E-judge: Application of AHP in Identifying Conspirators.

Content

1 Introduction

1.1 Restatement of the problem

In the ICM question of year 2012, we are required to build a model that could lead the procurator to the most suspicious work staffs that have committed a criminal act within a company.

The data we have are: the name list and index numbers of all 83 staffs, the message list that record all 400 message links of the work staffs as well as the topic list that summarizes 15 topics contained in the staffs' messages. The mission of our team is to use those data to carry out the suspect list as well as the names of the criminals who are likely to be the leaders of this conspiracy. To help us build an accurate model to solve this case, our supervisor provide us a scenario she worked on a few years ago that is similar but much smaller than the current case we are dealing with.

1.1.1 Definitions and parameters

Table 1

Definitions and Parameters	Denotation	Definitions and descriptions
Suspicious information		Message content that is likely to be part of the conspiracy
Suspicious message		Messages that contains suspicious information
i	Index of staff	The node number of a specific staff
L	Index of level	The index number of a set of messages, which are considered to have the same level of conspiracy
O_i	Outgoing message Amount	The amount of messages a staff send to others
I_i	Incoming message Amount	The amount of messages a staff received from others
T_i	Total message amount	The amount of messages a staff sends or receives
O_{ci}	Message amount sent to conspirators	The amount of messages of a staff that are only sent to the known conspirators
I_{ci}	Message amount received from conspirators	The amount of messages of a staff that are only received from the known conspirators
T_{ci}	Total message Amount with	The amount of messages of a staff that are only sent to or received from the known

	conspirators	conspirators
T_{Li}	L level messages Amount	The amount of a staff's L level messages
T_{cLi}	L level messages amount with conspirators	The amount of a staff's L level messages that are only sent to or received from the known conspirators
W_L	Weight of level L	The weight of an L level of messages that should be added in calculation
$\sum_L W_L \bullet T_{Li}$	Suspicious information amount	The amount of suspicious information contained in the messages.
$\sum_L W_L \bullet T_{cLi}$	Suspicious information amount with conspirators	The amount of suspicious information contained in the messages that are only sent to the known conspirators.
D_{si}	Suspicious degree	The degree to which the staff is considered to be a conspirator only according to the content of the text messages
A_i	Activity level	Depict how frequent the staff is in terms of texting
ρ_{oi}	Outgoing ratio	The ratio of the outgoing message in a staff's total message
ρ_{ci}	Ratio of Messages with conspirators	The ratio of the message that are only sent to or received from the known conspirators in a staff's total message
P_i	Potential to be conspirator	The likelihood a staff should be considered to be a part of the conspiracy
P_{DH}	Higher discriminate value	The P_i value of the higher discriminate line
P_{DL}	Lower discriminate value	The P_i value of the lower discriminate line

1.1.2 Assumptions

✧ In reality the factors we considered will certainly affect each other. But in order to simplify the model we assume that the final likelihood can be approximated as a linear function of the factors that have been discussed.

✧ There are two 'Elsie' in the name list, and we assume that they are the same people whose data has not been recorded completely at one time. So we add the data of the two Elsie (at node 7 and node 37) together as the data of Elsie's. We also apply this analogy to other staffs sharing the same name, such as Gretchen and Jerome.

2 Model building and algorithm explanation

2.1 Problem analysis

The likelihood of a person getting involved in a commercial crime is quite a complex issue and can be hard to be described by an objective and descriptive method. Obviously, the job calls for information and knowledge in different fields as well as sources. Therefore, we decided to use an AHP (Analytic Hierarchy Process) to derive a relatively warranted and comprehensive evaluation on each person involved in the case.

After considering several relevant factors that can have different influences from minor to dramatic, we concluded that to work out the likelihood of each node to being a conspirator, two dimensions should be considered:

- **Network dimension:** describe the topological characteristics and the role of each node in the whole network;
- **Information Flow dimension:** describe the quantized flow of suspicious information of a node.

The following diagram shows the 4 factors in the 2 dimensions that contribute to the overall potential to be conspirators of each node:

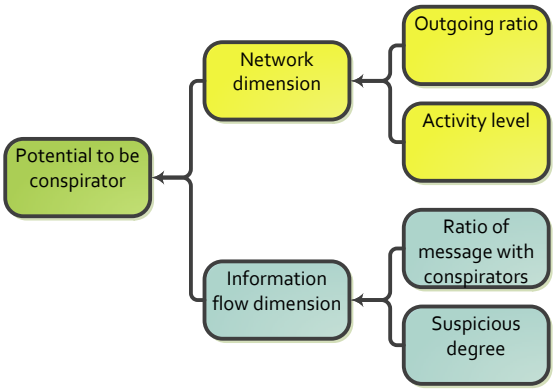

Figure 1 The 4 main factors influencing the overall likelihood.

