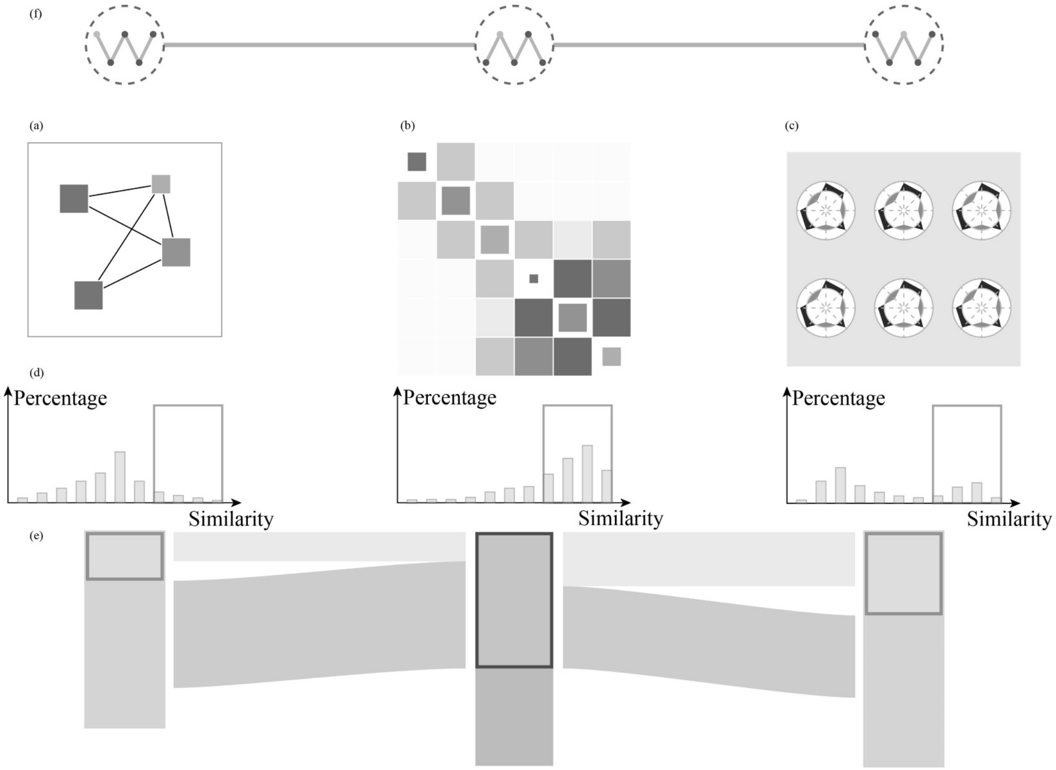


Fig. 4 Visual encoding in the topology evolution explorer and attributes evolution explorer: (a) node-link diagram of the topology of a rare category; (b) matrix representation of the topology of a rare category; (c) Z-glyphs of vertices attributes in a rare category; (d) similarity bar charts; (e) similarity sankey diagram; (f) temporal pattern glyph

similarity) and is highlighted by the lighter point, while similarities at other time stamps are represented by the darker points.

Designing considerations Timeline mapping, instead of animation, is used to visualize the dynamics of the topology, attribute, and similarities of substructures. According to Tversky et al. [59], animation increases memory burden and the difficulty of comparing the difference among different time stamps. As users may need to analyze and compare the topology and attribute patterns at each time stamp, we choose timeline mapping to lower the user burden in the visualization.

The node-link diagram and the matrix representation are the two most common graph visualization techniques. Ghoniem et al. [60] stated that matrix representations have better performance than the node-link diagram when the number of vertices is larger than 20. We also find out that our collaborators adapt better to the node-link diagram than the matrix representation through discussions. Besides, the number of vertices in rare categories should not be large. Otherwise, they are not rare. Thus, we decide to provide both the node-link diagram and the matrix representation and let users choose their preferable visualization form.

Furthermore, we use a list of baseline-based bar charts to show attributes of vertices in a rare category in the first version of iNet. However, when a category contains a large number of vertices, the height of a single bar chart is limited, which decreases the effectiveness of the identification of the distribution of the attributes. Therefore, we decide to use the Z-glyph design for the compactness and space efficiency in the second version of iNet.

Last, we use line charts to show the dynamics of substructure topology and attributes along time. Analysts can observe the dynamics by directly comparing the matrices, node-link diagrams, and z-glyphs at each time stamps. However, this way is neither intuitive nor convenient. Therefore, we add the line charts on the timeline to provide analysts an intuitive way to analyze the dynamics of substructures.

5.3 Interactions in analysis loop

As mentioned in Section 3, we design a series of interactions in each visualization components to keep users in the analysis loop. In iNet, the procedure of analyzing irregular transitions includes three stages, including the query stage, the identify stage and the interpret stage.

In the query stage, users first select two adjacent time stamps on the timeline (Fig. 2 (3)) to initialize the DIRAD algorithm by clicking on the circle button between two similarity line charts. Then, the DIRAD algorithm detects substructures in the network. At the beginning of the entire analysis procedure, users can select time stamps they are interested in by brushing on a time axis in the parameter panel. Users can also adjust the weights of topology and attributes in DIRAD by the slider in the parameter panel and re-query candidates.

