Non-conspirator？Conspirator ！

Content

1 Introduction

1.1 Background and Analysis of the Problems

With the development of crimes, organized crimes appeared. Those people who are involved in these kinds of crimes always connected with each other because of the same goal and some restrictive factors. These crimes bring more harm to the society than individual crimes. Our organization, ICM, is investigating a conspiracy. The 83 people who are investigated are members of a software company, which mainly markets with banks and credit card companies. At present, ICM has already known some information. Those investigators think that the information will help them to find out the most-possible people selected of the ambiguous conspirators and unknown leaders. The goal of molding is to find out who are the most-possible conspirators in the complicated office.

- According to those known conspirators, 15 topics (three of which have been deemed to be suspicious), and 400 message links, our goal is to assure who are conspirators, who are leaders, prioritize the 83 nodes by likelihood of being part of the conspiracy, and determine a discriminate line separating conspirators from non-conspirators.
- We should also think about what to do if some of those known conditions changed (According to the text, Chris becomes the conspirator, Topic one becomes suspicious.) We should get what changes the result finally has.
- When dealing with the problem, we get more information (the original massage text), it is needed to be known how our model will finally be enhanced.
- When the model is used more widely, it must be made sure that it can be applied into any condition.

1.2 Methods of the Analysis

As investigators, we now know well about the 83nodes, 400 links over 21,000 words of message traffic, 15 topics (three have been deemed to be suspicious), 7 known conspirators, and 8known non-conspirators. We now call the known information about the company INTELLIGENCE. From the practice, the method which is always used in the intelligence work is a combination of qualitative reasoning and quantified reasoning, that is called Analytic Hierarchy Process (called AHP in the later part).

At the same time, we also use Structural Model Analysis (SMA).

1.3 Assumptions

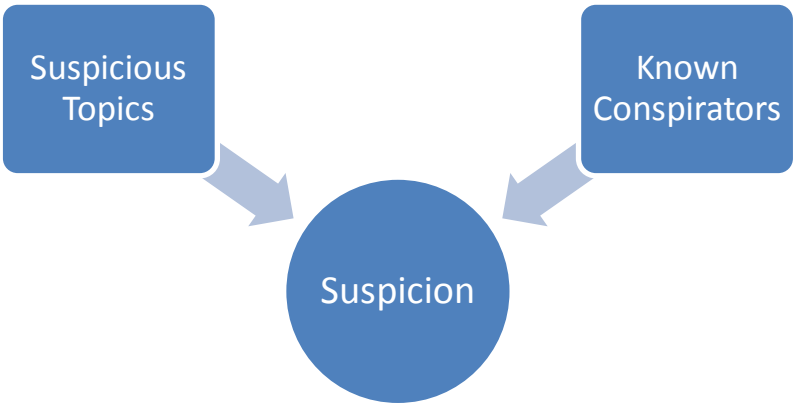
- There is only linguistic communication between people. Body language never exists.
- People can talk to each other freely, without the limit of distance.
- The information of known conspiractors and non-conspiractors is correct.
- The talk happens only between two people. Talks between 3 or over 3 members never exists.

- Suspected topics are only those claimed in the text. Other topics never become suspected.
- There is no false information.
- The suspicion is never influenced by gender, age and any other personal information.

2. Model Approach

2.1 The confirmation of the indexes judging one's suspicion

As a group of conspirators, they connect each other through communication, so there must be interaction of information between them. The three suspicious topics are also included in the information. Considering suspicious topics and conspirators, we consider that the indexes judging the suspicion are the three points below:



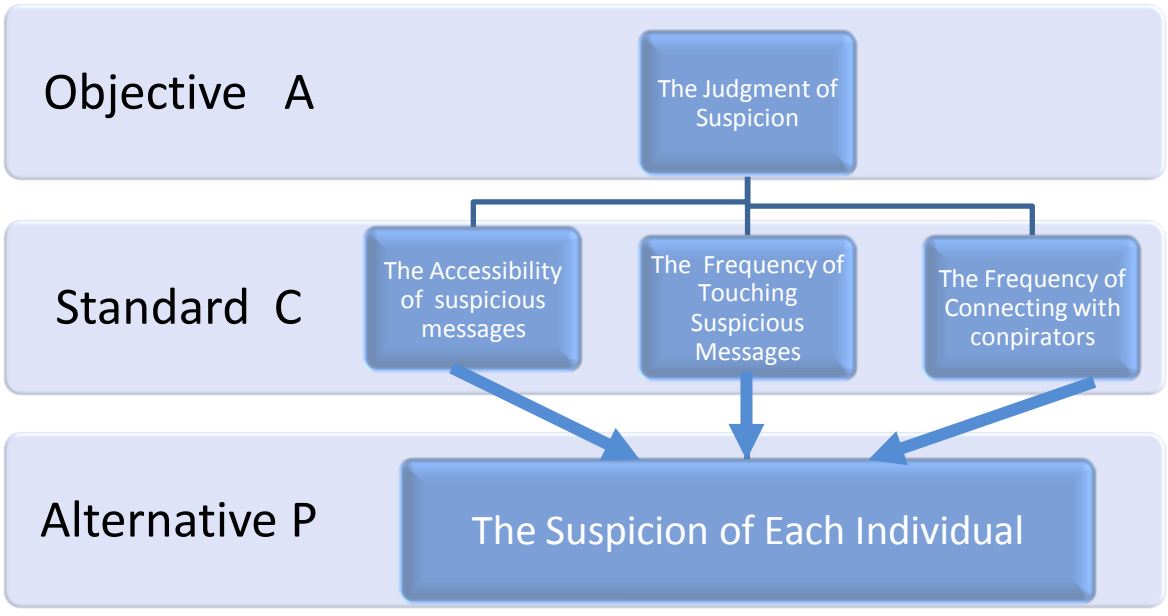
①The frequency individuals send and receive suspicious topics.

②The frequency individuals interact with conspirators.

③The accessible level to suspicious messages.

2.2 Molding Using AHP

2.2.1 The Establishment of the Model



2.2.2 The Construction of the Judging Matrix

2.2.2.1 Judging Matrix A-C (Compare the Significance of Each Standard In Order to Judge the Suspicion)

Except for the special circumstances, people in the same company communicate with each other unavoidably. Therefore, the significance of the connecting with conspirators is the least important standard, while the accessibility is the most important. We consider the order of the significance of the three standards is: $C_1>C_2>C_3$. Thus the judging matrix A-C is:

A	C_1	C_2	C_3
C_1	1	2	4
C_2	1/2	1	3
C_3	1/4	1/3	1

In the judging matrix, the meanings of the values of A_{ij} are as follows:
A_{ij}=1: Ai is as important as A_j;
A_{ij} =2: Ai is a little more important than A_j (suspicion)
A_{ij} =3: Ai is more important than A_j;
A_{ij} j=4: Ai is extremely important than A_j.

Either factor in the matrix (A_{ij})stands for how more important A_i is than A_j.
According to the document, it can be calculated by the formula below:
Normalization of each line of the judging matrix:.

$$\overline{A_{ij}} = \frac{A_{ij}}{\sum_{k=1}^{3} A_{kj}}, i, j = 1,2,3$$

Sum up each judging matrix that has already been normalized in each row:

$$\overline{W_i} = \sum_{j=1}^{3} \overline{A_{ij}}, j = 1,2,3$$

.

Normalize the vector:. $W = \frac{\overline{W_i}}{\sum_{j=1}^{3} \overline{W}_j}, i = 1,2,3,$

Thus the characteristic vector can be got:
$W = (W_1, W_2, W_3)^T = =(0.5584 \quad 0.3196 \quad 0.1220)^T$.
Then the characteristic root can also be calculated:

$$\lambda_{\max} = \sum_{i=1}^{n} \frac{(AW)_i}{nW_i} = 3.0183, \quad BW = \lambda_{\max} W$$

., $(AW)_i$ stands for the i-th component of vector AW.

The last step is consistency check. The formula is $CI = \frac{\lambda_{\max} - n}{n-1}$, in which RI stands for the index of average random consistency. It is in the chart below:

Order	1	2	3	4	5	6	7	8	9
RI	0.00	0.00	0.58	0.90	1.12	1.24	1.32	1.41	1.45

the average random consistency index of the first order to ninth order matrix

CR = CI/RI=0.0176<0.10, thus the matrix gets the satisfied consistency.

At last, the weight of each standard, which is the value of characteristic vector. C_1 is 0.5584. C_2 is 0.3196. C_3 is 0.1224.

2.2.2.2 Judging Matrix C_1-P (Compare the relative level of suspicion only considering the accessible level of suspicious messages from each node.)

