ACCDS: A Criminal Community Detection System Based on Evolving Social Graphs

Xiaoli Wang, Meihong Wang(✉), and Jianshan Han

Software School of Xiamen University, Xiamen, China
{xlwang, wangmh}@xmu.edu.cn, 919769245@qq.com

Abstract. This paper presents an intelligent criminal community detection system, called ACCDS, to support various criminal event detection tasks such as drug abuse behavior discovery and illegal pyramid selling organization detection, based on evolving social graphs. The system contains four main components: data collection, community social graph construction, criminal community detection and data visualization. First, the system collects a large amount of e-government data from several real communities. The raw data consist of demographic data, social relations, house visiting records, and sampled criminal records. To protect the privacy, we desensitize the real data using some data processing techniques, and extract the important features for profiling the human behaviors. Second, we use a large static social graph to model the social relations of all residents and a sequence of time-evolving graphs to model the house visiting data for each house owner. With the graph models, we formulate the criminal community detection tasks as the subgraph mining problem, and implement a subgraph detection algorithm based on frequent pattern mining. Finally, the system provides very user-friendly interfaces to visualize the detected results to the corresponding user.

Keywords: Evolving social graphs · Criminal community detection
Subgraph mining · Data visualization

1 Introduction

Nowadays, the community security management system has been widely used in China, and massive resident information has been collected, such as resident demographic data, visiting records, social relationships, criminal records, etc. Many existing works have focused on analyzing such complex data to detect anomaly events. The branch of data mining has wide applications in security, finance, and many others. These methods can be categorized into two groups based on various data models: high-dimensional method (e.g., [1, 2]) and graph based method (e.g., [3–5]). As graph is recently widely used to model real objects, this paper also focuses on the anomaly detection problem based on graph models. Different from existing work concerned with general criminal events, this paper focuses on some urgent real problems such as drug abuse behavior discovery and illegal pyramid selling organization detection. We develop a community data analysis system containing four main components. The components and our main contribution can be summarized as follows.

C. Woo et al. (Eds.): ER 2018 Workshops, LNCS 11158, pp. 44–48, 2018.
https://doi.org/10.1007/978-3-030-01391-2_10

- We collect massive real community data from some cities in China, and use several data cleaning techniques to obtain high-quality data. The sensitive profiles are desensitized before we store the data into the database.
- We use evolving social graphs to model the house visiting data, and employ a subgraph mining algorithm to solve the criminal events detection problem.
- We develop a powerful visualization system to display the evolving social graphs, the warning messages, and the detected criminal communities.

2 System Architecture and Demonstration

Our system is implemented based on J2EE platform and the system architecture is shown in Fig. 1. The system has four main components, including data collection, community social graph construction, data analysis, and result visualization.

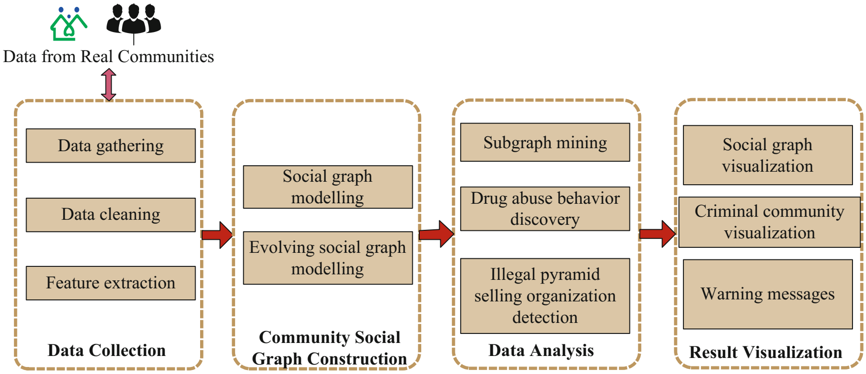

Data from Real Communities
Data gathering
Data cleaning
Feature extraction
Data Collection
Social graph modelling
Evolving social graph modelling
Community Social Graph Construction
Subgraph mining
Drug abuse behavior discovery
Illegal pyramid selling organization detection
Data Analysis
Social graph visualization
Criminal community visualization
Warning messages
Result Visualization

