Jia-cheng PAN, Dong-ming HAN, Fang-zhou GUO, Da-wei ZHOU, Nan CAO, Jing-rui HE, Ming-liang XU, Wei CHEN, 2020. RCAnalyzer: visual analytics of rare categories in dynamic networks. *Frontiers of Information Technology & Electronic Engineering,* 21(4):491-506. https://doi.org/10.1631/FITEE.1900310

RCAnalyzer: visual analytics of rare categories in dynamic networks

Key words: Rare category detection; Dynamic network; Visual analytics

Corresponding author: Wei CHEN
E-mail: chenvis@zju.edu.cn
ORCID: https://orcid.org/0000-0002-8365-4741

Motivation

1. Finding anomalous changing behaviors of network structures is valuable.
2. Most of the existing anomaly detection algorithms are automatic, and do not take human insights into account. Visualization could be helpful in terms of supporting interactive data exploration and providing a rich context representation.
3. Challenges exist in designing such a visualization system to support the process of rare category detection (RCD) in a dynamic network. None of the existing techniques is developed to support the visualization of rare categories. Capturing the changing structures of rare categories in the context of a big dynamic graph is challenging.

Main idea

1. A large dynamic network in the form of a series of connected triangular matrices is represented to show snapshots.
2. A hierarchical clustering algorithm and a tree cut algorithm are developed to produce an adaptive “focus+context” view, so that a large graph can be fully displayed while showing the detailed structures of potential rare categories.
3. An integrated visual analysis system is presented to support the detection of rare categories and facilitate rare category labeling.

Method

We identify the following design requirements that RCAnalyzer should fulfill based on the analytical tasks:

1. Provide an overview of dynamic networks.

2. Capture the changing structures of rare categories in the context of dynamic networks.

3. Reveal the features of detected examples.

4. Reserve the context of labeled rare categories.

5. Enable users to set and reset the labels of candidates.

Method

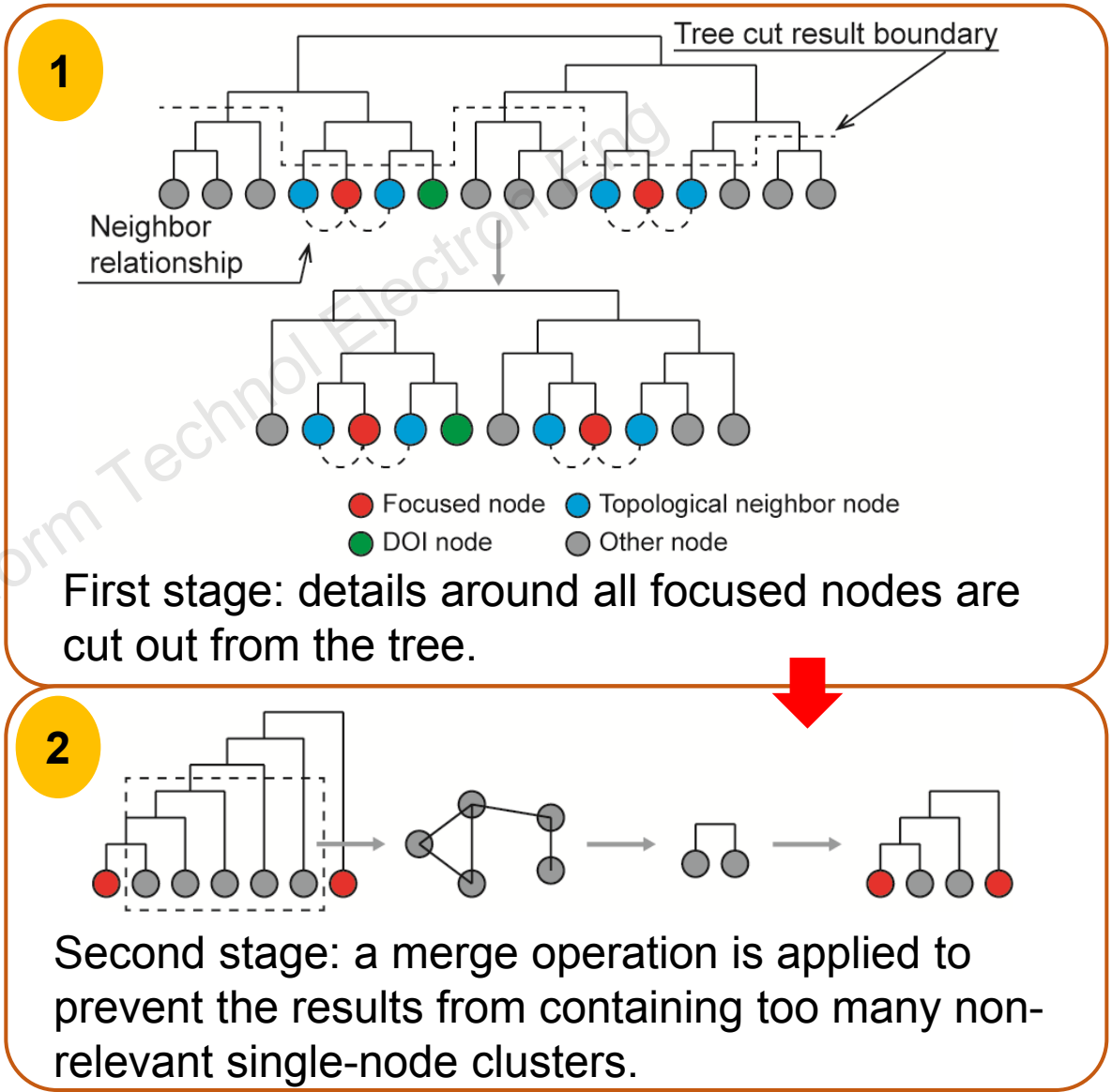
Following the design requirements, we design RCAnalyzer using batch-update incremental RCD (BIRD), which is the first (and only) work designed for detecting rare categories in dynamic networks.