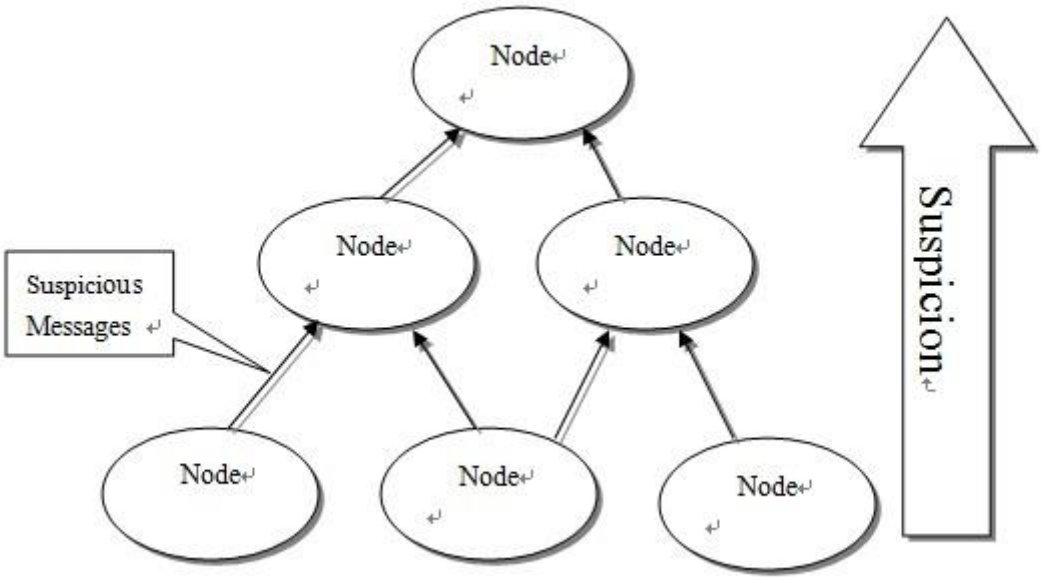
The accessible level of suspicious messages is the one's frequency of receiving suspicious messages. Apparently, the more suspicious access one receives, the more people pass suspicious topics to him or her. So this node gets more suspected degree. Therefore, the suspected level can be characterized by each of the factors and Y_j in the accessible matrix. Those Y_js are got by structurally analyzing the accessible matrix.

It can be seen in the picture. The adjacency matrix can be worked out through how those messages are converted. "L_{ij}=0" means suspected messages cannot be converted to j. "L_{ij}=1" means suspected messages can be converted to j. The matrix I is an 83*83 unit matrix.

$R = I + L + L^2 ... L^n = (I + L)^n$, I is unit matrix.

Make out the sum of each line: $\mathrm{Yj} = \sum_{i=1}^{83} Rij$

Y_j and the factors in the accessible matrix characterizes the amount of people who can convert suspected messages. By using the result the suspected degree can be characterized.



There are 83 objectives in this model. These 83 objectives can be sorted into five sets by considering factors in line and Y_j. It is considered that every person in the same set have the same suspected level. For those people who are in the different sets, the more the Y-value of the set they are in, the more their suspected degrees are.

And the more the Y-values different, the more their suspected degrees differ form each other.

It is defined that b_4 is the relative importance between the i-th (i=1,2,…,83) objective and the ht j-th (j=1,2,…,83) objective. It is recorded as C1-P=(bij)83*83. b_{ij}=1, b_{ji}=1/b_{ij}.

b_{ij} in the judging matrix is valued as follow:

b_{ij} =1: means B_i is of the same suspected degree as B_j;

b_{ij} =3: means the suspected degree of B_i is lightly more important than that of B_j;

b_{ij} =5: means the suspected degree of B_i is more important than that of B_j;

b_{ij} =7: means the suspected degree of B_i is largely more important than that of B_j;

b_{ij} =9: means the suspected degree of B_i is extremely more important than that of B_j.

The judging matrix C_1-P is as follow:

C $_1$-P	**B_1**	**B_2**	**…**	**B_{83}**
B_1	**b_{11}**	**b_{12}**	**…**	**$b_{1,83}$**
B_2	**b_{21}**	**b_{22}**	**…**	**$b_{2,83}$**
...	...	...	**…**	...
B_{83}	**$b_{83,1}$**	**$b_{83,2}$**	**…**	**$b_{83,83}$**

Then make the hierarchical order separately.

