Hindawi Publishing Corporation
Abstract and Applied Analysis
Volume 2013, Article ID 308675, 10 pages
http://dx.doi.org/10.1155/2013/308675

Research Article

Crime Busting Model Based on Dynamic Ranking Algorithms

Yang Cao,[1] Xiaotian Xu,[2,3] and Zhijing Ye[1]

[1] *College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210023, China*
[2] *College of Overseas Education, Nanjing University of Posts and Telecommunications, Nanjing 210023, China*
[3] *School of Engineering and Computing Sciences, New York Institute of Technology, Old Westbury, NY 11568-8000, USA*

Correspondence should be addressed to Yang Cao; caoyeacy@njuptsast.org

Received 28 May 2013; Accepted 11 June 2013

Academic Editor: Xinsong Yang

This paper proposed a crime busting model with two dynamic ranking algorithms to detect the likelihood of a suspect and the possibility of a leader in a complex social network. Signally, in order to obtain the priority list of suspects, an advanced network mining approach with a dynamic cumulative nominating algorithm is adopted to rapidly reduce computational expensiveness than most other topology-based approaches. Our method can also greatly increase the accuracy of solution with the enhancement of semantic learning filtering at the same time. Moreover, another dynamic algorithm of node contraction is also presented to help identify the leader among conspirators. Test results are given to verify the theoretical results, which show the great performance for either small or large datasets.

1. Introduction

Nowadays, many crimes are committed by collaboration of conspirators. Therefore, with the interconnections of conspirators, a complicated conspirator network can be spawned. However, many conspirators still sustain their social ties with the outside, and thus conspirator network often hides in a greater social network.

To identify the hidden conspirators' network from a complex social network, it calls for us to discover the hidden conspirators' network and analyze its unique features to detect the leader. Those features of networks can be captured from various information, such as topological properties of the network, semantic network analysis of their messages interactions, and other prior knowledge, which contains known conspirators, known nonconspirators, and background of the entire social network.

To study this network, manual approach is the most comprehensive method. However, it would become extremely ineffective and inefficient with large database. From many pervious work [1–4], to deal with the problem of large database, people have used the graphic-based centrality measures of network to study the characteristics of conspirators. Criminals with high betweenness centrality are usually brokers, while those with high degree centrality appreciate better profit by running higher risks [1]. Morselli also proposed that leaders of a criminal organization tend to balance profit and risk by making a careful trade-off between out-degree and betweenness centrality [2].

However, those static centrality approaches, which only utilize graphical properties, tend to overlook many other imperative analytical information such as the network topology, the sematic meaning of people's interactions. Therefore, the idea of complex network analysis, including subnetwork detection and block-modeling, has been introduced to detect the inner patterns of interactions between social actors [3]. Despite they shed light on the internal structures of networks, these approaches are still burdened with intimidating complexity with large databases.

Inspired by the discussions above, the suspicious ranking system must be modified carefully to combine these pieces of information with network topology and centrality. Meanwhile, topics with higher frequency and more contacts to known conspirators may increase the crime probability. To provide a better solution to this problem, an advanced dynamic network mining approach with semantic network

analysis will be introduced in this paper. Notably, based on modifying the definition of centrality, dynamic cumulative nominating algorithm to measure each people's suspicion will be adopted to serve our model better, which will reduce computational expensiveness than most other topology-based approaches. In the meantime, with the enhancement of semantic learning, the accuracy of solution will be also increased. Furthermore, with deeper analysis of the structure of the possible conspirator network, node contraction algorithm will be presented to help identify the leader among conspirators. As a final point, the traditional centrality approach is also performed to verify the inner agreement and connection with our approach. Through that way, a priority list based on the possibility of the suspect and suspect's leader rank will be dynamically adjusted with new clues added. Eventually, the model will be promoted into a more common place which can be applied in other networks.

The algorithms will be tested with the data from ICM 2012 problem [5] which shows the great performance for either small or large datasets.

2. Overall Assumptions and Preliminaries

2.1. Overall Assumptions

(i) A conspirator knows all other members in the conspiracy.

(ii) A nonconspirator does not know who conspirators are and hence treats conspirators and nonconspirators equally.

(iii) A conspirator is reluctant to mention to an outsider topics related to crime.

(iv) Conspirators tend not to talk frequently with each other about irrelevant topics.

(v) The leader of the conspiracy is the most inseparable of the whole conspirators' network.

(vi) The information of known conspirators and nonconspirators is valid.

(vii) The information offered in materials is complete and reliable. All the messages and the topics represent their thoughts, ignoring that someone lies during the eavesdropping.

2.2. Models Preliminaries

2.2.1. Degree. Degree is defined as the number of edges linked to a node in graph. It can be written as $\deg(v)$. In directed graph, the number of incident edges is input degree $\deg_{\text{in}}(v)$. The number of emergent edges is output degree $\deg_{\text{out}}(v)$.

2.2.2. Centrality. Centrality of nodes indicates the relative importance of nodes within a graph. It can be utilized to determine the center of the suspicious network. Here are three popular types of centrality.

(i) Degree Centrality. Degree centrality refers to the centrality of a node with respect to other adjacent nodes. In suspicious network, it reflects activeness of a member. More links to a member means more possible the member be the leader. For a given graph $G := (V, E)$ with V set of nodes and E set of edges, the normalized degree centrality of node i is

$$C_D(i) = \frac{\sum_{j=1}^{N} K\left(e_{ij}\right)}{N-1}, \quad i \neq j, \tag{1}$$

where $K(e_{ij})$ = the number of conversations between v_i and v_j. e_{ij} = binary variable represents whether there is a link between two nodes. If there is one or more conversation from v_i to v_j, $e_{ij} = 1$, otherwise the value is 0; $N = \text{count}(v_j)$.