The values of all the factors are dimensionless values, which means that they can be added together to form the final result.

2.2 Overall algorithm of our program

$$P_i = \rho_{oi} + \rho_{ci} + D_{si} + A_i \tag{1}$$

The algorithm of the E-judge model is been illustrated as follow:

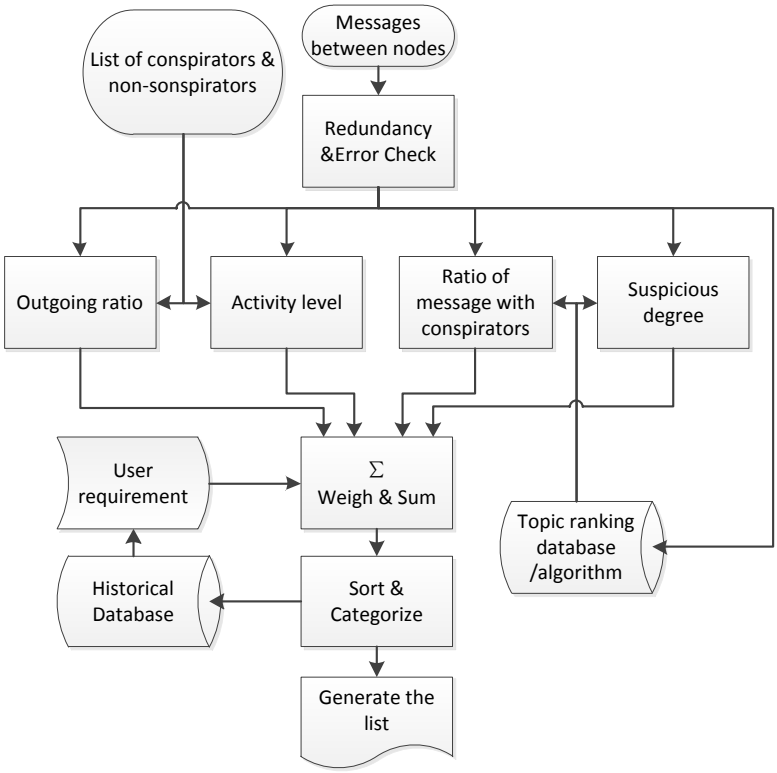

Figure 2

2.3 Defining the factors

2.3.1 Information flow dimension (microscopic view)

This type of factors mainly takes the information content of the messages into consideration.

2.3.1.1 Ratio of Messages with conspirators

$$\rho_{ci} = \frac{T_{ci}}{T_i} \quad , \quad \rho_{ci} \in [0,1] \tag{2}$$

Denotation:

$$\text{Ratio of messages with conspirators} = \frac{\text{Total message amount with conspirators}}{\text{Total message amount}}$$

The reason why we choose this factor and explanation:

- In this situation, the more frequent a staff contact with the known conspirators the more likely he is part of the conspiracy.
- This factor indicates how frequent a staff contact with conspirators

- To obtain the value of T_i and T_{ci}, we came up with an algorithm. In this algorithm we firstly defined several matrix:

Message sparse matrix:

$$N_p = \underbrace{\begin{Bmatrix} 0 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ \vdots & \cdots & \cdots & \vdots \\ 0 & 0 & \cdots & 0 \end{Bmatrix}}_{\text{i senders}} \text{j receivers} \tag{3}$$

$$\left(0 \le i, j \le A-1 \quad i \ne j\right)$$

A is total the number of nodes.

N_p refers to one message indexed by 'p', whose column represents the message source and row index stands for the message destination for each message. All the values in N_p are zeros except for only one element carrying a 1 representing one message. For example, $N_p(i, j)$ refers to the message sent to j by i.

Total conspiracy interconnection matrix:

$$M = \begin{pmatrix} m(0,0) & \cdots & m(0,j) & \cdots & m(0,A-1) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m(i,0) & \cdots & m(i,j) & \cdots & m(i,A-1) \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m(A-1,0) & \cdots & m(A-1,j) & \cdots & m(A-1,A-1) \end{pmatrix} \tag{4}$$

$$\left(0 \le i, j \le A-1 \quad i \ne j\right)$$

A is the number of nodes.

This matrix describes all the messages in the network. For example, $m(i, j)$ represents the amount of message sent to j by i.

The conspiracy interconnection matrix M can be easily generated by computer program according to the following equation:

$$M = \sum_{p=0}^{N} N_p \tag{5}$$

Where N is the total message.

Conspirators' vector:

$$C = \begin{pmatrix} C_0 & C_1 & \ldots & C_{A-1} \end{pmatrix} \tag{6}$$

This vector represents the known conspirators, for example, if the i-th person is a known conspirator, C_i is 1, otherwise is 0.