In the identify stage, users identify rare categories among the substructures detected by DIRAD. If no rare category is found, users return to the query stage, and query substructures again. After the detected candidates are shown in the list, users can roughly determine whether there are any rare categories in these candidates. Users can directly skip substructures that are impossible to be rare categories and explore the details of substructures that are potentially rare categories by clicking on them. As users cannot predict where they will find a clear boundary of a rare category candidate, they are enabled to adjust the max number of vertices in substructures in the parameter panel.

In the interpret stage, users can first identify topology and attribute patterns of rare categories and interpret how it is abnormal. After the analysis is done, the result will be returned to DIRAD, and the next iteration will be executed. It contains five major interactions: expanding/collapsing of time stamps, dragging, switching, and hovering. When users demand to explore the topology evolution or attributes the evolution of a rare category, they can expand the time stamps they interested in by clicking on the vertical lines on the time axis. By clicking on the lines again, the corresponding time stamps will be collapsed. When multiple time stamps are expanded, and the number of vertices in the rare category is large enough, the screen cannot accommodate the sequence of vertices. Thus, we add a dragging interaction on the sequence, which allows users to explore the entire sequence by moving it left and right. Users can freely switch the visualization form at a specific time stamp between matrix representation, node-link diagram, and multiple baseline bar chart through radio buttons on the top of the time stamp (as shown in Fig. 2). When users put the mouse on a vertex at a time stamp, the same vertex at other time stamps will be highlighted to help users track the evolution of the vertex easier.

In this section, we first demonstrate the usability of DIRAD by detecting rare categories with DIRAD on three real datasets with ground truth, then conduct a use scenario to show the effectiveness of iNet based on a dataset without ground truth, and last conduct a user study to demonstrate the usability of iNet by a synthetic dataset. The prototype system is a web application. The front-end visualization is implemented by AngularJS, D3, and CSS. The back-end server is implemented by Python with Flask, numpy, scipy, and networkx. Use scenarios and the user study run on a PC with Intel(R) Core(TM) i7-4770 CPU, 20 GB RAM, and Windows 10.

6 Algorithm performance

We test our DIRAD algorithm on three real-world dynamic multivariate networks, including review networks, Q&A networks, and collaboration networks. The Epinion [61] dataset is a who-trust-whom network derived from Epinion, where each vertex represents a user, and each edge indicates whether one user trusts another user at a certain time stamp. The Stackoverflow[1] dataset is collected from the question and answer site Stack Overflow, where each vertex represents a Stack Overflow user, and each edge indicates one comment from one user to another. The DBLP[2] dataset is generated based on the IEEE Visualization publications from 1990-2015.

[1] https://archive.org/details/stackexchange

[2] http://www.vispubdata.org/site/vispubdata/

In DBLP, each vertex represents a paper, and an edge exists if and only if when one paper cites another paper.

In the following experiments, we compare DIRAD with the following methods: (1) GRADE-G: the graph-based rare category detection method GRADE [52] that is built purely on evolving graphs. Here, GRADE is short for Grading of Recommendations Assessment, Development and Evaluation, which is generally adopted by organisations worldwide. (2) GRADE-A: the GRADE algorithm that is built on evolving vertices' attributes; (3) RANDOM: the random sampling on the given network at each time stamp. The effectiveness comparison results regarding the number of queries are shown in Fig. 5. We have the following observations: (1) when the topology domain and feature domain are consensuses, our DIRAD algorithm could balance the information that is extracted from these two domains and enhance the performance, i.e., our DIRAD algorithm requires fewer queries in most cases in Figs. 5(a) and 5(b); (2) when the RCD model built on topology domain or feature domain is biased (e.g., GRADE-G performs worse than random sampling when $t = 1,2$ in Fig. 5(c)), our multi-view learning framework reduces the negative impact and ensures our model to achieve robust performance in the noisy scenarios.

7 Use scenario: collaboration network in computer science

This dataset is extracted from DBLP[3]. We first extract 89767 papers published in 41 conferences and journals, then we separate them into 12 research fields. The papers are written by 6758 researchers. We model the collaborations among researchers to a dynamic network: if two researchers collaborate for N times before a time stamp T, we add an edge with weight N to link the two researchers at the time stamp T. Attributes of researchers are the papers they published in different research fields. Thus the attribute of a researcher is a temporal (k) dimension vector. During the detection of anomalies, users should traverse all time stamps and find anomalies at each time stamps. However, considering the limitation of space, we only demonstrate some representative findings at some time stamps here.

In Fig. 6, a major category in both topology dimension and attribute dimension are found. The topology explorer shows that vertices in this category form a constant sub-network which consists of three small communities. In each community, a center vertex exists and thus makes the topology of these communities similar to star structure. The similarity bar charts and the Sankey diagram shows that the similarities among the category and around 50% of sub-networks are 0.6 or more at each time stamp. Z-glyphs in attribute explorer shows that researchers in this category are very productive in a specific field. Meanwhile, 12/13 (92%) researchers have publications in other fields, which indicates that researchers in this category have a major field and at least one secondary field. The similarity bar charts and the sankey diagram show that more than 50% of sub-networks are similar (similarity > 0.5) to this category in attribute dimension in 2016 and 2017.

