ATTENet: Detecting and Explaining Suspicious Tax Evasion Groups

Qinghua Zheng[1,2], **Yating Lin**[1,2], **Huan He**[1,2], **Jianfei Ruan**[1,2] and **Bo Dong**[3,4]
[1]MOE Key Laboratory of Intelligent Networks and Network Security
[2]School of Electronic and Information Engineering, Xi'an Jiaotong University
[3]School of Continuing Education, Xi'an Jiaotong University
[4]National Engineering Lab for Big Data Analytics, Xi'an Jiaotong University
qhzheng@mail.xjtu.edu.cn, linyating@stu.xjtu.edu.cn, hehuan@mail.xjtu.edu.cn,
xjtu.jfruan@gmail.com, dong.bo@mail.xjtu.edu.cn

Abstract

In this demonstration, we present ATTENet, a novel visual analytic system for detecting and explaining suspicious affiliated-transaction-based tax evasion (ATTE) groups. First, the system constructs a taxpayer interest interacted network, which contains economic behaviors and social relationships between taxpayers. Then, the system combines basic features and relationship features of each group in the network with network embedding method node2vec, and then detects suspicious ATTE groups with random forest algorithm. Last, to explore and explain the detection results, the system provides a ATTENet visualization with three coordinated views and interactive tools. We demonstrate ATTENet on a non-confidential dataset which contains two years of real tax data obtained by our cooperative tax authorities to verify the usefulness of our system.

1 Introduction

Each year billions of dollars slip through the government as a consequence of tax evasion, which is detrimental to public welfare and services [IRS, 2012; Ferrantino *et al.*, 2012; Balafoutas *et al.*, 2015]. Among the many tax evasion methods for transferring profits, affiliated-transaction-based tax evasion (ATTE) is a new strategy that is carried out via legal-like transactions between a group of companies that have complex interactive relationships [Ruan *et al.*, 2019]. Since this tax evasion strategy is hidden in a group of companies, existing individual-based tax evasion detection methods cannot effectively detect such tax evasion companies.

To address this issue, we propose a network embedding based approach for detecting ATTE groups. First, since the ATTE strategy is implemented through corporate transactions between multiple companies in a group, the topological network of transactions captures these relationship features. Therefore, we construct a taxpayer interest interacted network (TPIIN) to describe a relationships between companies and related persons [Tian *et al.*, 2016] and divided this network into tax groups. Then, we use network embedding method node2vec to automatically extract the various relationship features of each tax group and use random forest algorithm to detect the suspicious ATTE groups. A detailed description of the ATTE detection algorithm can be accessed at https://doi.org/10.1016/j.ins.2018.11.008.

Although the above approach can detect the suspicious ATTE score of each tax group, the detection results are difficult to be explained to users. In particular, the values in relationship feature vectors have not physical meanings in real word, users cannot understand the relationship between these values and the actual structure in TPIIN. Therefore, inspired by existing studies on visualization of tax evasion systems [Didimo *et al.*, 2018; Goumagias *et al.*, 2018], we develop a visual analytic system based on our proposed approach, **ATTENet**, to show both the overview of TPIIN and the detection results of ATTE groups with visual explanation.

Our main contributions are as following: 1) We propose a network embedding based approach for detecting ATTE groups in TPIIN; 2) We present a novel visual analytic system based on our proposed approach for exploring the detection results and explaining the relationship features in ATTE group.

2 System Architecture

We describe our system illustrated in Figure 1. ATTENet consists of three components. The first component is TPIIN construction, which collects all the data of tax and taxpayer from official tax database and build the TPIIN network. This component also divides the TPIIN network into tax groups according to the relationships among persons and companies. The second component is the suspicious group detection. It extracts group features and detects suspicious groups. The last component offers visual exploration of the results of suspicious ATTE group.

TPIIN Construction. The tax data collected from our cooperative tax authorities includes the detail information of companies and the transaction records, such as company name, transaction amount, and industry category, etc. Since there are millions of companies and billions of transaction records in raw data, we develop a data preprocessing script to merge the transaction records of the same time period by the source company and target company to improve analysis performance. Then, we construct the TPIIN by using person and company entities in the tax data as vertices and generating edges between vertices based on the transaction data and and company data. As shown in Figure 1, four types of

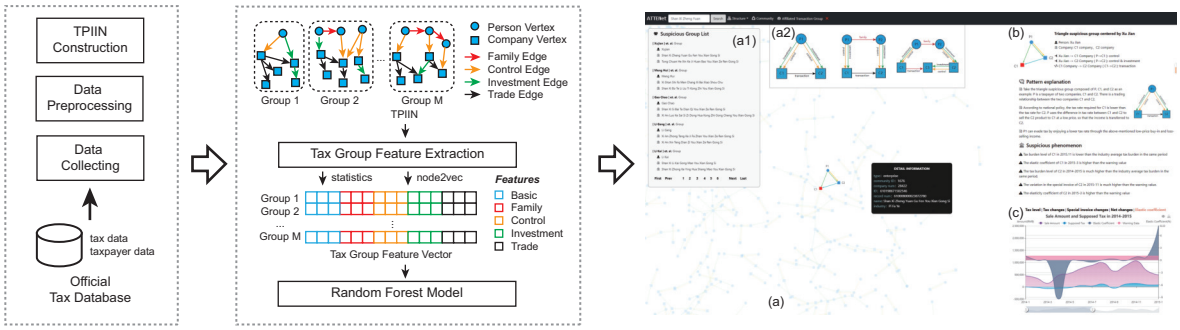

Figure 1: The blocks of ATTENet system

edges are generated to describe four relationships [Tian *et al.*, 2016]: (a) the person-to-person family edge; (b) the person-to-company control edge; (c) the company-to-company investment edge; and (d) the company-to-company trade edge. Moreover, the edges of type (a) and (b) have no weight, while the edges of type (c) and (d) are weighted by the amount of investment and trade.