(ii) Betweenness Centrality. Betweenness centrality measures how much a node acts as a medium along the shortest path between two other nodes. It helps analyzing who has bigger possibility to be an intermediary to exchange information between two other members. Member with high betweenness centrality also plays an important role in suspicious networks. The normalized betweenness centrality is

$$C_B(i) = \frac{\sum_{j=1}^{N} \sum_{k<j}^{N} \omega_{j,k}(i)}{N-1}, \quad i \neq j, \tag{2}$$

where $\omega_{j,k}$ shows whether the shortest path between two nodes passing node i.

2.2.3. Symbol Chart. For some symbol and their meaning, see Table 1.

Table 1: Symbol.

Symbols	Meaning
$\deg(v)$	Degree of a node
$C_D(i)$	Degree centrality
$C_B(i)$	Betweenness centrality
$\omega_{j,k}$	The shortest path between two nodes passing node i
l	Number of topics in one conversation
$K(e_{ij})$	The number of conversations with same person
$\widetilde{q_i^{n+1}}$	The nomination score of the node v_i after $(n+1)$ iteration
q_i^n	The normalized nomination score of the node v_i after n iteration
a_{ij}	The element in the adjacent matrix of the effect network
w_t	Empirical weight of topic's effect, in our case $w_t = 15$
$T(v_{ij})$	Unified topic's suspicion degree, from v_j to v_i

3. Analysis of Suspicious Topics

3.1. Statistical Analysis. Topics among known conspirators are important information which can be utilized to analyse conspirator's characteristics on choosing topics. According to the statistical characteristics, some unknown conspirators can be unearthed and some people's suspicion can be eliminated.

Table 2: Topics among known conspirators.

	Jean	Alex	Elsie	Paul	Ulf	Yao	Harvey
Jean		**11***			8		14
Alex			1	**13***	**11***	3, **7***	
Elsie		**11***			**13***		
Paul	**11***		**7***		**7***		4
Ulf		**7***, **11***, **13***				**13***	
Yao	**13***	**7***, **11***, **13***	**7***, 9		**13***		2, **7***
Harvey						**13***	

Suspicious topics are in bold with *.

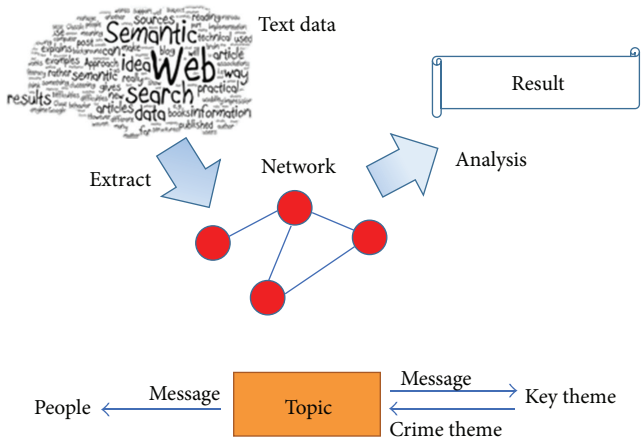

Figure 1: Semantic network analysis procedure.

After analyzing the test data from ICM 2012, Table 2 displays the topics relation matrix among known conspirators, and then Table 3 illustrates the comparison of suspicious topic among different kinds of groups. As we know, conspirators usually talk about known suspicious topics (topics 7, 11, 13) and rarely mention irrelevant topics to other conspirators. Therefore, a suspect person is more likely to be a conspirator if conspirators often talk about suspicious topics with him. Conversely, a person who often receives irrelevant information is probably innocent.

To sum up, the conclusion is as follows.

(i) Topic 13 has more conspiracy possibility than topics 11, 7.

(ii) Topic 11 has more suspects involved. It is an active topic among the suspects and conspirators.

(iii) Topic 7 has larger frequency among suspects.

3.2. Semantic Network Analysis

3.2.1. Background. Semantic network is a network which represents semantic relations between concepts. This is often used as a form of knowledge representation. It is a directed or undirected graph consisting of vertices, which represents concepts and edges. Semantic network analysis, a machine learning technique to analyse large amount of messages in the foundation of semantic network, is commonly used in search engines. It can predict user's identity and inclination according to the frequently used searching words of the user. So, semantic analysis can be used to separate given topics into more detailed parts [6].

The term text analysis describes a set of linguistic, statistical, and machine learning techniques that model and structure the information content of textual sources for business intelligence, exploratory data analysis, research, or investigation. It attaches importance only to the text itself [7, 8]. Semantic network analysis represents human's knowledge and text analysis puts an emphasis on textual data processing. The procedure of semantic network analysis is demonstrated in Figure 1.

3.2.2. Basic Assumptions

(i) The meaning of a word could be represented by the set of its verbal associations [9].

(ii) Social network analysis is based on an assumption of the importance of relationships among interacting units [10].

3.2.3. Step Procedures and Results

Step 1. Series of suspicious words and themes related to crime will be extracted from the known suspicious messages in Topic.xls.

After machine learning of topic semantic diffusion [11], four suspicious factors are chosen by the system as follows:

(1) *Economic information*

(2) *Spanish words*

(3) *Codes*

(4) *Known conspirators' names.*

Step 2. It is easy for us to analyze the connection between the original topics and suspicious topics. Then we sort out topics that are related to these pieces of criminal information exacted. So, more topics are related to crime. We called the topics that were not suspicious topics and did not contain the key theme or word as normal topic.

Step 3. Since the number of conspiratorial topics has increased, in order to distinguish the degree of suspicion, we assign different topics to different weights again, depending on how much they are related to crime. In the process of calculating the topic weight, 4 variables i_1, i_2, i_3, i_4 are promoted to represent these 4 factors. The value of the four variables is either 1 or 0. Then, the suspicion degree of topics based on four factors can be presented as

$$\deg_s = k_1 i_1 + k_2 i_2 + k_3 i_3 + k_4 i_4, \quad (3)$$

where k_1, k_2, k_3, k_4 are coefficients of each factor and $k_1 + k_2 + k_3 + k_4 = 1$.

In order to simplify analysis, we assign same importance to the four factors, which means $k_1 = k_2 = k_3 = k_4 = 0.25$.

Table 3: Suspicious topics frequency comparison.