Figure 7(a) shows a rare category with topology anomaly in snapshots. The topology explorer shows that the category contains two small connected cliques. The similarity bar charts and the sankey diagram show that there are almost no sub-networks being similar to this category. Figure 7(b) shows a rare category with attribute anomaly in snapshots. Z-glyphs in the attribute explorer shows that researchers in this category have a major field, which is the same as the major category. However, 19/26 (73%) researchers seldom publish in other fields, which indicates that researchers in this group are very focusing researchers and might seldom cooperate with researchers in other fields. The similarity bar charts and the sankey diagram show that very few sub-networks are similar to them (similarity > 0.5).

Figure 8 shows a rare category with topology anomaly in network dynamics. In 2014 and 2015, there are very few connections among the vertices, although these vertices already exist in the network. The matrix representation and the node-link diagram in the topology explorer show that a few links are added from 2014 to 2015, which indicates that the pattern of the category in 2014 and 2015 can be identified as a constant or an enhancing near star pattern. However, the topology pattern significantly changes to a clique in 2016 and remains the same pattern in 2017. Therefore, a change happens in 2016 in the topology dimension, and thus the vertices belong to a rare category with topology anomaly in network dynamics.

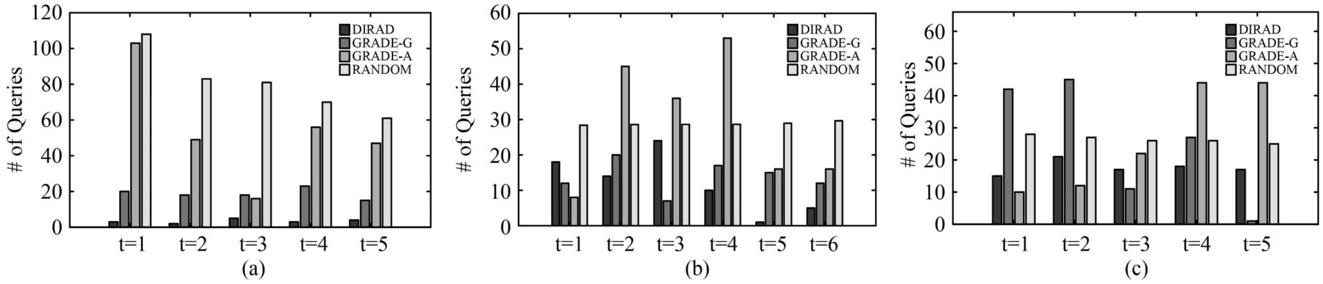
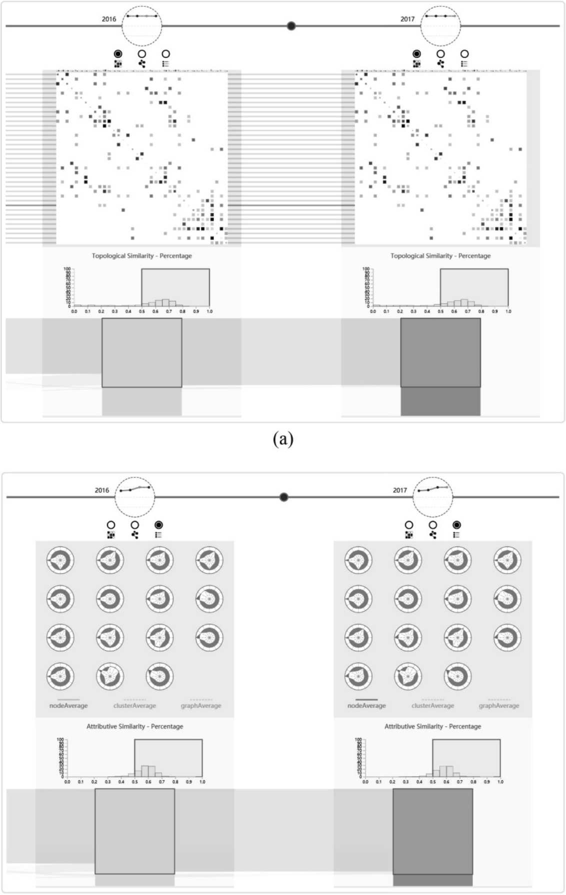


Fig. 5 Effectiveness Analysis. Our DIRAD method and three other methods are used on three real-world datasets. Since the total query time includes calculation whose time is short enough to be ignored and communication of network which is unstable and relatively much longer. We use the query times instead of the total query time to represent the efficiency of the algorithms. As shown in the three bar charts respectively, our DIRAD method requires fewer queries in most cases, which shows a better performance. (a) Epinion; (b) Stackoverflow; (c) DBLP