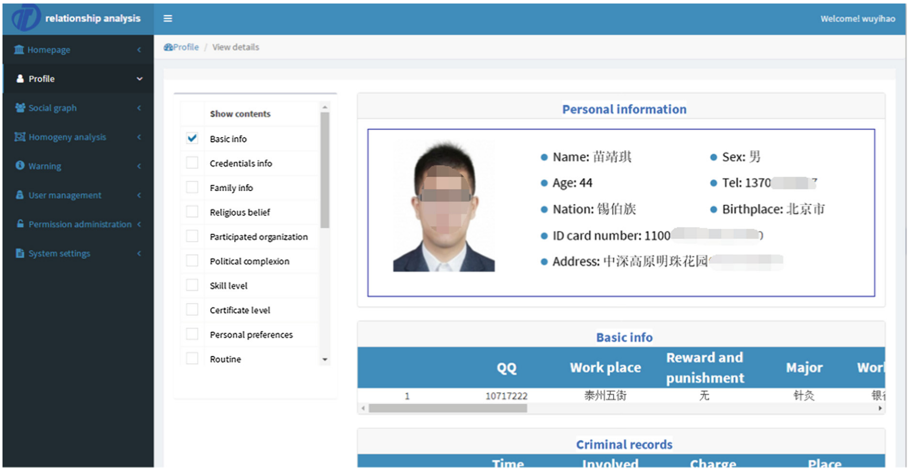
Fig. 1. The system architecture

2.1 Data Collection

We gather data from several real communities in China. The raw data contain massive information, such as resident demographic data, house visiting records, social relationships, and historical criminal records. We do data cleaning first to delete all the noisy data. Then, to protect residents' privacy, we desensitize the demographic data by removing explicit properties and randomly generating some synthetic profiles. Furthermore, we extract the most important features for profiling the human behaviors. The extracted profiles of a resident are shown in our system as in Fig. 2.

2.2 Community Social Graph Construction

We use a large static social graph to model the social relations of all residents, as shown in Fig. 3. In the social graph, each resident is represented as one node. If two residents have relationship such as family relation or friendship, there will be one edge being



Show contents

☑ Basic info
☐ Credentials info
☐ Family info
☐ Religious belief
☐ Participated organization
☐ Political complexion
☐ Skill level
☐ Certificate level
☐ Personal preferences
☐ Routine

Personal information

- Name: 苗靖琪
- Sex: 男
- Age: 44
- Tel: 1370[redacted]
- Nation: 锡伯族
- Birthplace: 北京市
- ID card number: 1100[redacted]
- Address: 中深高原明珠花园[redacted]

Basic info

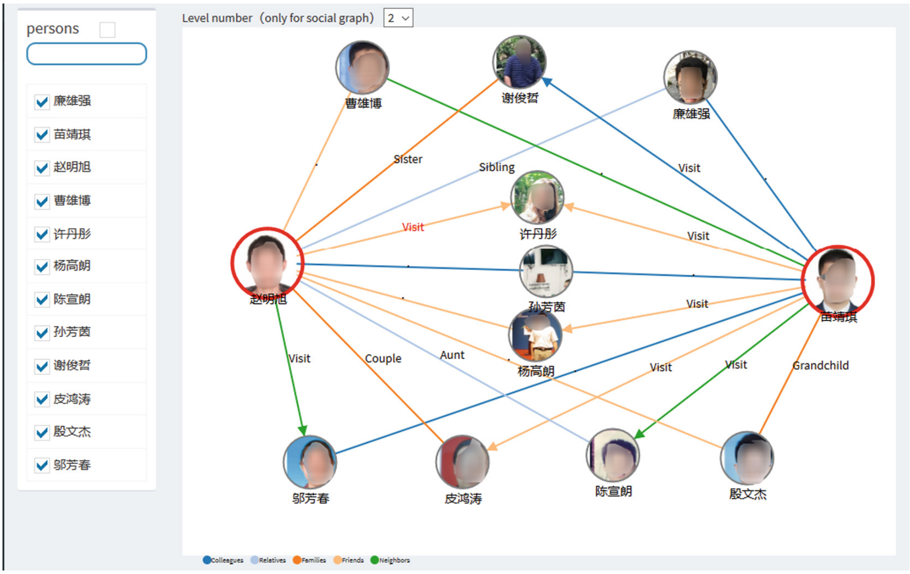
	QQ	Work place	Reward and punishment	Major	Wor
1	10717222	泰州五街	无	针灸	银