	Innocent-innocent	Innocent-conspirators	Innocent-suspect	Suspect-conspirators	Conspirators-conspirators
Topic 7*	0%	0%	8.03%	12.28%	**25%**
Topic 11*	0%	0%	4.16%	8.77%	**21.43%**
Topic 13*	0%	9.09%	0%	5.26%	**28.57%**
Sum	0%	9.09%	12.5%	26.3%	**75%**

Table 4: The suspicion degree of topics.

Topic	Suspicious keywords	$\deg_s$
1	Stock price (1)	0.25
2	Spanish (2)	0.25
3		0
4	Paige (4)	0.25
5	Security (3), Chris (4)	0.5
6	Paige (4)	0.25
7*	Spanish (2), codes (3), (4)	0.75
8		0
9	Jean (4)	0.25
10		0
11*	Accounting (1), flaws (3), (4)	0.75
12	Spanish (2)	0.25
13*	Key in the conspiracy plan	1
14	High price (1)	0.25
15	Computer security (3), Paige (4)	0.5

Table 5: 5 types of topics.

	Type 1	Type 2	Type 3	Type 4	Type 5
Topics	13*	7*, 11*	5, 15	1, 2, 4, 6, 9, 12, 14	3, 8, 10
$\deg_s$	1	0.75	0.5	0.25	0

According to former equation, the suspicion degrees of all 15 topics are shown in Table 4.

Based on the suspicion degree of each topic, 15 topics can be divided into 5 types as shown in Table 5.

Similar to pervious statistical analysis, Tables 4 and 5 suggest that suspicious topics 7, 11, 13 are highly relevant to crime for their bigger weight. Topics 5, 15 show some clues about crime and other topics indicate little connections with crime.

Step 4. In order to simplify the situation of multitopic and multiconversation with the same person, a unified topic suspicion degree of topic is defined in terms of each topic's suspicion degree. For conversation with multitopics, the suspicion degree of each topic is sum up to as the numerator of the unified topic suspicion degree. For the situation of multiconversation with same person for more than one time, an average value will be calculated. Therefore, the unified topic suspicion degree can be represented as

$$T\left(v_{ij}\right) = \frac{\sum_{k=1}^{K(v_{ij})} \sum_{l} \deg_s}{K\left(e_{ij}\right)}; \quad v_{ij} \neq 0 \text{ or } K\left(v_{ij}\right) \geq 1, \tag{4}$$

where l = number of topics in one conversation and $K(e_{ij})$ = the number of conversations between v_i and v_j.

This suspicion degree is a significant indicator to determine a person's identity.

4. Cumulative Nominating Algorithm

4.1. Algorithm Descriptions. The likelihood of conspirators' nodes in the social network can be regarded as the reputation in the small network of conspirators. Thus, the priority list can be obtained from the algorithm of cumulative nominating. The nomination scores indicate the importance of a particular node in the small network of conspirator, which reflect the suspicious degree of a suspect. This algorithm can be specifically described as 4 simple principles [12].

(1) The new nomination of a node not only includes its pervious nomination scores but also contains the effect of the pervious nomination scores of nodes which it is connected to. To sum up, the nomination score includes 2 parts: pervious scores and effect of others [13, 14].

(2) The effect of others includes other's scores, topic, and the manage identity.

(3) The initial nomination scores are the conditions of the problem, which is conspirators = 1, suspect = 0.5, and innocent people = 0. Normalization is done after each term of nomination in order to adapt to different size networks.

(4) After enough time of nomination cycle, when the nomination scores of suspect are enough for discrimination but not higher than the score of known conspirator, we believe that the irritation can be stopped and the higher cumulative nomination scores, the higher suspicious degree.

The entire cumulative process can be express as

$$\widetilde{q_i^{n+1}} = q_i^n + \sum_j a_{ij} q_j^n, \tag{5}$$

where $\widetilde{q_i^{n+1}}$ = the nomination score of the node v_i after $(n+1)$ iteration, q_i^n = the normalized nomination score of the node v_i after n iteration, and a_{ij} = the element in the adjacent matrix of the effect network.

Considering the effect of topic and the manage identity, a_{ij} can be defined as

$$a_{ij} = \frac{w_t \cdot T\left(e_{ji}\right) - w_m \cdot e_{ji} \cdot M\left(v_i\right)}{\deg_{\text{out}}\left(v_j\right) + 1}, \tag{6}$$