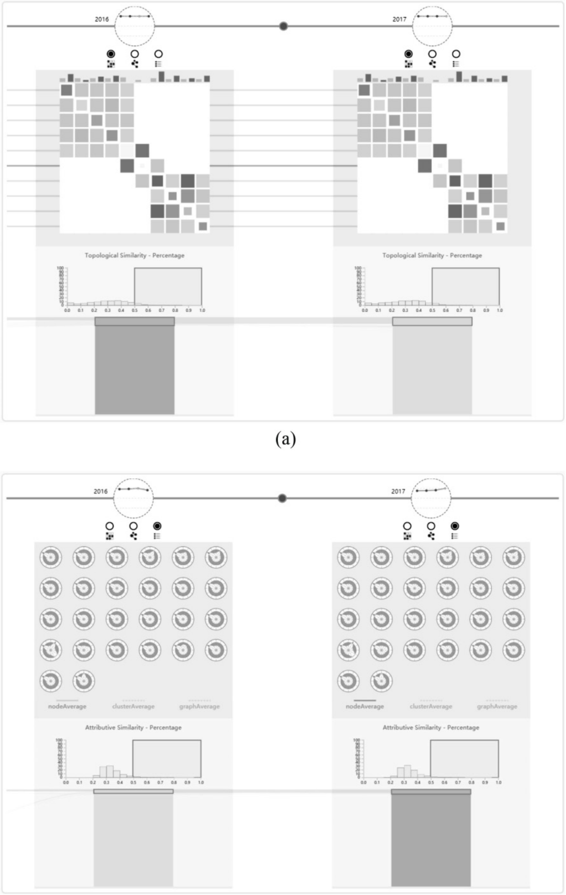
[3] http://dblp.uni-trier.de/



(a)

(b)

Fig. 6 A major category in a collaboration network with its patterns and similarities. A major category can be found in both topology dimension (a) and attribute dimension (b). (a) In topology pattern, it shows three small communities formed by the vertices in this category, with each community containing a center vertex. As a result, we can view the topology similarly as star structures. The corresponding similarity bar charts and the Sankey diagram shows that around 50% of sub-networks are 0.6 or more at each time stamp, which indicates the similarity of the category. (b) In attribute pattern, the Z-glyphs show that researchers in the above major category are very productive in a specific field according to our definition of the network. Its corresponding similarity bar charts and the Sankey diagram show that more than 50% of sub-networks are similar (similarity > 0.5) to this category in attribute dimension in 2016 and 2017



(a)

(b)

Fig. 7 Two rare categories with type 1 anomalies shown in topology dimension (a) and attribute dimension (b) respectively. (a) In topology pattern, it shows that the category contains two small connected cliques. The similarity bar charts and the sankey diagram show that there are almost no sub-networks being similar to this category. (b) In attribute pattern, the Z-glyphs show that researchers in this category have a major field, which is the same as the major category. However, most of them seldom publish in other fields, which indicates that they are very focusing. The similarity bar charts and the Sankey diagram show that very few sub-networks are similar to them (similarity > 0.5)

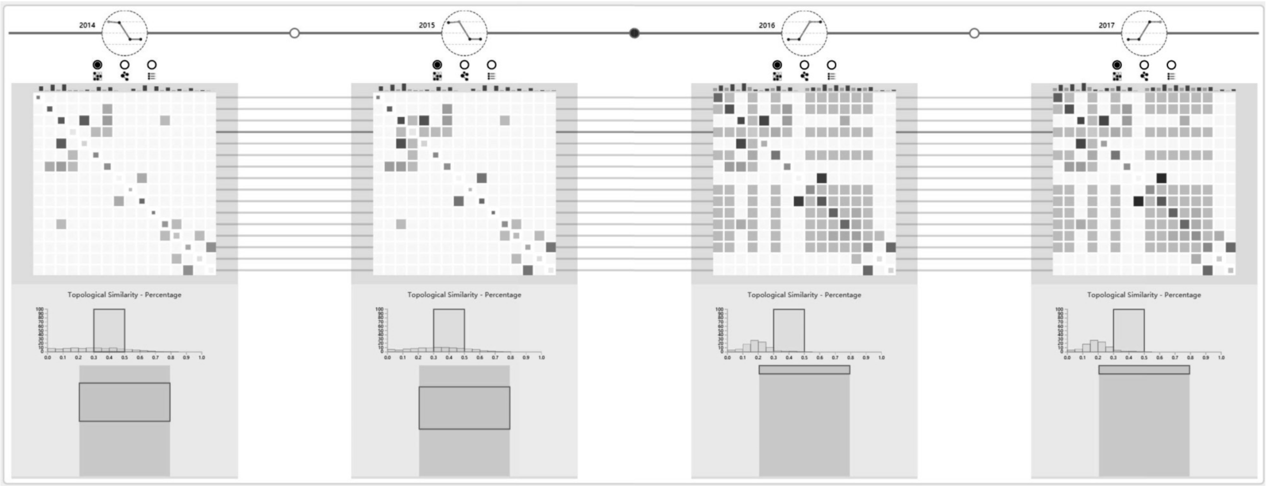


Fig. 8 Topology anomaly in network dynamics: the topology structure pattern significantly changed between 2015 and 2016

8 User study

Participants In this study, we invited 15 researchers to participate in our user study to estimate the effectiveness of our prototype system in analyzing anomalies in multivariate dynamic networks. Two of them are senior researchers with more than 5 years of visualization experiences, and others are junior researchers with more than 3 years of visualization experiences.