Now we can find our T_i and T_{ci} :

$$T_i = \sum_{j=0}^{m-1} M_{i,j} \tag{7}$$

This represents the sum of a column (or a sender), where $M_{i,j}$ is the element at row i and column j in matrix M.

$$T_c = C \times M \tag{8}$$

Where our T_{ci} is the elements of the row matrix T_c.

2.3.1.2 Suspicious degree

$$D_{si} = \frac{\sum_L W_L \bullet T_{cLi}}{\sum_L W_L \bullet T_{Li}} \quad , \; D_{si} \in [0,1] \tag{9}$$

Denotation:

$$\text{Suspicious degree} = \frac{\text{Suspicious information amount with conspirators}}{\text{Suspicious information amount}}$$

The reason why we choose this factor and explanation:

- The more suspicious information is contained in a staff's message, the more likely he will be a part of the conspiracy.
- If we only calculate the suspicious information amount contained in the messages sent to or received from all others, then those who are innocent but communicate with all other people very frequently may be wrongly accused if they happened to include some of the suspicious information in their messages by chance. In this situation, the more frequent a staff contact with the known conspirators the more likely he is part of the conspiracy.
- To solve the problem mentioned above, we define the **ratio** of suspicious information amount in the messages sent to or received from the known conspirators in all messages of a staff as the suspicious degree.
- Since different messages have different amount of suspicious information, we introduced W_L as the weight of a specific level of messages that should be added in calculation.
- We use a simple text analysis program to divide all the messages into several levels (indexed by L) according to their likelihood of containing conspiracy information, and this program is going to be a very powerful and useful tool when

processing large amounts of text messages. Our team has developed such a tool, and this will be explained in detail in 4.4.

2.3.2 Network dimension (macroscopic view)

This type of factors mainly assesses the role of each node in the network.

2.3.2.1 Outgoing ratio

$$\rho_{oi} = \frac{O_i}{T_i} \quad , \quad \rho_{oi} \in [0,1] \tag{10}$$

Denotation:

$$\text{Outgoing ratio} = \frac{\text{Outgoing}}{\text{Total}}$$

The reason why we choose this factor and explanation:

- The more information a staff has to tell others, the more likely he is one part of the conspiracy. For example, if a staff is actual a part of the conspiracy, he will be responsible to either generate information or pass information to others. In contrast, an innocent man doesn't need to send messages frequently.

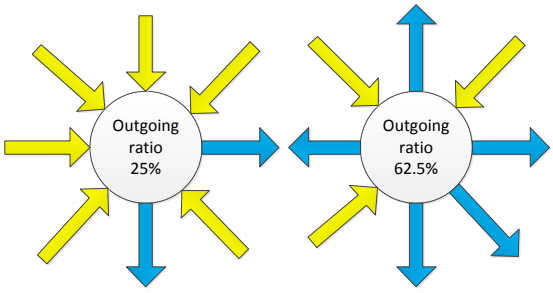

Figure 3 Comparison between nodes with the same number of message yet different outgoing ratio

The figure above shows how two nodes with the same amount of message flow may differ from each other when we take the direction of message into account. In a communication network the out-degree of a node can reveal the role it plays in the whole network, likewise the outgoing ratio we use here also represents whether a node acts more like an information source or destination. In criminal network analysis, those who always generates information more often than listen to the others should be paid more attention to for that this kind of node show actually more positive and significant influence onto the network.

- Although this factor cannot determine the likelihood that a staff is a conspirator alone, it can work well along with other factors.

2.3.2.2 Activity level

$$A_i = \frac{T_i}{\max\{T_0, T_1, \cdots, T_n\}} \quad , \quad A_i \in [0,1] \tag{11}$$

Denotation:

$$\text{Activity level} = \frac{\text{Total message amount}}{\text{maximum total message amount}}$$

The reason why we choose this factor and explanation:

- Activity level can measure the vitality of nodes in the network. An active node has much more influence to others, thus should be regarded as a key point of the whole network.
- Normalization is achieved by dividing a staff's total message by the maximum total message amount. It can eliminate the difference of diverse network.

3 Testing and optimizing the E-judge model using the EZ case

By applying our E-judge model mentioned above, we obtain the following result.