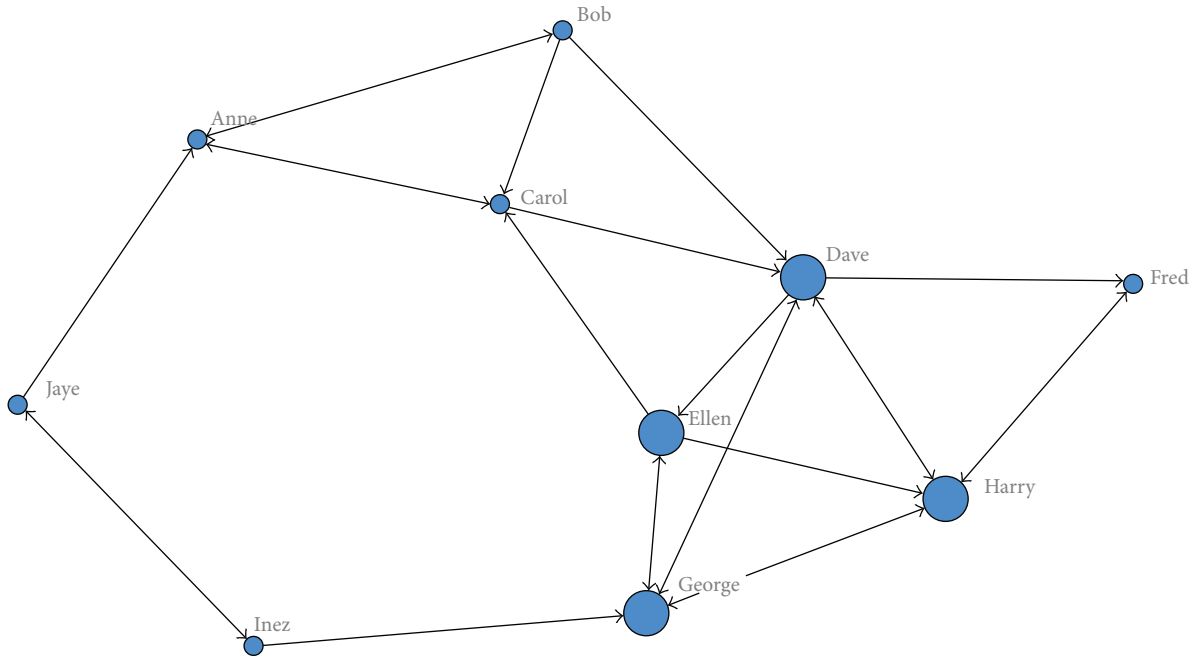

Figure 2: Topological clustering results of EZ case.

where w_t = empirical weight of topic's effect, in our case $w_t = 15$, $T(e_{ji})$ = unified topic's suspicion degree, from v_j to v_i, and w_m = empirical weight of manage identity's effect, in our case $w_m = 1$; $M(v_j)$ = binary keying variable, if v_j is a senior manager, $M(v_j) = 1$, otherwise the value is 0. (The senior manager is able to talk more sensitive topics, while others who talk those topics seem suspicious.)

In order to adapt the model to network with different size, normalization is required to be done after each term of iteration. Because of this, the normalized nomination function can be expressed as

$$q_i^{n+1} = \begin{cases} 1 & \widetilde{q_i^{n+1}} \geq 1 \\ N \cdot \widetilde{q_i^{n+1}} & 0 < \widetilde{q_i^{n+1}} < 1 \\ 0 & \widetilde{q_i^{n+1}} \leq 0 \end{cases} \tag{7}$$

whereby $\sum_i q_i^n = \sum_i q_i^0$, $q_i^n \in [0, 1]$;

$$\text{Normalization parameter } N = \frac{\sum_i q_i^0 - \text{count}\left(q_i^{n+1} = 1\right)}{\sum_{q_i^{n+1} \neq 1} \widetilde{q_i^{n+1}}}. \tag{8}$$

Known conspirators $q_i^0 = 1$, suspect $q_i^0 = 0.5$, and known innocent people $q_i^0 = 0$.

During the iteration, if $\text{count}(q_i^{n+1} = 1) > \text{count}(q_i^0 = 1)$, the iteration can stop and the final priority list is obtained based on the nomination score q_i^n.

4.2. Case Validation. Here, we use two given cases from ICM 2012 to verify our algorithm.

4.2.1. EZ Case. EZ case can be utilized to verify our model because of its similar suspicious network, small data quantity, and known result. Given data of EZ case will be substituted into the cumulative nominating algorithm in our model. The supervisor has offered some information:

(1) Dave, George are known conspirators and Anne, Jaye are known nonconspirators;

(2) 28 messages with 5 topics among 10 people;

(3) Ellen, Carol were found based on supervisor's analysis, but Carol was misjudged;

(4) Bob admitted his involvement in conspiracy;

(5) supervisor was pretty sure that Inez was involved.

After degree clustering, the topological results of EZ case are shown in Figure 2. In addition, based on semantic analysis, topics 1, 3, 5 are considered as suspicious topics and, respectively, weighted as 1, 1, and 1. Other topics are weighted as 0.

Test result is shown in Figure 3.

The test result shows that Dave, George, Inez, Jaye, and Ellen are the top five of all 10 people. This result is partly same to supervisor's result.

Moreover, Inez is identified by our model and Carol is also misjudged. The test result also corresponds with known result.

Bob shows low possibility in our result because he has few conversations with other people especially known conspirators and he often talks about nonsuspicious topics.

Therefore, our analysis model displays higher accuracy than supervisor's model.

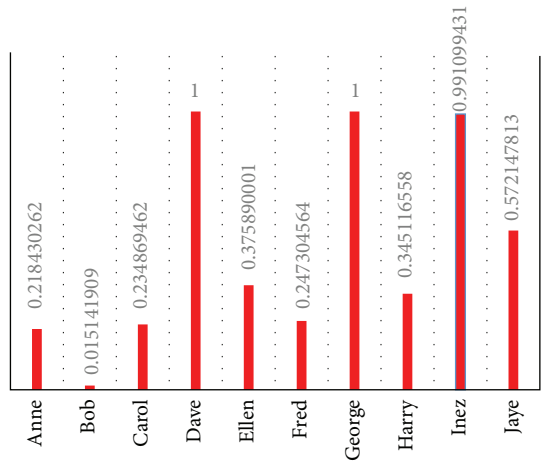

Figure 3: Priority list of EZ case.

4.2.2. Complex Case. The following are information that ICM 2012 has previously provided to us [5]:

(i) all 83 office workers' names;

(ii) 15 short descriptions of the topics;

(iii) 364 links of the nodes that transmit messages and the topic code numbers;

(iv) 7 known conspirators: Jean, Alex, Elsie, Paul, Ulf, Yao, and Harvey;

(v) 8 known nonconspirators: Darlene, Tran, Jia, Ellin, Gard, Chris, Paige, and Este;

(vi) senior managers of the company: Jerome, Delores, and Gretchen.

Two new clues are added as follows:

(i) topic 1 is also connected to the conspiracy;

(ii) Chris is one of the conspirators.

(1) Without New Clues. For complex case, ICM has offered us some information as shown below:

(i) 364 links of the nodes that transmit messages and the topic code numbers;

(ii) 7 known conspirators: Jean, Alex, Elsie, Paul, Ulf, Yao, and Harvey;

(iii) 8 known nonconspirators: Darlene, Tran, Jia, Ellin, Gard, Chris, Paige, and Este;

(iv) senior managers of the company: Jerome, Delores, and Gretchen.