Dataset Due to the high complexity of real datasets used in previous evaluations, we construct an ideal synthetic dataset to run the user study, which brings three benefits: first, synthetic datasets have more valid ground truth, which increases the accuracy of the study, second, the size and the number of time stamps is controlled in synthetic datasets and thus reduces the analysis procedure, third, anomalies in synthetic datasets are more clear, and thus participants have less burden during the study. The synthetic dataset is constructed in the following steps: 1) generate a 25*40 grid network in which vertices have consistent attributes at each time stamps; 2) add stars, cliques, and bipartite cores into the network at the last time stamp; 3) add different attribute pattern to vertices in those special structures; 4) add temporal presence of topology and attributes of special structures along time. Information on major categories and rare categories is shown in Table 1.

Tasks During the user study, participants were asked to discover anomalies of rare categories as many as possible between time stamp 5 and time stamp 6 of the synthetic dataset. To ensure that participants fully explore the functionality of the prototype system, the participants were required to complete the following tasks.

T1 **Identify the major categories and describe the corresponding pattern.**

T2 **Identify three kinds of anomalies of rare categories in network snapshots and describe their corresponding patterns.**

T3 **Identify two kinds of anomalies of rare categories in network dynamics and describe their corresponding patterns.**

Procedures The user study has three stages. First, we introduce the basic concepts in this work, including the “rare category”, patterns in multivariate dynamic networks, and major features of the system, with a 10 minutes tutorial. Then, we let participants to explore the functionality of the system for ten minutes. During the exploration, participants can asked any questions about the system. Last, participants are asked to explore the synthetic dataset in the system, complete the tasks, and answer the corresponding questions. The whole exploration is timed. During the study, participants are allowed to ask moderators any questions about the prototype system to avoid confusion. Participants are told that there are a major category, three kinds of type 1 rare categories, and two type 2 rare categories. When users think they find all rare categories, the study is finished. The time spent, number of queries, false positive rare, and false negative rare are recorded and calculated after the study is finished.

Results The results is shown in Tables 2 and 3. Most participants managed to complete all the tasks within 30 minutes (1-2 minutes per vertex). In total, participants are able to distinguish the major category and rare categories with $FPR = 0$ and $FNR = 6.25\%$ on average. For interpretation of rare category anomalies, the average accuracy of identification of topology pattern, temporal presence of topology pattern, attribute pattern, and temporal presence of attribute pattern are 90%, 84%, 90%, and 90%. Although most participants recognized subjects with type 2 (d) and type 2 (e) as rare anomalous categories, some participants failed to correctly figure out the changes between topology patterns and attribute patterns, as the accuracy of identification of temporal presence of topology in T3 (d) and temporal presence of attribute in T3 (e) are both 70%. Also, some of the participants failed to distinguish a bipartite structure and a clique structure in T2 (a), which might because some of the participants lack experience in analyzing graph structure.

9 Discussion and future works

Direct clustering vs. our method A common confusion about our method is why do not we direct use some data mining techniques such as clustering to find anomalies in multivariate dynamic networks. As we stated in Section 4, the features of rare categories in multivariate dynamic networks are compact, bordered, and size-sensitive, which indicate that a small group of vertices which form a more compact structure inside a large group of compact vertices can also be identified as a rare category. However, common clustering methods can

Table 1 Categories in the synthetic dataset

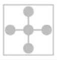
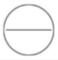
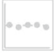
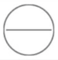
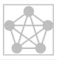
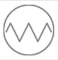
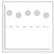
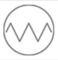
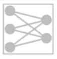
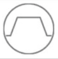
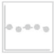
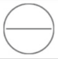
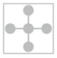
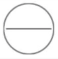
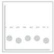
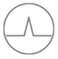
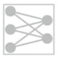
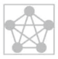
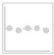
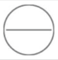
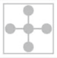
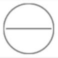
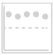
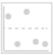
Categories		T	t	A	t	PCT
Major						75%
Snapshot	a					6%
	b					4%
	c					5%
Dynamics	d	 	-			5%
	e			 	-	5%

Table 2 Result of identification of rare anomalous category

	TP	FP	TN	FN	FPR	FNR
Anomalous Rare Category	75	0	40	5	0	6.25%

Table 3 Results of the interpretation of category patterns

		T	t	A	t
T2	a	75%	85%	90%	80%
	b	95%	85%	100%	100%
	c	100%	100%	80%	100%
T3	d	90%	70%	90%	95%
	e	100%	90%	80%	70%
Avg.		90%	84%	90%	90%

hardly detect clusters in this microscope. Instead, DIRAD aims to find vertices on the boundary of areas with density changes both in the topology dimension and attribute dimension and thus are more proper in our scenario.