Table 2

Likelihood	Name	T_i	O_i	ρ_{oi}	T_{ci}	ρ_{ci}	D_{si}	A_i	P_i	**Tag**
Low	Jane	3	2	0.667	0	0	0	0.429	1.095	I
	Anne	5	2	0.4	0	0	0	0.714	1.114	I
	Carol	6	2	0.333	1	0.167	0.25	0.857	1.607	I
	Fred	3	1	0.333	1	0.333	0.667	0.429	1.762	I
High	Bob	4	3	0.75	1	0.25	0.5	0.571	2.071	C
	Harry	6	3	0.5	3	0.5	0.333	0.857	2.190	I
Very high	Inez	3	2	0.667	1	0.333	0.9	0.429	2.329	C
	Ellen	5	3	0.6	3	0.6	0.846	0.714	2.760	C
	Dave	7	3	0.429	7	1	1	1	3.429	C
	George	7	3	0.429	7	1	1	1	3.429	C

C: conspirator I: innocents

Violet: known conspirators Blue: conspirators worked out by supervisor

As we can see, the P_i values of the known conspirators are high, while those of the known innocent people are low.

3.1 Determine the discriminate lines:

Due to irregularities of the data, the certain discriminate line that can distinguish the innocents from the criminals is impossible to obtain precisely. In order to get a

relatively accurate result, we roughly introduced two discriminate lines to distinctly categorize the people in three groups according to their likelihood to be a conspirator:

The upper discriminate line has a higher P_i value (denoted as P_{DH}), above which all the staffs should be considered as criminals;

The lower discriminate line has a lower P_i value (denoted as P_{DL}), below which all the staffs should be considered as innocent;

As for those whose P_i values meet $P_{DL} < P_i < P_{DH}$, their true identities cannot be determined rigidly. We call this area as uncertain zone. They have to be analyzed separately according to other attributes, and the procurator may need to gather some more data if possible.

In this case, as is shown in the table above, Bob (known conspirator) has a lower P_i value than Harry (known innocent), which means that these two people are in the uncertain zone and that the boundary value of is probably around here.

The P_i value of the lower discriminate line should be in this range:

$$1.762 < P_{DL} < 2.071 \tag{12}$$

The P_i value of the higher discriminate line should be in this range:

$$2.190 < P_{DH} < 2.429 \tag{13}$$

Then we drew the two lines in the table above, which are marked with red and orange.

However, since the data of this case is limited and the gaps of the P_i value between the adjacent members in the ranking list are too big, the positions of those two lines should be analyzed precisely in different circumstances with the help of more sufficient data if possible.

3.2 How we overcome the shortage of the supervisor's model:

In the supervisor's model, Bob and Inez were missed and Carol was falsely accused.

In our E-judge model, we designed specific algorithm to prevent the members from being falsely judged.

For people like Bob:

This type of people send messages more than receive, which means that they have a lot of information to be transmitted to others, and the ratio of outgoing messages is much higher than that of the incoming. Thus we defined ρ_{oi} , outgoing ratio, which has been explained in detail in 2.3.2.1

For people like Inez:

This type of people doesn't contact with too many people, but most of their contacts are known conspirators. As we all know, the more a staff contacts with the known

conspirators and the more suspicious information is contained in the messages, the more likely they are a part of the conspiracy.

Thus, we introduced two factors to assess this feature. One is suspicious degree D_{si}; one is ratio of messages with conspirators ρ_{ci}. We use these two factors to assess the information content of the messages, which are explained in detail in 2.3.1.1 and 2.3.1.2.

For people like Carol:

Using the factors we defined it's easy to identify the true identity of this type of people.

4 Solve the current case according to the four requirements

4.1 Information Redundancy Pre-check

4.1.1 Reason for Redundancy Check

As the social network can be extremely complicated in the real interpersonal relationship, there are lots of redundant messages that need extra work to be identified and screened off. Such useless information always doubles the work in analyzing network relations and sometimes even hides the fact underlying the raw data.

In the commercial crime case we are dealing with, the 80 conspirators and 400 messages which form rather a complex relations-network. To identify the extra information that increases the job burden, computing complexity and the storage space to hold all the necessary data, we introduce a Hop-count method from a concept of Internet protocol determining the correlation between individuals in the network and the conspirators.

Although introducing such a method may not have an instant effect on cases with small amount of information (e.g. the EZ case), its benefit of saving resource and simplifying computing will probably arise if the data amount increases dramatically, since the complexity of computing in this case is approximately in proportion to the square of node number in the network, i.e. $C \propto N^2$, where N is the node number.

4.1.2 Concept of Hop-count

Hop-count is a value which stands for the logical distance between two nodes in a network. It can help people understand how close two nodes are or how close relation the two modes share.

In the scenario of determine the likelihood of crime, Hop-count can tell the degree of the co-relation between a specific person and a conspirator.

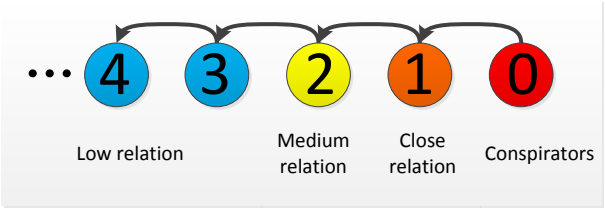

Figure 4

Figure: Hop-count illustrates the relation between ordinary nodes & conspirators.