After degree clustering, the topological results of complex case are shown in Figure 4. According to the given data in "Messages.xls", "Names.xls", and Topics.xls, a priority list based on cumulative nominating algorithm that shows the likelihood of one's being conspirator is obtained in Figure 5.

(2) With New Clues. New clues: topic 1 is also connected to the conspiracy and Chris is one of the conspirators.

After changing Chris's $q^0 = 1$, and changing Topic 1's $\deg_s = 1$, the result is obtained as Figure 6.

Even though the new clues cannot coincide with the original one, the former actually fluctuates slightly surrounding the latter. That is to say, they are quite similar to some extent. As to the big difference of the last node (Chris), it is a result when Chris becomes a known conspirator. Therefore, the model still works well and the list remains stable even if some conditions have changed.

After the iterative computation, the known conspirators' scores remain equal to one, while some known innocent persons' scores change rapidly, which is reasonable because in later clues some known innocent persons may also change into the conspirators, like Chris.

However, this algorithm cannot distinguish the possible leader from the known conspirators and high suspicious suspects because the nomination scores of known conspirators are all the same equal to one.

5. Node Contraction Algorithm

5.1. Algorithm Description. The model based on cumulative nomination algorithm cannot clearly differentiate nodes with high score. Hence, a new analyzing method should be introduced to compare the difference among nodes within conspiracy. The leader can be finally identified by this method.

Node contraction method [15–17] combines the node to be measured with its adjacent nodes into a new node and compares the importance of each contracted node based on its network agglomeration degree. This method suggests that the most important node is the one whose contraction leads to the largest increase of the networks agglomeration. Both degree and position are considered in node contraction.

Assume v_i is a node within graph $G = (V, E)$, the adjacent nodes of v_i with total number of k_i will combine with v_i, and then a new node v_i' substitutes the original $k_i + 1$ nodes.

As shown in Figure 7, if a node's agglomeration degree is $\Phi(G)$, the contracted agglomeration degree can be represented as $\Phi(G * v_i)$. If v_i is really important, the contracted network can be well agglomerated [18, 19].

The agglomeration degree of network depends on two factors, the connection ability among each node and the amount of nodes within the network N. The connection ability can be measured by an average distance L which refers to the mathematical average of distance between two nodes (node pair). The agglomeration degree of network is defined as

$$\Phi(G) = \frac{1}{N \times L} = \frac{N-1}{\sum_{i \neq j} d_{ij}}, \tag{9}$$

where $N \geq 2$, d_{ij} = distance between v_i and v_j, $0 < \Phi \leq 1$.

Then, the importance of node v_i is expressed as

$$\mathrm{IMC}(v_i) = \left[1 - \frac{\Phi(G)}{\Phi(G \cdot v_i)}\right] \cdot q_i, \tag{10}$$

where $G \cdot v_i$ = graph after contraction of v_i.

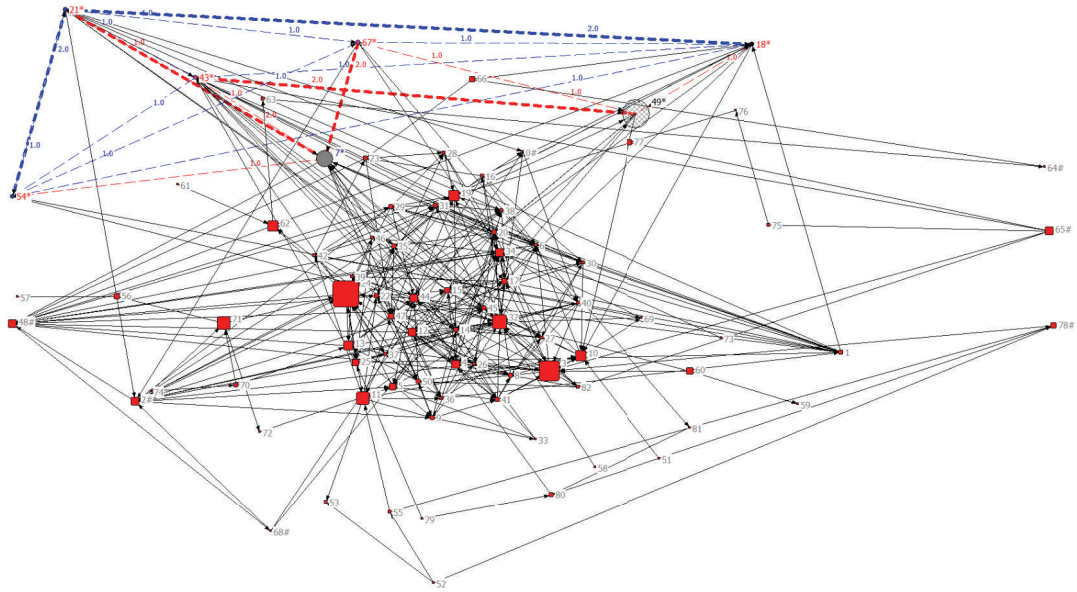

Figure 4: Topological clustering results of complex case.