Limitations Although the prototype system can assist users in identifying, exploring, and interpreting the rare categories in multivariate dynamic networks, limitations still exist in several aspects. First, more interactions are required to support more accurate analysis. If users are enabled to brush some vertices in topology explorer and attribute explorer and re-calculate the similarity, users can find a rare category more accurately. Besides, it will be more intuitive if a visual comparison is provided to enable users to directly compare the rare category and other equal-size sub-networks with specific similarity. Second, the similarity calculation is a time-consuming process. As the similarity is calculated for every vertex in the network, the process can be accelerated by parallel computation. We plan to improve the interactions of the prototype system and re-implement similarity calculation in GPU in the next version.

Future works First, we will improve the DIRAD to make it detecting rare category candidates more efficiently. Moreover, we will design a classification-based rare category detection method for multivariate dynamic networks. Second, we plan to add more components to show domain context information, which is generally changing case by case but can help users to understand the network data better. Last, we plan to add more data operation interactions, including vertex querying and vertex filtering. With these interactions, users can focus on vertices in which they interested in and reduce their workload.

10 Conclusion

In this paper, we present a novel rare category detection method, called DIRAD, which detects rare category candidates in multivariate dynamic networks and introduce a prototype visualization system which enables users to identify true rare categories among candidates, explore and interpret the anomalies of rare categories. We first summarize possible patterns in three dimensions of multivariate topology networks, including topology dimension, attribute dimension, and time dimension. Evaluations are conducted to evaluate our method in three aspects: 1) effectiveness of DIRAD in detecting rare category candidates; 2) effectiveness of the prototype system in identifying and analyzing rare categories; 3) effectiveness of the prototype system in assisting users to complete analytical tasks.

Acknowledgements This work was supported by National Key Research and Development Program (2018YFB0904503), the National Natural Science Foundation of China (Grant Nos. 61772456, U1866602, 61761136020, U1736109).

References

1. Zhao Y, Luo X, Lin X, Wang H, Kui X, Zhou F, Wang J, Chen Y, Chen W. Visual analytics for electromagnetic situation awareness in radio monitoring and management. IEEE Transactions on Visualization and Computer Graphics, 2020, 26(1): 590–600
2. Beck F, Burch M, Diehl S, Weiskopf D. The state of the art in visualizing dynamic graphs. In: Proceedings of Eurographics Conference on Visualization. 2014, 83–103
3. Chandola V, Banerjee A, Kumar V. Anomaly detection: a survey. ACM Computing Surveys, 2009, 41(3): 1–58
4. Ranshous S, Shen S, Koutra D, Harenberg S, Faloutsos C, Samatova N F. Anomaly detection in dynamic networks: a survey. Wiley Interdisciplinary Reviews Computational Statistics, 2015, 7(3): 223–247
5. Zhou D, He J, Cao Y, Seo J S. Bi-level rare temporal pattern detection. In: Proceedings of IEEE International Conference on Data Mining Series. 2016, 719–728
6. Mei H, Chen W, Wei Y, Hu Y, Zhou S, Lin B, Zhao Y, Xia J. Rsatree: distribution-aware data representation of large-scale tabular datasets for flexible visual query. IEEE Transactions on Visualization and Computer Graphics, 2019, 26(1): 1161–1171
7. Pelleg D, Moore A W. Active learning for anomaly and rare-category detection. In: Proceedings of the 17th International Conference on Neural Information Processing Systems. 2004, 1073–1080
8. Liu Z, Chiew K, He Q, Huang H, Huang B. Prior-free rare category detection: more effective and efficient solutions. Expert Systems with Applications, 2014, 41(17): 7691–7706
9. He J, Tong H, Carbonell J. Rare category characterization. In: Proceedings of IEEE International Conference on Data Mining. 2010, 226–235
10. Zhou D, Wang K, Cao N, He J. Rare category detection on timeevolving graphs. In: Proceedings of IEEE International Conference on Data Mining. 2015, 1135–1140
11. Cheng Z, Chang X, Zhu L, Kanjirathinkal R C, Kankanhalli M. MMALFM: explainable recommendation by leveraging reviews and images. ACM Transactions on Information Systems, 2019, 37(2): 16
12. Liu A A, Nie W Z, Gao Y, Su Y T. Multi-modal clique-graph matching for view-based 3d model retrieval. IEEE Transactions on Image Processing, 2016, 25(5): 2103–2116
13. Liu A A, Su Y T, Nie W Z, Kankanhalli M. Hierarchical clustering multi-task learning for joint human action grouping and recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39(1): 102–114
14. Huang H, He Q, Chiew K, Qian F, Ma L. Clover: a faster priorfree approach to rare-category detection. Knowledge and Information Systems, 2013, 35(3): 713–736
15. Zhou D, Karthikeyan A, Wang K, Cao N, He J. Discovering rare categories from graph streams. Data Mining and Knowledge Discovery, 2016, 31(2): 1–24
16. He J, Carbonell J. Prior-free rare category detection. In: Proceedings of SIAM International Conference on Data Mining. 2009, 155–163
17. Huang H, He Q, He J, Ma L. Radar: rare category detection via computation of boundary degree. In: Proceedings of Pacific-Asia Conference on Advances in Knowledge Discovery and Data Mining. 2011, 258–269
18. Vatturi P, Wong W K. Category detection using hierarchical mean shift. In: Proceedings of International Conference on Knowledge Discovery and Data Mining. 2008, 847–856
19. Pan J, Han D, Guo F, Zhou D, Cao N, He J, Xu M, Chen W. Rcanalyzer: Visual analytics of rare categories in dynamic networks. Frontiers of Information Technology and Electronic Engineering, 2020, 21(4): 491–506
20. Kind A, Stoecklin M P, Dimitropoulos X. Histogram-based traffic anomaly detection. IEEE Transactions on Network and Service Management, 2009, 6(2): 110–121
21. Luo X, Yuan Y, Zhang K, Xia J, Zhou Z, Chang L, Gu T. Enhancing statistical charts: toward better data visualization and analysis. Journal of Visualization, 2019, 22(4): 819–832