The individuals that have direct links with the conspirators are assigned with highest level of co-relation with the conspirators (to verify the difference between them, see Suspicious Degree) hence are more likely to be involved in the conspiracy. Those who are indirectly linked with the conspirators are of relatively lower possibility according to how many hops their connections actually need. Therefore, the larger Hop-count to a closest conspirator the less likely a person would be involved in the conspiracy.

So we group all the individuals in the network and group them by their Hop-count to conspirators. Those who need 3 or even larger Hop-count is considered to have weak link with the crime and are screened off before they enter the list of suspicions.

4.1.3 The Accuracy of Redundancy Check

As a matter of fact, in order to prove the accuracy of this pre-check, we didn't actually delete them in the list of suspicious (but they are labeled with flags), and finally find out that these individuals who have a relatively lower correlation with conspirators tend to have lower possibility in the final result compared with the others .

The Following table shows all the nodes with Hop-count greater than (or equals to) 3 and their value in the final ranking.

Table 3 the result of redundancy check

Number	Name	Raking(out of 83)
52	Vind	50
53	Chara	81
55	Olina	68
56	Cha	78
57	Sheng	83
58	Lao	55
61	Le	56
63	Quan	73
72	Andra	77
74	Gard	51
75	Bariol	76
76	Cole	75
80	Fanti	80

Varying from the 50th to 83rd, several nodes that have little influence in the network are successfully screened off by the redundancy check algorithm.

4.2 Requirement 1—Appling E-judge to the current case

We fit our E-judge model in the data of the current case. And finally we got a list of possible conspirators ranked according to P_i.

Table 4 Top 25 of the priority list

	Node	Name	ρ_{oi}	ρ_{ci}	D_{si}	A_i	P_i	Rank	Tag
very high	73	Carina	1	1	1	0.0434	3.0434	1	
	81	Seeni	0.8	0.6	1.3103	0.2173	2.9277	2	
	21	Alex	0.4	0.55	0.7977	0.8695	2.6172	3	C
	43	Paul	0.6316	0.3684	0.7836	0.8260	2.6097	4	C
	67	Yao	0.6	0.6	0.7272	0.6521	2.5794	5	C
	7	Elsie	0.375	0.125	0.4535	1.3913	2.3448	6	C
	18	Jean	0.5556	0.3333	0.6729	0.7826	2.3444	7	C
	60	Lars	1	0.3333	0.8641	0.1304	2.3279	8	
	54	Ulf	0.3	0.8	0.7537	0.4347	2.2885	9	C
high	32	Gretchen	0.4054	0	0	1.6086	2.0141	10	SM
	49	Harvey	0.4545	0.4545	0.5947	0.4782	1.9820	11	C
	34	Jerome	0.5	0.1071	0.0275	1.2173	1.8520	12	SM
	36	Priscilla	0.4444	0.2222	0.7905	0.3913	1.8484	13	
	33	Kim	0.75	0.5	0.3922	0.1739	1.8161	14	
low	2	Paige	0.4545	0.1363	0.2444	0.9565	1.7918	15	I
	40	Douglas	0.6923	0.3076	0.1718	0.5652	1.7370	16	
	0	Chris	0.6667	0.25	0.2654	0.5217	1.7038	17	I
	24	Franklin	0.5238	0.0952	0.0883	0.9130	1.6204	18	
	79	Phille	1	0.5	0	0.0869	1.5869	19	
	10	Dolores	0.5	0.1428	0.3266	0.6086	1.5782	20	SM
	3	Sherri	0.4762	0.0952	0.0649	0.9130	1.5494	21	
	30	Stephanie	0.4615	0.1538	0.3529	0.5652	1.5335	22	
	48	Darlene	0.4	0.25	0.0057	0.8695	1.5252	23	I
	20	Crystal	0.5333	0.2667	0.0718	0.6521	1.5240	24	
	50	William	0.5454	0.1818	0.2610	0.4782	1.4665	25	

C: conspirator SM: senior manager I: innocent
Violet: known conspirators Blue: known innocents Yellow: senior managers

As we can see in the data above, we prioritize the 83 nodes by likelihood of being a part of the conspiracy and select the top 25 nodes for discussing.