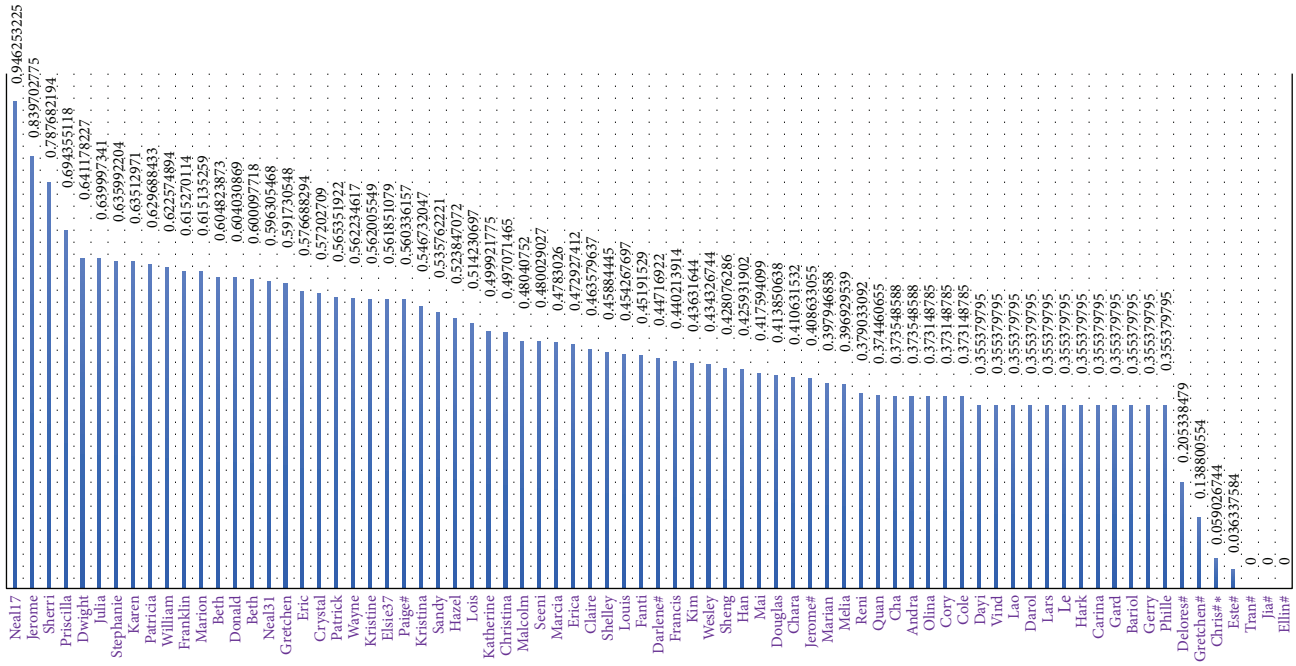

Figure 5: Suspect nomination priority list without new clues.

The node v_i plays more important role if it has more links k_i, because more other node pairs' shortest paths pass through it and the average distance after contraction will be highly reduced. In a conspiracy, the leader often connects with more people than other conspirators and messages between two subordinates often pass through the leader. This two properties can indicate a person's importance.

Basic steps of node contraction are shown below.

Step 1. Calculate the distance between a node pair (v_i, v_j), which is d_{ij}.

Step 2. Calculate the initial agglomeration degree $\Phi(G)$ of the network.

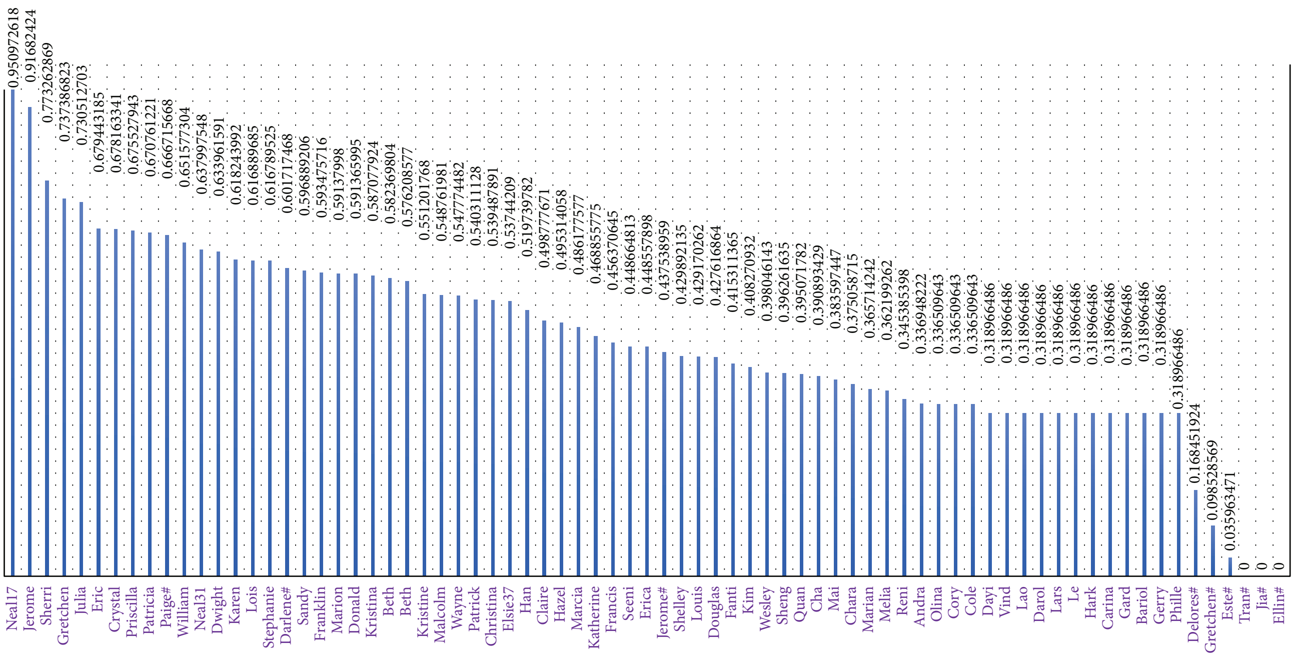

Figure 6: Suspect nomination priority list with new clues.