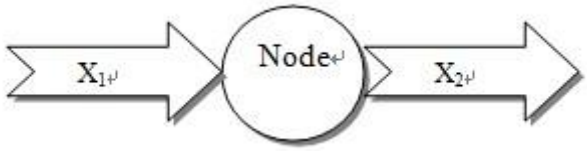
According to 2.2.1.1, the characteristic vector can be worked out. Wn, the value of this vector, cam indicate the suspected degree of n, using which can make the order.

2.2.2.3 Judging Matrix C_2-P(Compare the relative suspicion of each people thinking about only the connect of suspected messages.)

The frequency individuals send and receive suspicious topics

Define: $X_i=X_1+X_2$, i=1,2 , … , n ,

X_1 indicates how many links（ involving any known suspicious message topics)starts at this node , X_2 indicates how many links（ involving any known suspicious message topics ）ends at this node



We consider that suspicious topics are only No.7, No.11, and No.13. In the next step, we should screen out all the messages that talks about either one of those suspicious topics, and then calculate X, indicating the intensity of contact between every node and either suspicious messages.

We assume that suspected topics are only No.7, No.11, and No.13. Firstly, using the 400 messages between these people(links), we can screen out messages including

either suspected topic. We can then make out how many times each node sends and receives messages, standing for the level each node connects with suspected messages. The sum is recorded as X.

The next step is to divide the 83 objectives into five sets according to the sum of sending and receiving of each node, just like what has been done in 2.2.2.2. Then we should list the judging matrix C_2-P, and then do what we have done in 2.2.2.1 to make out the characteristic vector W. cengcidanpaixu!!

2.2.2.4 The Judging Matrix C_3-P (Compare the relative suspicion of each node thinking about only the connect of suspected messages)

So far, it is known that Jean, Alex, Elsie, Paul, Ulf, Yao and Harvey are conspirators. Therefore, we can make out the amount of interactions between every node and every known conspirator, using the 400 messages between these people. The amount is called Z, which means the interactive level between individuals and conspirators. Therefore, we consider that the more frequently nodes connected with conspirators, the more suspicious they are.

The next step is to divide the 83 objectives into five sets according to the sum of sending and receiving of each node, just like what has been done in 2.2.2.2. Then we should list the judging matrix C_3-P, and then do what we have done in 2.2.2.1 to make out the characteristic vector W. Then make the hierarchical order separately.

2.2.3 The Total Hierarchical Order

From 2.2.2.1, we know that the value of those weights. C_1 is 0.5584. C_2 is 0.3196. C_3 is 0.1220. Supposed that the weight of one's suspicion of C_1 is W_{C1}, C_2 is W_{C2}, and C_3 is $W_{C3.}$ We can make out the weight of the total order by the formula below.

$$U = W_{C1} * 0.5584 + W_{C2} * 0.3196 + W_{C3} * 0.1220$$

We can make out the final suspicion order according to U.

2.3 Conclusions

2.3.1 The Final Suspicion Order of the Eighty-three Nodes:

According to the value of U, we can make the list of the suspicion of 83 nodes, as the chart following.

Rank	Node									
1~10	21*	67*	7*	54*	10	43*	17	13	18*	49*
11~20	81	3	38	50	6	32	16	30	44	4
21~30	11	42	34	28	20	37	15	2#	47	22
31~40	48#	36	41	19	29	46	33	35	31	14
41~50	5	80	27	82	65#	12	57	75	45	8
51~60	9	40	60	69	24	79	23	39	51	56
61~70	72	78#	1	77	26	52	53	55	68#	58
71~80	59	70	74#	25	66	73	0#	61	62	63
81~83	64#	71	76							

(* indicates prior known conspirators, # indicate prior known non-conspirators)

2.3.2 Discriminate Line Separating Conspirators from Non-conspirators.

Do cluster analysis of the U of 83 nodes,the result of which indicates that the 63 nodes of lower suspicion, so that the dividing line should be set up between No.4 Gretchen and No.11 Francis.

2.3.3Who Is the Leader

According to the speculation, Dolores is the leader. Firstly, his suspicion is very big. Secondly, he is a manager. He communicates with others is more frequently than others.