User interface of RCAnalyzer

(a) the timeline view shows a high-level overview;
(b) the matrices view shows the aggregated adjacency matrix of dynamic networks and the details of the selected nodes' neighborhood;
(c) the instance view shows the feature of candidates and the query history of BIRD;
(d) the sub-network view shows the sub network of the query result;
(e) the label result view shows the historical label results.

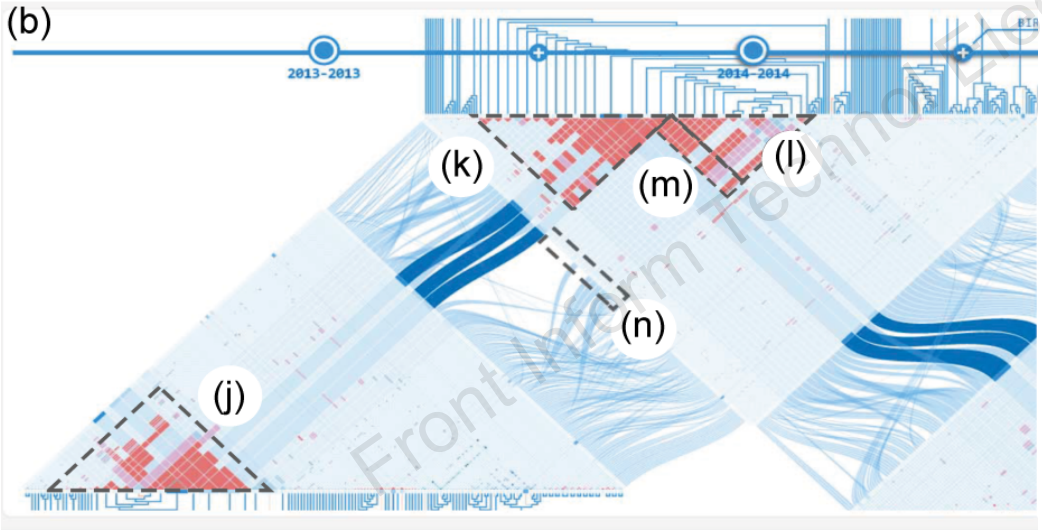
Method

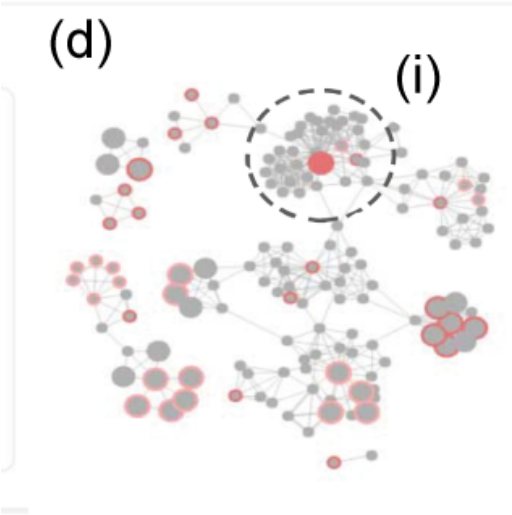
A multi-focus tree cut algorithm is designed to enable the matrices view to provide fine-grained details around the selected nodes and coarse-grained details around the unrelated nodes. It involves two stages.