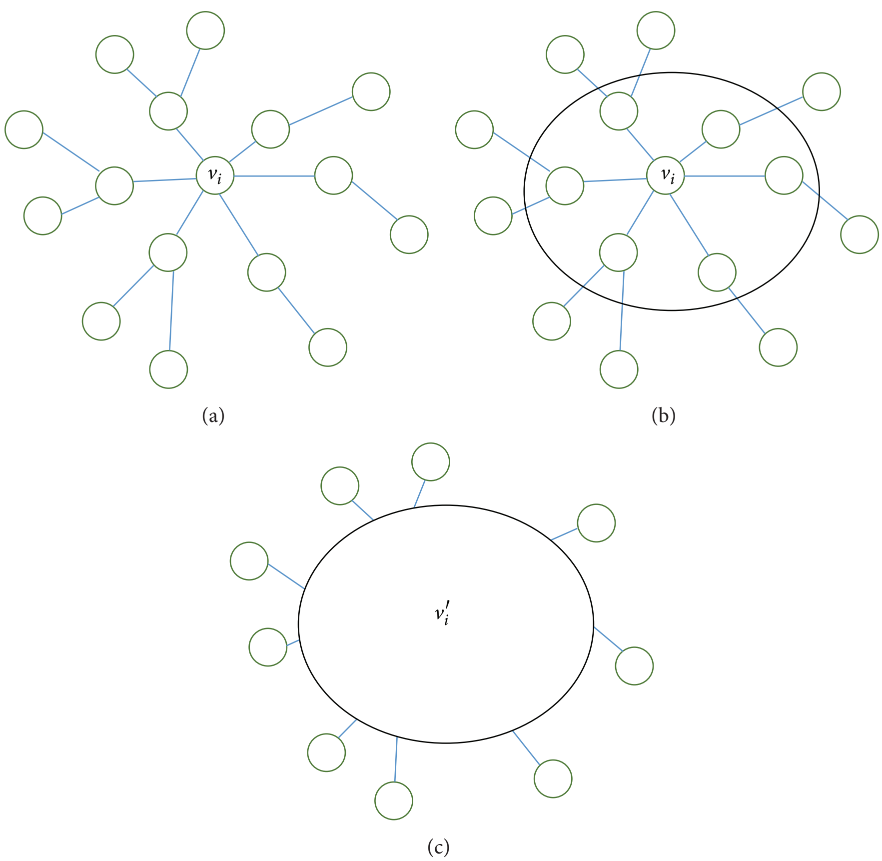

Figure 7: Illustration of node contraction.

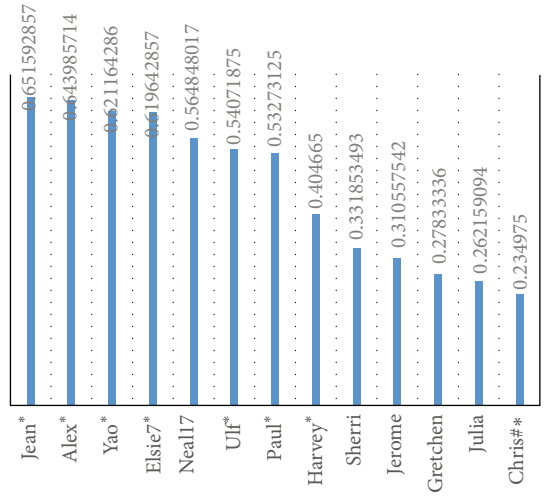

Figure 8: Leader Rankbased on IMC.

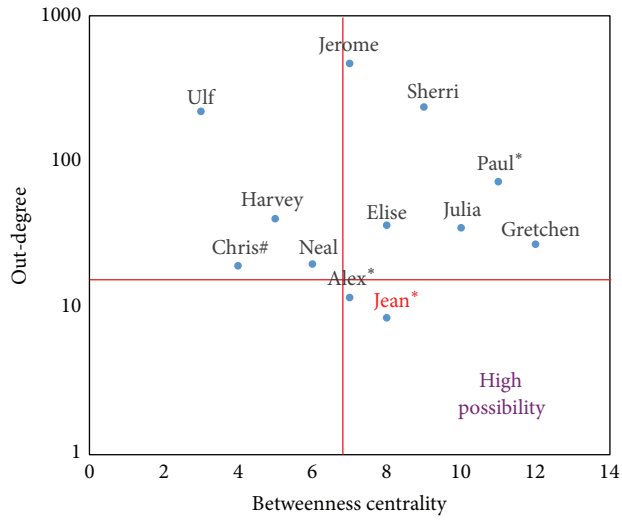

Figure 9: Joint distribution of betweenness centrality and out-degree. (Yao's out-degree is 0, which means he could not be the leader empirically.)

Step 3. Calculate the importance of v_i.

(a) Calculate d'_{lj} of all node pairs (v_l, v_j) after contraction.

(b) Calculate $\Phi(G * v_i)$.

(c) Calculate $\mathrm{ICM}(v_i)$.

5.2. Leader Rank Result. Suspicion score q_i of each node can be obtained from results of former model. Nodes with suspicion score bigger than 70% can be taken into node contraction method. Because the final purpose is identifying the leader, we only consider people with high suspicion score and compare the importance of these people. The person with biggest importance value is judged as leader of the conspiracy. The result is shown in Figure 8.

From Figure 8, Jean ranks first among all 13 people. So, Jean can be considered to be the leader of the conspiracy network.

5.3. Centrality Theory Support. Centrality-based analysis of criminal networks finds that a leader of a criminal organization tends to carefully balance out-degree and betweenness-centrality. It has been proposed that the leader usually maintains a high betweenness centrality but a relatively low out-degree, for enhancing efficiency while ensuring safety [2].

In Figure 9, Jean has high relatively betweenness centrality with relatively low out-degree, which is in accord with his identity of a leader. Thus, our conclusion that Jean is the leader is thus empirically supported by centrality theory.

6. Model Promotion

Networks have a typical pattern that they all consist of nodes and links. In most cases, nodes and links contain many related information. So, such kind of network can be analyzed by mathematical method. For our model takes full account of interactions among nodes and weights of different related information, it can also be utilized in other similar cases, such as: social network and biological network.

A common approach of network analysis is presented below.

(1) Observe the characteristics of the network and transform the complex relations among each individual into an abstract mathematical network with each individual as nodes and relations as links or edges.

(2) If detailed information of the network is unknown, figure out the basic parameters of each node within the network, degree and centrality. These parameters can roughly reflect the importance of each person.

(3) For some certain cases, semantic network analysis should be applied to weigh the suspicion degree to increase accuracy. What is more, the effect of interconnection among individuals should be considered. Combining suspicion degree with interaction, a cumulative score of each node can be calculated. After iteration, one can identify, prioritize, and categorize similar nodes in a network database.