2.4 Model Testing

2.4.1 Sensitivity Testing

In the new condition that Chris becomes the conspirator and Topic 1 becomes suspicious, we can use the model above to get the suspicion rank of these 83 people, as the chart following,

Rank	Node									
1~10	48#	34	21*	67*	7*	54*	5	43*	17	18*
11~20	2#	10	32	13	15	22	49*	20	81	47
21~30	3	28	31	41	50	4	14	27	44	0*
31~40	6	19	37	38	16	9	11	29	30	33
41~50	45	46	25	36	69	1	12	24	42	35
51~60	82	56	80	66	57	63	75	40	65#	8
61~70	60	68#	23	26	39	51	55	62	64#	71
71~80	72	78#	79	61	73	52	53	58	59	70
81~83	74#	76	77							

(* indicates prior known conspirators, # indicate prior known non-conspirators)

Analysis: According to the list, we can get that the ranks of some people has little changes. Some people get big changes in their ranks, who are close connected with Chris and Topic 1. It substantiates that the model has high sensitivity.

2.4.2 Accuracy Testing

Using the model to deal with EZ case, we can get the suspicion rank as following.

George*	Dave*	Ellen	Harry	Bob	Carol	Fred	Inez	Anne#	Jaye#

(* indicates prior known conspirators, # indicate prior known non-conspirators)

Analysis: Bob ranks fifth, while Carol ranks sixth. It mostly coincides with the eventual crime condition, which substantiates that the model is accurate.

3. Model Enhancement

3.1 Semantic Network Analysis

For the third requirement, it is mentioned in the former part of the essay that the team get a more powerful method (semantic network analysis). In this condition, we can do a lot to enhance our model. Semantic network analysis is a kind of idea in artificial intelligence. According to the description in the published documents, we get that the most important factor related to SNA is word frequency of the text of those messages. Therefore, we conclude the enhanced model and some significant points as following.

- We should first of all manage the statistics of the word frequency and make a list of it from the highest to the lowest.
- Through the frequency list, many of the words are meaningless, such as "I", "you", and "to". In the next step, we should clear away all of these null words. After it, the further selection should be connected with those topics. Then words in the new list are all about what nodes are talking about.
- After working out the new word frequency list, we should estimate the weight of those words in the new list. When doing the new estimate, words ranking higher get more weight than those of lower ranking.
- After getting the new weight, we can eventually work out the list of the suspicious degree. Then the result will be more accurate.

3.2 Application of the Model

For the forth requirement, it is claimed that our team should further develop the model so that our model can be used no matter what the condition is. We conclude the points as below.

- Firstly, we should do cluster analysis. After itemizing those messages into several different parts, we can use our model to deal with the problem easily. So no matter how complicated the data are, they can be categorized into different parts reasonably.
- If the problem involves some factors such as people's thoughts and emotions, we can still use the model but just add more standards when establishing the model. For example, if it should be considered in the model about nodes' emotions, we can set the parameters ourselves. We can define some words as happy emotions and some as sad ones, some as satisfied emotion and some as unsatisfied ones.
- If there are some information about semantics, network, and messages, we can assure more parameters, which means weight, by deeply analyzing them.
- As for many other network problems, if there's any polytropic factors in it, we can deal with it by valuing those parameters to assure the each of those weights, just like the model above in this essay.

4. Conclusions

4.1 Strengths

- Each weight in AHP can influence on the result directly or indirectly. The

influence of every factor is quantized. The model is systematic, and very clear.

- The decision method is simple and practical. The method combines quantitative one and qualitative one, to resolve the complicated system to make problem easier. Also, the model can convert multi-objective problems into single-objective ones.
- What's more, the combination of SMA and AHP enhances the model greatly.

4.2 Weakness

- It is of greatly subjective. This mostly makes errors. And the model will not be that convincing. Especially when the problem is about human, or full of human emotions, the results will be particularly un convincing.
- When the data are great, we need more hierarchies to deal with them. For example, when we have many index, we may have great difficulties in judging the significances between two index.

References

[1] Lidong Sun1 Yunfa Xiong2 , Showing Analysis Method of Public Security Intelligence, Journal o f Chinese People`s Public Security University,2007, Sum52.

[2]Xinxian Liu , Options and Judgement—AHP Decision-Making , Shanghai Science press ,1990

[3]Youjun Zheng , Criminal Psychology, Sichuan University press,2008

[4]http://wenku.baidu.com/view/b21baa4a767f5acfa1c7cd17.htm

Interpretive Structural Modeling