THE XXX OF XXX

Contents

1 Introduction

1.1 Background

In a information-based society, the development of data network is getting increasingly faster. As a result, the relations between people who are involved into the social network become more complicated. In order to methodize the massive data in the network, the network analysis model comes into being. These models of connection between people describes mutual relations in the aspect of point set, entropy and macro-statistics, which largely simplify the sophisticated social phenomenon and provides a way to forecast the future development. Now the network analysis model has been widely applied to analyse social individuals' connections, comprehend the social structure[1][2] and investigate the communication system in large group of organization[3][4].

Among them, crime network receives particular attention in de-

termining the criminal organizations[5]. This paper mainly focuses on identifying criminal conspirators on the background of a conspiracy in a software company.We are going to find out the potential criminals with provided pieces of evidence.

1.2 Out Work

2 Assumptions

We make these following assumptions about the criminal data in this paper.

- We assume that the resource of message traffic is reliable and all the information is true.
- In order to simplify the model, we neglect the time factor in conversations.
- According to the basic information we receive, we assume that only one single criminal organization exists in the process of our model construction.

3 Symbol Descriptions

4 Crime Detection

After researching several papers in relevance of crime network field, we draw some regular common conclusions on this special kind of network[?]

1. The diameter of crime network usually isn't great. Thus, most networks enjoy a high density.
2. The level of information sharing is equal.
3. Most networks have radial center.

With Feature (1) and Feature (2) we can analyse the criminal potentiality in the crime network. Besides, Feature (2) can help us to identify the conspiracy leaders in the whole relations. Through these common features, we apply these into practical problems and establish the following models to tackle with the problem. Before modeling, to avoid ambiguity, we define the two concept here:

1. Suspicion parameter: we employ to measure the suspicious degree of one node.
2. Crime centrality: This parameter is used to quantity one node's closeness to the criminal center.

4.1 Model One: Criminal Digging Algorithm Based on Connection Law

Connection law is a method to determine the high frequency of communication between two people under the circumstances of concentrative distribution of data. In this case, the investigators has formed the people and message network. Thus, we are able to carry on a series of deeper data mining. In order to find suspicious degree of each person in this company, according to graph theory, we define the nodes as people and the sides as the communication among these people. Thus, we establish network model which can reflect the relation among people in this company. To evaluate the suspicious degree of each peson in the network model, we discuss its measuring method in the following section.

4.1.1 Suspicion Detection via the importance of node in the network

In this model, we introduce the conception of potential relation degree with the conspirators as A_{ij}. Through the number of A_{ij} nodes, we can build the connection path from node i to node j. The proportion of indirect connection of nodes is the proportion of sides with same nodes and the joint nodes, which can be expressed as:

$$A_{ij} = \frac{P_i \bigcap P_j}{P_i \bigcup P_j} \tag{1}$$

$$AA_i = \frac{1}{n}\sum_{j=1}^{n} A_{iS_j} \tag{2}$$

In this equation, O is the row vector with each value being 1, P_i represents relevant vectors of node i. refers to the number of node which represents the people in the network. However, after reviewing the cases provided by our supervisor, we notice that the nodes which enjoy high level of connection with others in the network don't directly mean that they are under high level of suspicion. Take Carol for example, although he has direct conversations with most people in the company, the relevant topics are not about the known suspicious topics. The single use of connection degree to measure the importance of node may indicate errors. Therefore, we can conclude that the potential connection with criminals can represents the direct connecting ability with its suspicious node, but it cannot perfectly reflect the criminal message he receives .As a result, we decide to use the information similarity to evaluate one person's suspicious degree.

4.1.2 Determining Suspicion Degree with Information Similarity with Conspiracy Plotting

Here,we introduce several parameters to help us measure the suspicion degree.We use $W(k)$ to represent the weight of topic k.To help us quantity the value of connection, we use the function

send(i,j,k), which is defined as follows: If node i sends messages about topic k to node j, then the value of send(i,j,k) is 1, or it is 0.

$$W(k) = \sum_{i=1}^{n} \sum_{j=1}^{n} send(i,j,k) \tag{3}$$

And, We employ the concept of $informationsimilarity$ as B_{ij}. It is defined as:

$$B_{ij} = \frac{(V_i \bigcap V_j) \times W^T}{(P_i \bigcup P_j) \times W^T} \tag{4}$$

V_{ij} represents the node i receives or sends topic j. So V_i can express the relevant eigenvector of node i concerning all the topics. In the same way as the calculation of AA_i, we can get the value of BB_i, which describes mean of the information similarity.