Suspicious ATTE Group Detection. In this component, we first divide the TPIIN into tax groups according to the relationship type between each vertices. Each tax group is a subgraph of TPIIN with no edges between other groups. Then, we extract five types of features to represent each tax group. The basic features include include basic information and statistics on members of the tax group, such as registration time and total registered capital. As color-encoded legends in Figure 1, the other four types of feature describe the inner relationships in the group, which correspond to the four edge types in the TPIIN. We use node2vec to learn the representation vector of each relationship respectively, and then combine the representation vectors and the basic features as the tax group feature vector. The tax group feature vector is used as input to random forest algorithm to get the suspicious ATTE score of each tax group.

ATTE Groups Visualization. To facilitate users to intuitively explore the huge TPIIN network and analyze the results suspicious ATTE group detection, we two coordinated views: the TPIIN overview (Figure 1(a)), the explanation view(Figure 1(b)), and the statistics view (Figure 1(c)). In the TPIIN Overview, each circle represents a person vertex, while each square represents a company vertex. The edges are color-encoded by the four types of edges. When clicking on any element of a tax group, the corresponding details will be displayed in other two views. In addition, users can browse the suspicious group list by specifying the relationship pattern (Figure 1(a2)). The explanation view and the statistics view show the detail information of the selected suspicious ATTE group. Each panel in these two views depicts the abnormity phenomenon of this tax group, such as suspicious profit chain, tax burden level, tax variation, net price, etc. In the explanation view, the trading company's abnormal information is displayed, which can help users to check the abnormal relationships with other companies in the suspicious group. To facilitate exploration, the ATTENet also support checking detail information of every visual element in the TPIIN overview by clicking on the vertices.

3 Demonstration

We demonstrate ATTENet through a live demonstration with a case of detecting suspicious ATTE groups in non-confidential dataset.

First, the user can select a suspicious ATTE group in the "Suspicious Group List" (Figure 1(a1)) or select a basic relationship pattern (Figure 1(a2)) to update the list. Then, the user can view the selected group in the TPIIN view to see the relationships in this tax group (Figure 1(a)). Next, the user can check the detail information of each element in this tax group, including the persons' information, companies' information and the detailed relationship between each element (Figure 1(b)). Lastly, the user can check the statistics on each company of this tax group to understand the basic features of a tax group.

4 Conclusion

We present ATTENet, a novel visual analytics system for detecting and explaining the suspicious ATTE groups. It extracts both tax group features and relationship features and detects ATTE groups with node2vec and random forest algorithms. It enables research and practice on tax evasion group detection and visual explanation to the detection result.

Acknowledgments

This research was partially supported by "The Fundamental Theory and Applications of Big Data with Knowledge Engineering" under the National Key Research and Development Program of China with Grant No. 2018YFB1004500, the MOE Innovation Research Team No. IRT17R86, the National Science Foundation of China under Grant Nos. 61721002 and 61532015, the Project of China Knowledge Centre for Engineering Science and Technology, and the consulting research project of Chinese academy of engineering "The Online and Offline Mixed Educational Service System for 'The Belt and Road' Training in MOOC China.

References

[Balafoutas *et al.*, 2015] Loukas Balafoutas, Adrian Beck, Rudolf Kerschbamer, and Matthias Sutter. The hidden costs of tax evasion.: Collaborative tax evasion in markets for expert services. *Journal of Public Economics*, 129:14–25, September 2015.

[Didimo *et al.*, 2018] Walter Didimo, Luca Giamminonni, Giuseppe Liotta, Fabrizio Montecchiani, and Daniele Pagliuca. A visual analytics system to support tax evasion discovery. *Decision Support Systems*, 110:71–83, June 2018.

[Ferrantino *et al.*, 2012] Michael J. Ferrantino, Xuepeng Liu, and Zhi Wang. Evasion behaviors of exporters and importers: Evidence from the U.S.–China trade data discrepancy. *Journal of International Economics*, 86(1):141–157, January 2012.

[Goumagias *et al.*, 2018] Nikolaos D. Goumagias, Dimitrios Hristu-Varsakelis, and Yannis M. Assael. Using deep Q-learning to understand the tax evasion behavior of risk-averse firms. *Expert Systems with Applications*, 101:258–270, July 2018.

[IRS, 2012] IRS. IRS Releases 2006 Tax Gap Estimates | Internal Revenue Service, 2012.

[Ruan *et al.*, 2019] Jianfei Ruan, Zheng Yan, Bo Dong, Qinghua Zheng, and Buyue Qian. Identifying suspicious groups of affiliated-transaction-based tax evasion in big data. *Information Sciences*, 477:508–532, March 2019.

[Tian *et al.*, 2016] F. Tian, T. Lan, K. Chao, N. Godwin, Q. Zheng, N. Shah, and F. Zhang. Mining Suspicious Tax Evasion Groups in Big Data. *IEEE Transactions on Knowledge and Data Engineering*, 28(10):2651–2664, October 2016.