22. Xia J, Ye F, Zhou F, Chen Y, Kui X. Visual identification and extraction of intrinsic axes in high-dimensional data. IEEE Access, 2019, 7(1): 79565–79578
23. Inselberg A. Parallel Coordinates, 1st ed. New York: Springer, 2009
24. Cao N, Gotz D, Sun J, Qu H. Dicon: interactive visual analysis of multidimensional clusters. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(12): 2581–2590
25. Zhao Y, Wang L, Li S, Zhou F, Lin X, Lu Q, Ren L. A visual analysis approach for understanding durability test data of automotive products. ACM Transactions on Intelligent Systems and Technology, 2019, 10(6): 70
26. Xu P, Mei H, Liu R, Wei C. Vidx: visual diagnostics of assembly line performance in smart factories. IEEE Transactions on Visualization and Computer Graphics, 2017, 23(1): 291
27. Corchado E, Herrero Á. Neural visualization of network traffic data for intrusion detection. Applied Soft Computing, 2011, 11(2): 2042–2056
28. Tsai C F, Hsu Y F, Lin C Y, Lin W Y. Intrusion detection by machine learning: a review. Expert Systems with Applications, 2009, 36(10): 11994–12000
29. Thom D, Bosch H, Koch S, Wörner M, Ertl T. Spatiotemporal anomaly detection through visual analysis of geolocated twitter messages. In: Proceedings of Pacific Visualization Symposium. 2012, 41–48
30. Zhao J, Cao N, Wen Z, Song Y, Lin Y R, Collins C. Fluxflow: visual analysis of anomalous information spreading on social media. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(12): 1773–1782
31. Cao N, Shi C, Lin S, Lu J, Lin Y R, Lin C Y. Targetvue: visual analysis of anomalous user behaviors in online communication systems. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 280–289
32. Rufiange S, McGuffin M J. Diffani: visualizing dynamic graphs with a hybrid of difference maps and animation. IEEE Transactions on Visualization and Computer Graphics, 2013, 19(12): 2556–2565
33. Bach B, Pietriga E, Fekete J D. Graphdiaries: animated transitions andtemporal navigation for dynamic networks. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(5): 740–754
34. Frishman Y, Tal A. Online dynamic graph drawing. IEEE Transactions on Visualization and Computer Graphics, 2008, 14(4): 727–740
35. Brandes U, Nick B. Asymmetric relations in longitudinal social networks. IEEE Transactions on Visualization and Computer Graphics, 2011, 17(12): 2283–2290
36. Oelke D, Kokkinakis D, Keim D A. Fingerprint matrices: uncovering the dynamics of social networks in prose literature. In: Proceedings of Computer Graphics Forum. 2013, 371–380
37. Burch M, Schmidt B, Weiskopf D. A matrix-based visualization for exploring dynamic compound digraphs. In: Proceedings of International Conference on Information Visualisation. 2013, 66–73
38. Vehlow C, Burch M, Schmauder H, Weiskopf D. Radial layered matrix visualization of dynamic graphs. In: Proceedings of the 17th International Conference on Information Visualisation. 2013, 51–58
39. Bach B, Pietriga E, Fekete J D. Visualizing dynamic networks with matrix cubes. In: Proceedings of Annual ACM Conference on Human Factors in Computing Systems. 2014, 877–886
40. Zhao J, Liu Z, Dontcheva M, Hertzmann A, Wilson A. Matrixwave: visual comparison of event sequence data. In: Proceedings of Sigchi Conference on Human Factors in Computing Systems. 2015, 259–268
41. Li J, Chen S, Zhang K, Andrienko G, Andrienko N. Cope: interactive exploration of co-occurrence patterns in spatial time series. IEEE Transactions on Visualization and Computer Graphics, 2019, 25(8): 2554–2567
42. van den Elzen S, Holten D, Blaas J, van Wijk J J. Reducing snapshots to points: a visual analytics approach to dynamic network exploration. IEEE Transactions on Visualization and Computer Graphics, 2016, 22(1): 1–10
43. Vehlow C, Beck F, Auwärter P, Weiskopf D. Visualizing the evolution of communities in dynamic graphs. Computer Graphics Forum, 2015, 34(1): 277–288
44. Cui W, Wang X, Liu S, Riche N H, Madhyastha T M, Ma K L, Guo B. Let it flow: a static method for exploring dynamic graphs. In: Proceedings of IEEE Pacific Visualization Symposium. 2014, 121–128
45. Hlawatsch M, Burch M, Weiskopf D. Visual adjacency lists for dynamic graphs. IEEE Transactions on Visualization and Computer Graphics, 2014, 20(11): 1590–1603
46. Ying Z, She Y, Chen W, Yutian L, Xia J, Chen W, Liu J, Zhou F. Eod edge sampling for visualizing dynamic network via massive sequence view. IEEE Access, 2018, 6(1): 53006–53018
47. Zhou D, Weston J, Gretton A, Bousquet O, Schölkopf B. Ranking on data manifolds. In: Proceedings of the 16th International Conference on Neural Information Processing Systems. 2003, 169–176
48. Woodbury M A. Inverting modified matrices. Memorandum Report, 1950, 42(106): 336
49. Zhu M, Chen W, Xia J, Ma Y, Zhang Y, Luo Y, Huang Z, Liu L. Location2vec: a situation-aware representation for visual exploration of urban locations. IEEE Transactions on Intelligent Transportation Systems, 2019, 20(10): 3981–3990
50. Zhou D, He J, Candan K S, Davulcu H. Muvir: multi-view rare category detection. In: Proceedings of International Joint Conferences on Artificial Intelligence. 2015, 4098–4104
51. He J, Carbonell J G. Nearest-neighbor-based active learning for rare category detection. In: Proceedings of the 21st Annual Conference on Neural Information Processing Systems. 2007, 633–640
52. He J, Liu Y, Lawrence R. Graph-based rare category detection. In: Proceedings of Industrial Conference on Data Mining. 2008, 833–838
53. Silverman B W. Density Estimation for Statistics and Data Analysis, 1st ed. New York: Routledge, 1998
54. Andersen R, Chung F, Lang K. Local graph partitioning using pagerank vectors. In: Proceedings of IEEE Symposium on Foundations of Computer Science. 2006, 475–486
55. Guo S, Xu K, Zhao R, Gotz D, Zha H, Cao N. Eventthread: visual summarization and stage analysis of event sequence data. IEEE Transactions on Visualization and Computer Graphics, 2017, 24(1): 56–65
56. Chen W, Guo F, Han D, Pan J, Nie X, Xia J, Zhang X. Structure-based suggestive exploration: a new approach for effective exploration of large networks. IEEE Transactions on Visualization and Computer Graphics, 2018, 25(1): 555–565
57. Cao N, Lin Y R, Gotz D, Du F. Z-Glyph: visualizing outliers in multivariate data. Information Visualization, 2018, 17(1): 22–40
58. Shervashidze N, Schweitzer P, Leeuwen E J V, Mehlhorn K, Borgwardt K M. Weisfeiler-lehman graph kernels. Journal of Machine Learning Research, 2011, 12(3): 2539–2561
59. Tversky B, Morrison J B, Betrancourt M. Animation: can it facilitate? International Journal of Human-Computer Studies, 2002, 57(4): 247–262
60. Ghoniem M, Fekete J D, Castagliola P. A comparison of the readability of graphs using node-link and matrix-based representations. In: Proceedings of IEEE Symposium on Information Visualization. 2005, 17–24
61. Tang J, Gao H, Liu H. mTrust: discerning multi-faceted trust in a connected world. In: Proceedings of ACM International Conference on Web Search and Data Mining. 2012, 93–102