In this case, according to the results we acquired above:

$$1.762 < P_{DL} < 2.071, \quad 2.190 < P_{DH} < 2.429$$

We drew the lower discriminate line (red line) only above the known conspirator with the highest P_i value, Paige with node number 2, in order to include as many suspects into investigation as possible (otherwise some conspirators may get off). Thus $P_{DH} \approx 2.2$

We drew the higher discriminate line (orange line) between Ulf and Gretchen, for that there is the largest gap of P_i value between them. Thus $P_{DL} \approx 1.8$

Those with $P_i > 2.2$ are considered to have very high likelihood to be a part of the conspiracy. In the result it's clear that most of the known conspirators (marked with violet) are included in the highly suspicious list. Besides we found another three conspirators, they are:

Table 5

Node	Name	Priority rank
73	Carina	1
81	Seeni	2
60	Lars	8

Those with $P_i < 1.8$ are thought to be innocent people. According to our marks, it's clear that most of the known innocents (marked with blue) are excluded from the highly suspicious list

Those with $1.8 \le P_i \le 2.2$ are difficult to categorize rigidly. If we are only taking this data as reference rather than determination, the following people should be considered to have a high possibility to be a part of the conspiracy:

Table 6

Node	Name	Priority rank
32	Gretchen	10
49	Harvey	11
34	Jerome	12
36	Priscilla	13
33	Kim	14

Among the people in the table above, there are two senior managers (marked with yellow). We suggest that these two suspects been investigated further.

Although our E-judge model says that one of the known innocents (marked with violet), Harvey with node number 49, is not of the highest possibility to be a criminal, it is mainly caused by the uncertainty of the original data. Since either the data they are built upon or tested by are not assured to be perfect by anyone, the model is never ideal. Broadly speaking our E-judge model is valid enough to offer a relatively accurate result.

As for the leader of the conspiracy, we suggest that the suspects who rank the highest are highly likely to be the leaders we are looking for.

4.3 Requirement 2—Reactions to the input changes

4.3.1 Analysis on the changes

Given that Topic 1 is also connected to the conspiracy and that Chris is one of the conspirators, we change the database of our text analysis program and use our E-judge model and algorithm again to generate a new priority list. Comparing the differences of the two lists can also help us to analysis the sensitivity of our E-judge model.

Table 7 The comparison of two top 25 priority lists

Before change			After change		
likelihood	Node	Name	likelihood	Node	Name
very high	73	Carina	very high	73	Carina
	81	Seeni		81	Seeni
	21	Alex		7	Elsie
	43	Paul		21	Alex
	67	Yao		67	Yao
	7	Elsie		43	Paul
	18	Jean		54	Ulf
	60	Lars		18	Jean
	54	Ulf		60	Lars
high	32	Gretchen	high	32	Gretchen
	49	Harvey		0	Chris
	34	Jerome		49	Harvey
	36	Priscilla		2	Paige
	33	Kim		34	Jerome
low	2	Paige		36	Priscilla
	40	Douglas		33	Kim
	0	Chris	low	40	Douglas
	24	Franklin		68	Ellin
	79	Phille		20	Crystal
	10	Dolores		24	Franklin
	3	Sherri		79	Phille
	30	Stephanie		3	Sherri
	48	Darlene		10	Dolores
	20	Crystal		30	Stephanie
	50	William		48	Darlene

As expected, the most significant change is that Chris's ranking has improved. The previous version before change says Chris is an innocent. The list also shows that he has low likelihood to be a conspirator. After changing the database according to the new conditions, Chris's likelihood becomes high. It can prove that our E-judge model is proper and it perceives change.

Another change is that the amount of suspects with high likelihood has slightly increased (marked with violet). It is reasonable, since if there are more conspirators in the network, people who are close to the new conspirator will become more suspicious. Paige is a good example in this situation (marked with pink): he has several direct communications with Chris, so his likelihood has increased.

We also acquire that most part of the list doesn't change too much. For the suspects at the very top of the list, their P_i value is already far higher than others, thus their rankings are unlikely to change too much. In a large network, an individual's influence is limited, as well as a single topic. In another word, the majority of the staff neither has close contact with Chris nor talks topic 1 too much. Thus the ranking does not change too much in a macroscopic view.

4.3.2 Stability and sensibility analysis:

According to the analysis above, we can also acquire that our E-judge model's sensitivity is adequate. It can correctly detect new conspirator; it can pick out the suspects with close contact to the new conspirator as well as those who talk the new

suspicious topic too much. What's more, our E-judge model is stable as well. A few changes do not influence the whole ranking, where only a few suspects will have qualitative change.

4.4 Error analysis

Since we are unable to draw a certain discriminate line to distinguish the innocents from the conspirators, there will always be a group of people in the uncertain zone. Insufficient investigation and irregularity of the data will cause error in this specific zone, which is also a mainly flaw of our model.