Criminal records

Time	Involved	Charge	Place

Fig. 2. The user profile page

connected between these two nodes. To support efficient criminal community detection, we use a sequence of time-evolving graphs to model the house visiting data for each house owner, as shown in Fig. 4. Different from the social graph, the edge in time-evolving graphs represent the visiting relationship. For example, if one resident of node A visited a house owner of node B, then there will be one edge connecting node A and node B.


persons ☐

☑ 廉雄强
☑ 苗靖琪
☑ 赵明旭
☑ 曹雄博
☑ 许丹彤
☑ 杨高朗
☑ 陈宣朗
☑ 孙芳茵
☑ 谢俊哲
☑ 皮鸿涛
☑ 殷文杰
☑ 邬芳春

Level number（only for social graph） 2

Fig. 3. The community social graph example for several selected residents

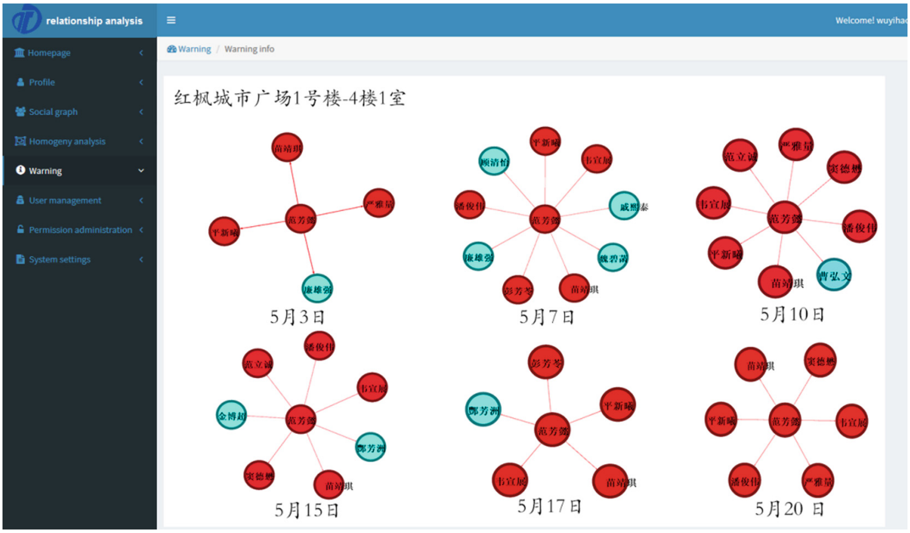


Fig. 4. The evolving visiting graphs for a selected address in May, and a highlighting result of the detected illegal pyramid selling organization (Color figure online)

2.3 Data Analysis and Visualization

Based on the graph models, we formulate the criminal community detection tasks as the subgraph mining problem, and implement a frequent pattern mining algorithm to solve it [6]. The details of the algorithm can be seen in our summited full research paper to a conference. We omit the details here as we will publish a technical report later online. As shown in Fig. 4, the substructures with red nodes and edges are the detected illegal pyramid selling organization. The system contains other pages to show other kinds of detected results. We omit them here for the space constraints.

Acknowledgment. This work is supported by the National Natural Science Foundation of China under Grant No. 61702432, the Fundamental Research Funds for Central Universities of China under Grant No. 20720180070, and the International Cooperation Projects of Fujian Province in China under Grant No. 2018I0016.

References

1. Gupta, M., Gao, J., Aggarwal, C.C., Han, J.: Outlier Detection for Temporal Data. Synthesis Lectures on Data Mining and Knowledge Discovery. Morgan & Claypool Publishers, San Rafael (2014)
2. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Data-driven process discovery - revealing conditional infrequent behavior from event logs. In: Dubois, E., Pohl, K. (eds.) CAiSE 2017. LNCS, vol. 10253, pp. 545–560. Springer, Cham (2017). https://doi.org/10.1007/978-3-319-59536-8_34

3. Akoglu, L., Tong, H., Koutra, D.: Graph based anomaly detection and description: a survey. Data Min. Knowl. Discov. **29**(3), 626–688 (2014)
4. Ranshous, S., Shen, S., Koutra, D., et al.: Anomaly detection in dynamic networks: a survey. Wiley Interdisc. Rev. Comput. Stat. **7**(3), 223–247 (2015)
5. Kaur, R., Singh, S.: A survey of data mining and social network analysis based anomaly detection techniques. Egypt. Inform. J. **17**(2), 199–216 (2016)
6. Kuramochi, M., Karypis, G.: Frequent subgraph discovery. In: IEEE International Conference on Data Mining, pp. 313–320 (2002)