Dongming Han is currently working towards the PhD degree with the State Key Laboratory of Computer Aided Design and Computer Graphics, Zhejiang University, China. His research interests include visualization and visual analytics.

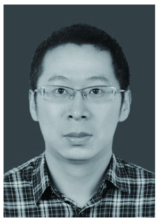
Jiacheng Pan is currently working towards the PhD degree with the State Key Laboratory of Computer Aided Design and Computer Graphics, Zhejiang University, China. He is a fan of visualization and visual analytics.

Rusheng Pan is currently working towards the PhD degree with the State Key Laboratory of Computer Aided Design and Computer Graphics, Zhejiang University, China. His research interests include visualization and visual analytics.

Dawei Zhou is a PhD candidate at Department of Computer Science and Engineering, Arizona State University, USA. His research interests are rare category detection, multi-view learning, and spatio-temporal learning.

Nan Cao is a professor at TongJi University in China, with the joint appointment at both College of Design and Innovation and College of Software Engineering. His primary expertise and research interests are information visualization and visual analysis.

Jingrui He is an assistant professor of computer science at Arizona State University, USA. Her research interests are heterogeneous machine learning, rare category analysis, active learning and semisupervised learning, with applications in social network analysis, healthcare, and manufacturing.



Mingliang Xu is a professor in the School of Information Engineering of Zhengzhou University, China. His current research interests include computer graphics, multimedia, artificial intelligence and virtual reality.

Wei Chen is a professor in State Key Lab of CAD&CG at Zhejiang University, China. He has performed research in visualization and visual analysis. His current research interests include visualization, visual analytics and bio-medical image computing.