4.5 Requirement 3—Semantic Text Analysis Model

Instead of having a full understanding of the detailed meaning of all the messages, deriving the degree to which the messages are related to the conspiracy is our ultimate goal.[1]

Our Semantic Text Analysis Model:

One of a powerful tool in this application is text categorization model whose result can be easily applied into further computing. Such algorithm can be implemented by programs easily, and our team has also developed one program, whose algorithm has been shown as follow:

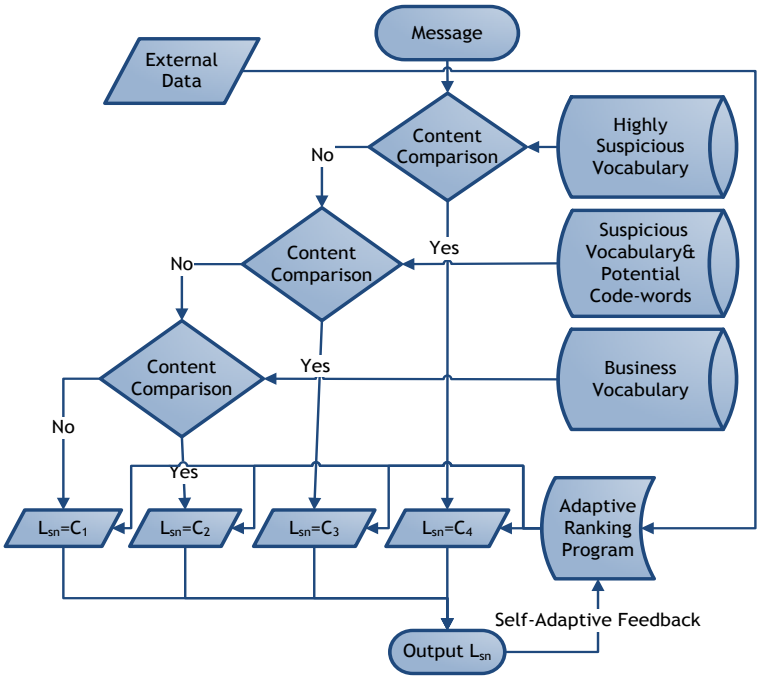

Figure 5 A possible program of semantic network analysis based on word categorization

We also build a data base for this model, contained all the possible vocabularies related to this model, for example, 'security hole' should be added to the suspicious vocabulary database.

How our Semantic Text Analysis Model can be improved:

Although such criteria can be adjusted according to the language characters and the specific circumstances to increase flexibility and adaptability, our program doesn't take consideration of the interrelationship between contexts. Such functionality limitation restrains the computers' comprehension performance.

Another kind of analytic tool using semantic neural network can provide more informative result in evaluating messages. For instance, Catpac^{TM}, a self-organizing, interactive artificial neural network used for text analysis *[2]* can find key words in the body of text and figure out the co-concurrence between each key words.*[3]*

By analyzing such co-concurrence between different words, the actual meaning of words in a specific language environment can then be determined by a computer program.

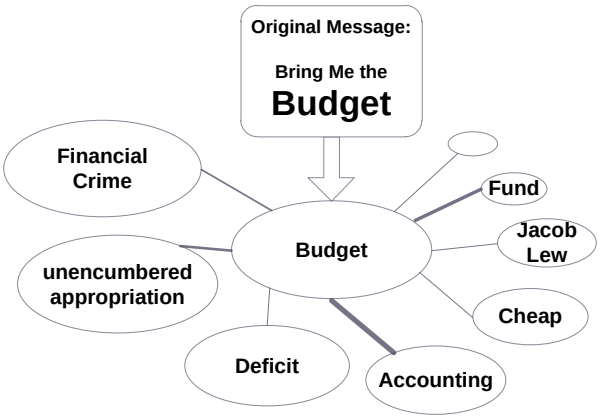

Figure 6 A set of possible semantic interconnections of keywords

(The size of a circle shows how the word is correlated with the case; the thickness of a line represents the magnitude of co-concurrence)

In order to support different ways of evaluating the content of a message the E-judge model contains a discrete module that can be replaced by other algorithm. The output interface of this module will always provide a normalized value so as to fit into the algorithms that take charge of computing the factors in the information flow dimension (ratio of message with conspirators& suspicious degree).

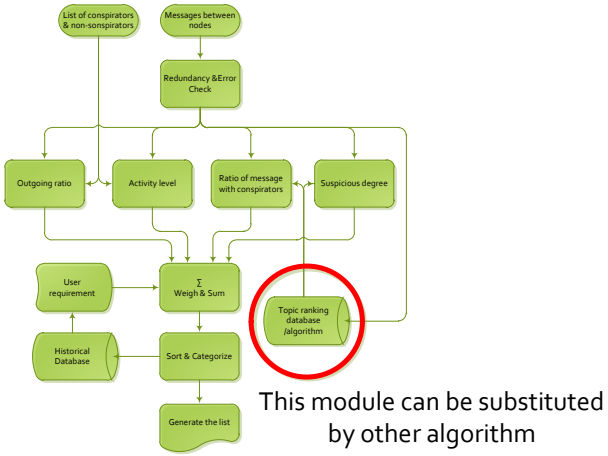

Figure 7 the discrete module taking charge of the lingual analysis in our E-judge model

To sum up, by applying Semantic and text Analysis into evaluating the messages, more valuable and objective results can be acquired comparing with the current evaluation results. Therefore, our E-judge model is designed to have a compatible and open interface for other modules so that it can be improved in further development.