4.1.3 Topsis Algorithm

Topsis algorithm see the parameters which may affect the results of the evaluation of N indicators as the N axis, constructing an $N-dimensional$ space, and then select the best value and the worst value of the index from all the objects to be evaluated. Then we use it to depict two points in N-dimensional space, which are respectively the best node and the worst node. Then we can calculate each the distance of objects to be evaluated to both worst and best points via the coordinates of the points, expressed as D_{best} and D_{bed}.

The smaller the value of result is, the better evaluation result one index has. In order to minimize the errors in analysis, we use the two index of suspicion parameter, AA_i and BB_i, and employ the Topsis Algorithm to determine the final suspicion parameter S.The results are shown in Table(???)

4.2 Model Two: Conspiracy Leader Model

As is expressed in the beginning of this part, the concept of criminal centrality is used here. Here, two parameters which respectively describes the closeness of connection of the nodes and the edges are introduced here as CE_i and CN_i. In the network built on the conversations of suspicious topics, we choose the child network whose center is node i and calculate its value of CE_i and CN_i.

$$CN_i = \frac{\sum CC_j}{sum_{k=all} CC_k}, dis(i,j)/le1 \tag{5}$$

And the calculation method of CE_i is the same as CN_i.

5 Semantic Network Analysis

Because the leader provides us with the text evidence about the case, we introduce semantic network analysis to obtain and understand this message traffic. Based on the conversation we

accumulate, we abstract the key words in each part. Then we sort out the simple semantic network with these key topics. The text connection of the people who have been identified are shown in

6 Strengths and Weaknesses

7 Model Design

呵呵

8 Conclusions

8.1 Strengths and Weaknesses

1. Strengths:

   (a) We take both the node position in the whole network and suspicious message it sends and receives into consideration for identifying and prioritizing conspirators. So, the solution of our model pursues high credibility, while reducing the misjudgment rate.

   (b) The result of simulation shows that our model can be applied to other field, so it is extendable.

(c) The result of our model match perfectly with the experience, which proves the reasonability of our model .

2. Weaknesses:

(a) Since there is no clear criteria for the classification, those conspirators who are slightly behind may be missed .

(b) We don ‘t take the Criminal Psychology into consideration while to mix the detective’ s judgement,what some people say may not true.

(c) In actual situations,criminal gang may be more handsome and the hidden boss may be hidden deeply so that our model has possibility to loss them.

3. Future work:

(a) Apply semantic network analysis to discover the potential linkage between the messages and scientifically classify them into different groups.

(b) Find a more reasonable criteria for the classification and consider the Criminal Psychology for dealing with more complex and specfic cases.

9 Model Extension

In our model , we use many network techniques to empower our model. And in common network analysis, we often use indexes about nodes and edges . What's more ,in a specific network, there are always related information about nodes and edges . We both use the common and feature of network to build our model, with the goal that making the estimating result more accurate.While dealing with the problem of crime busting, we take some specific features such as suspicious topics and persons .It is the same with similar problems like assuming the infection ability of different cells. Based on these conclusions above, we state a general approach, by which can identify, prioritize, and categorize similar nodes in a network as follows:

- First, collect datas of various properties of nodes in the network and analyze them to estimate general characteristics of that network.
- Then, according to features of specified network, transform them into indexes about nodes and edges just like what we do in our model.
- Finally, combine both of them together into a weighted average. And that is a score mesures nodes with the given standard.

With this score of result, we can identify, prioritize, and categorize similar nodes in that network.

References

[1] XU JJ, CHEN HC. CrimeNet Explorer: A Framework for Criminal Network Knowledge Discoveryp[J].ACM Transactions on Information Systems, 2005, 23(2).

[2] BERKOWTTZ SD. An Introduction on Structural Analysis : The Net work Approach to Social Research[M].Butterworth, Toronto, 1982.

[3] BREIGER RL. The analysis of social networks[A]. HARDY MA, BRYMAN A, eds. Handbook of Data Analysis. Sage Publications, London, UK, 2004.505 - 526.

[4] GARTON L, HAYTHORNTHWATTE C, WELLMAN B. Studying online social networks[A] . JONES S, Ed. Doing Internet Research[C]. Sage Publications, London, UK, 1999.75 - 105.

[5] 唐常杰，刘威，温粉莲，乔少杰。社会网络分析和社团信息挖掘的三项探索——挖掘虚拟社团的结构、核心和通信行为[J].
计算机应用，2006-2020-04

[6] 董宸，钱存乐，马建钧。基于社会网络分析李文的犯罪网络侦测方案设计[J]. 数学建模及其应用，2012-0072-11