(4) Use node contraction method to measure the importance of each individual within a small ensemble. The core of the network is the one with biggest importance.

In a contagion network, the source of disease can be found. Infected individuals and uninfected ones can be segregated by making use of our model. This model is beneficial for the institution of disease control and prevention to prevent contagion spreading.

7. Conclusions and Further Discussion

In this paper, we have proposed a crime busting model with two dynamic ranking algorithms—cumulative nominating

algorithm and node contraction algorithm in order to detect the likelihood of a suspect and their leader in a complex social network. The contributions and further discussion of our results are list as follows.

7.1. Contributions

(i) *Comprehensive*: we take both the message and node position into consideration for identifying, prioritizing, and categorizing. So, the solution of our model pursues high credibility, while reducing the misjudgment rate.

(ii) *Reasonable*: the result of our model matches perfectly with the experience, which proves the rationality and correctness of our model.

(iii) *Extendable*: the result of simulation shows that our model can be applied in other fields, not just crime busting.

(iv) *Flexible*: we cannot judge a person to be conspiratorial or innocent only based on the message traffic. Since everything may be an accident, our model has its false positive rate which allows the unexpected things to happen.

7.2. Further Discussion. In fact, our research on crime network is just remaining in the beginning, and many problems are waiting to be done. First, the criminal psychology is not taken into consideration while some simple examples show that some people may lie during the taping. Second, since there is no clear criteria for the classification, those conspirators who are slightly behind may be changed while the conspirators ranking in the front remain unchanged.

In near future, many areas should be studied further, for instance, how to apply semantic network analysis more efficiently to discover the potential linkage between the messages and scientifically classify them into different groups, how to dynamically and automatically select reasonable criteria for the classification, and so forth.

Appendix

Data Declarations for ICM 2012 Problem

(i) "Elsie" is given as one of the known conspirators. It is an important data in this problem. However, there are two "Elsie" with node number "7" and "37". According to some statistics about the message with suspicious topics, it suggests that Elsie with number 7 is more likely to be the known conspirator than the other Elsie. Therefore, we consider Elsie with node number 7 as known conspirator.

(ii) There are two "Gretchen" with node number "4" and "32." After analyzing some basic statistics, "Gretchen 32" is found to have more message exchanges than "Gretchen 4." For common sense, managers have more communication than others. So, Gretchen with node number 32 should be regarded as one of the senior managers in this problem.

(iii) "Topic 18" appears in line 215 of "Messages.xls," but "Topic.xls" only contains 15 topics. So, we ignore this data to correct the error.

(iv) "Dolores" is misspelled as "Delores" in "name.xls." This small error should be fixed.

References

[1] V. E. Krebs, "Mapping networks of terrorist cells," *Connections*, vol. 24, pp. 43–52, 2002.

[2] C. Morselli, "Assessing vulnerable and strategic positions in a criminal network," *Journal of Contemporary Criminal Justice*, vol. 26, no. 4, pp. 382–392, 2010.

[3] J. Xu and H. Chen, "Untangling criminal networks: a case study," in *Intelligence and Security Informatics*, vol. 2665, pp. 232–248, 2003.

[4] C. Arney and K. Coronges, "Judges' commentary: modeling for crime busting," *UMAP Journal*, vol. 33, no. 3, p. 293, 2012.

[5] COMAP, "2012 ICM problem," 2012, http://www.comap.com/undergraduate/contests/mcm/contests/2012/problems/.

[6] J. F. Sowa, "Semantic networks," in *Encyclopedia of Cognitive Science*, 2006.

[7] R. Feldman and J. Sanger, *The Text Mining Handbook: Advanced Approaches in Analyzing Unstructured Data*, Cambridge University Press, 2006.

[8] R. Bilisoly, *Practical Text Mining with Perl*, vol. 2, John Wiley & Sons, Hoboken, NJ, USA, 2008.

[9] M. Sharples, D. Hogg, C. Hutchinson, S. Torrance, and D. Young, *Computers and Thought: A Practical Introduction to Artificial Intelligence*, MIT Press, 1989.

[10] U. Gretzel, "Social network analysis: introduction and resources," 2001, http://lrs.ed.uiuc.edu/tse-portal/analysis/social-network-analysis/.

[11] Y. Freund, R. Iyer, R. E. Schapire, and Y. Singer, "An efficient boosting algorithm for combining preferences," *Journal of Machine Learning Research*, vol. 4, no. 6, pp. 933–969, 2003.

[12] A. Meaden and D. Hacker, *Problematic and Risk Behaviours in Psychosis: A Shared Formulation Approach*, Routledge, 2010.

[13] J. A. Zhang and X. E. Guo, "Trust model based on dynamic recommendation in P2P network," *Computer Engineering*, vol. 36, no. 1, pp. 174–180, 2010.

[14] F. Autrel, N. Cuppens-Boulahia, and F. Cuppens, "Reaction policy model based on dynamic organizations and threat context," in *Data and Applications Security XXIII*, vol. 5645 of *Lecture Notes in Computer Science*, pp. 49–64, Springer, 2009.

[15] Y. J. Tan, J. Wu, and H. Z. Deng, "Evaluation method for node importance based on node contraction in complex networks," *System Engineering Theory and Practice*, vol. 26, no. 11, pp. 79–83, 2006.

[16] M. Perkowitz and O. Etzioni, "Adaptive web sites: conceptual cluster mining," in *Proceedings of the 16th International Joint Conference on Artificial Intelligence*, pp. 264–269, July 1999.

[17] R. Cooley, B. Mobasher, and J. Srivastava, "Web mining: information and pattern discovery on the World Wide Web," in *Proceedings of the IEEE 9th IEEE International Conference on Tools with Artificial Intelligence*, pp. 558–567, November 1997.

[18] J. M. Kleinberg, "Authoritative sources in a hyperlinked environment," *Journal of the ACM*, vol. 46, no. 5, pp. 604–632, 1999.

[19] G. W. Flake, S. Lawrence, and C. L. Giles, "Efficient identification of web communities," in *Proceedings of the 6th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 150–160, August 2000.