4.6 Requirement 4—Scalability, expandability, analogy

We lay special stress on the scalability and the expansibility of the methodology.

4.5.1 Scalability

Futile information only makes the calculation complex and does not contribute to the final result, thus it is extremely important to screen out the information that is really useful for us to analyze the case.

Redundancy check

To solve this limitation, we designed a redundancy check step which is introduced in (4.1). Redundancy check can reduce the analyzing list and make the calculation more efficient and accurate. The benefit of redundancy check will become extraordinary when the data increase dramatically.

Using of database

Our E-judge model can implement screening and assessing using computers based on existing databases. It will reduce the requirement of the users. The usage of authoritative databases will ensure the validly and scalability of the model.

4.5.2 Expandability

In this technology developed era, all fields have closer relationships. An outstanding methodology should have expandability to be able to apply in many fields. Since different field do have some distinctiveness, the model should be flexible as well.

Universality

As we analyze this case from the point of view of the network, the model is suitable for all networks with similar structure. Necessary elimination is suitable for any case. For analysis of the rest of the candidate we mainly focus on two dimensions. One is network dimension representing the characters and the role of each node in the whole network, the other is information flow dimension quantizing the connection with nuclear nodes. This view can be used in to different types of cases.

Flexibility

What’s more, through self-adaptive feedback, our E-judge model can adjust itself automatically. It guarantees that the model can adapt to specific case properly. This flexibility will surely enhance the expandability.

4.5.3 Analogy

Biological network

In a biological network, processing method is exactly the same. By analyzing the image or chemical data based on professional knowledge, we can get a database of information that can influence the probability of infection which is similar to our message traffic. Redundancy check will remove the cells that have little possibility to be infected to simplify the case. Our E-judge model will consider both each cell's characters in the whole network and each node's connection with the identified infected nodes. The prioritize list can be a guideline that help us to identity infected cells.

With AIDS, for example, after infected, T-helper cell's function will decrease that will reduce the generation of r-interferon; lymphocyte decrease rapidly, concentration less than one of ten of the original concentration; the demand for amino acid changed; the generation of antibody suffocate*[4]*. All of these factors act as characters of the network. Thus our E-judge model can handle it successfully.

Portfolio Investment

E-judge can also be applied to help portfolio investors in finding a more profitable choice among candidate investees by analyzing the performance of different companies. Since the modern commercial affairs are becoming increasingly integrated and interrelated, the prediction of investment earnings will no longer depend solely on the annual income of a company. Therefore, the role a company plays in its business or stakeholder network will attribute more in the company's future, which makes E-judge helpful in estimating.

Like conspirators in the criminal network, there're companies who have been very profitable to be invested according to historical data, and there are those who are not worthy of investment like non-conspirators in our previous case. Likewise, the interaction between companies (Trading contacts, franchise, licensing, business alliance, etc.)which can make reasonable analogies with messages between individuals can also be categorized and scored according to their influence on the companies' performance.

By substituting the interaction information of all the candidate companies, E-judge can screen off those companies with less possibility to make profit and generate a ranking list which shows the priority of investment.

5 Strengths and weaknesses

Strength

We have already discussed the following strengths of our E-judge model in details before.

Stability and sensibility (4.3.2)

Scalability and expandability (4.5)

Weakness

Our E-judge model does not give a single explicitly boundary to distinguish conspirators and non-conspirators. With the principle of avoiding omitting, it gives a

fuzzy zone determined by two boundaries, candidates in it have high possibility to be a conspirator and they should be further surveyed.

6 References

[1] Christie M. Fuller, David P. Biros, Dursun Delen, An investigation of data and text mining methods for real world deception detection, Expert Systems with Applications, Volume 38, Issue 7, July 2011, Pages 8392-8398, ISSN 0957-4174, 10.1016/j.eswa.2011.01.032.

[2] Wikipedia Catpac http://en.wikipedia.org/wiki/Catpac

[3] MARYA L. DOERFEL, GEORGE A. BARNETT, A Semantic Network Analysis of the International Communication Association, Human Communication Research, Vol. 25 No. 4, June 1999 589-603 International Communication Association.

[4] Wikipedia AIDS. http://en.wikipedia.org/wiki/AIDS