Major results (use scenario)

We have extracted all co-authorships in the IEEE VIS dataset from 1990 to 2015. The matrices view (b) and the sub-network view (d) show the basic information of the network.

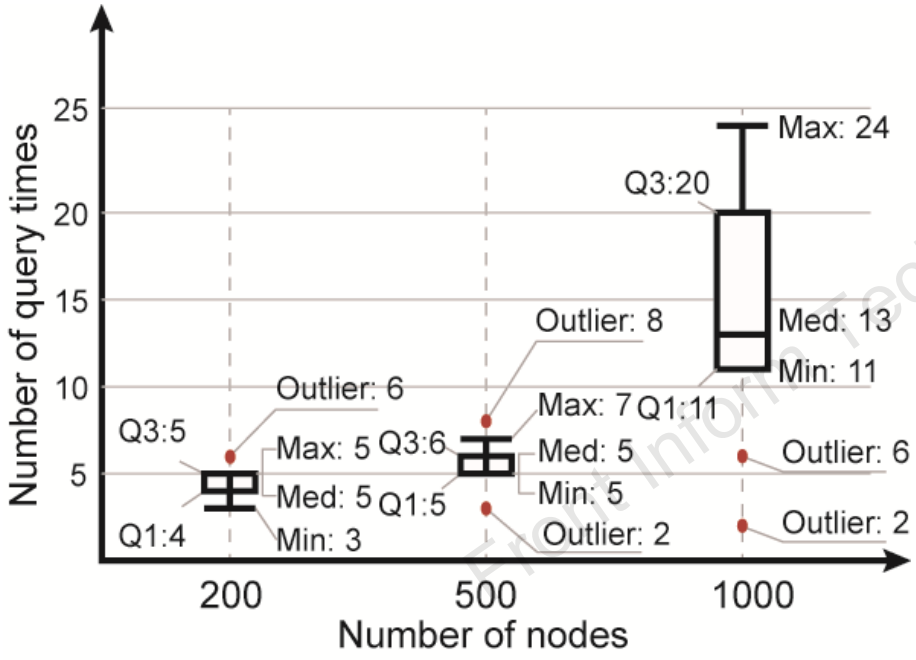




BIRD detects W. D., X. W., and H. L. between 2014 and 2015. Area (i) represents the compact neighborhood structures formed by them and their surrounding area in the sub-network view. Area (j) is the small community constituted by them and their surrounding areas in 2013. Area (k) is the same area as area (j) in 2014. Area (l) is a dense structure appearing beside area (k). Two nodes (area (m)) in area (k) have a lot of connections to nodes in area (l). Area (n) is the Sankey diagram which shows eight nodes in area (l) in 2014. Area (l) indicates the existence of a study with a lot of coauthors, which might be a result of multilateral cooperation. The abnormal changes of the surrounding areas of W. D., X. W., and H. L. make them a rare category

Major results (user study)

Query numbers of participants when labeling all rare categories in the datasets



Accuracy of labeling for four different special structures

Number of nodes	Clique	Bipartite graph	Star	Circle
200	91.67%	83.33%	83.33%	58.33%
500	83.33%	83.33%	100%	91.67%
1000	83.33%	91.67%	91.67%	83.33%
Average	86.11%	86.11%	91.67%	77.78%

Users gave some qualitative feedbacks:

1. The encodings in our matrices view were too complex.
2. The parameters were hard to comprehend.
3. Learnability and usability were both important problems which are hard to cover.

Conclusions

1. In this paper, we present RCAnalyzer, a novel visual analytics system which helps oracles analyze the results of RCD methods and label rare categories in dynamic networks.

2. We present a multi-focus tree cut algorithm and a tree-structure constrained lay-out optimization algorithm to support the comparison of instances in the context of their surrounding structures.

3.We use a use scenario and a user study to demonstrate the usability and effectiveness in analyzing rare categories in dynamic networks.