Chapter 24
Investigation of Key-Player Problem in Terrorist Networks Using Bayes Conditional Probability

D.M. Akbar Hussain

24.1 Introduction

The communication disruption in social networks occurs when it is fragmented by the removal of a set of nodes, which means the said set of nodes are the key actors in the network. The key player problem (KPP) has two contexts; in the first case KPP-1 we have to find a set of k-nodes which may be called as kp-set of order k and if we remove this kp-set it will severely damage communication among the remaining nodes. Whereas in the second case KPP-2 we also find kp-set of order k which is maximally connected to all other nodes [1]. In the first context KPP-1, fragmentation occurs or distance between nodes becomes very large so in realistic terms practically disconnected. In the second context KPP-2, how many and who can reach as many nodes as possible directly or alternatively through shorter path ways. It has been pointed out that the centrality measures alone from graph theory for example betweenness, degree and closeness are not enough to solve the key player problem [1, 2]. Consider for example the kite network as shown in Fig. 24.1 from David Krackhardt, it can be seen that if we remove node 1 which has the highest degree centrality the communication setup remains intact for the remaining nodes as seen in Fig. 24.2. On the other hand if we remove node 8 which has the highest betweenness centrality measure, although the network is fragmented in two components but still the major portion of the network remains intact as seen in Fig. 24.3. Selecting a set of nodes to solve KPP-1 or KPP-2 problem is not the same as the selection of equal number of individual nodes to get an optimal KPP solution [3]. The reason is that some time their is redundancy with respect to node's position and its removal does not necessarily change the fragmentation as the nodes are essentially connected via other nodes. Borgatti presented an optimal measure to solve KKP-1 problem by selecting a kp-set in such a way that it maximize fragmentation and distance. The fragmentation measure suggested by him counts the number

D.M.A. Hussain (✉)
Automation and Control, Department of Electronic Systems, Aalborg University, Niels Bohrs Vej 8, 6700 Esbjerg, Denmark
e-mail: akh@es.aau.dk

B. Furht (ed.), *Handbook of Social Network Technologies and Applications*,
DOI 10.1007/978-1-4419-7142-5_24, © Springer Science+Business Media, LLC 2010

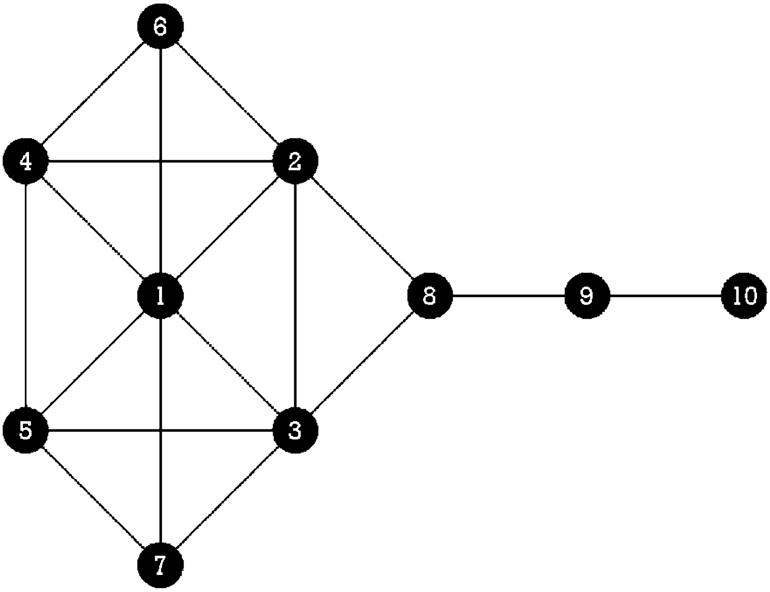


Fig. 24.1 Kite network by David Krackhardt

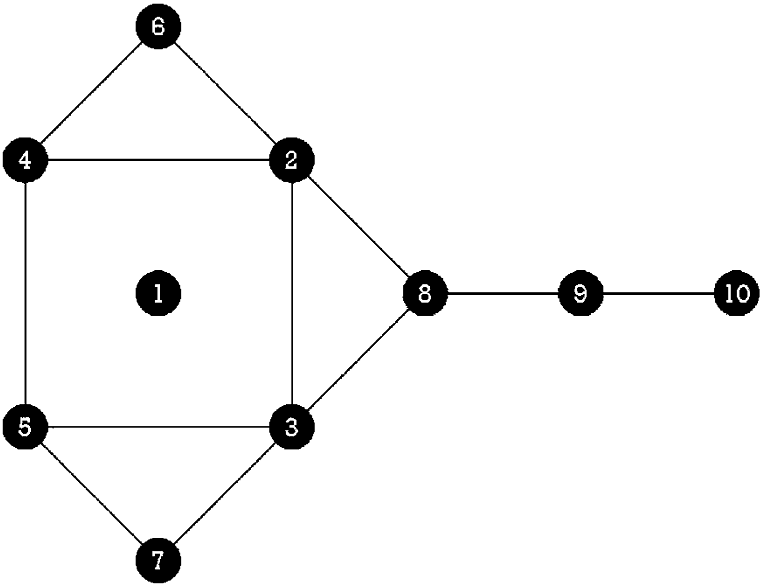


Fig. 24.2 Node with highest degree is removed

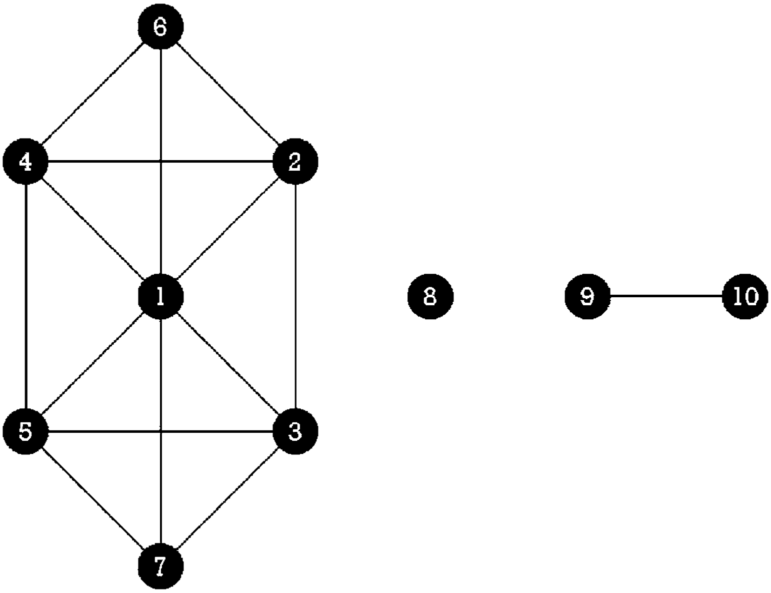


Fig. 24.3 Node with highest betweenness is removed

of pairs of nodes that are disconnected from each other for example a given matrix A in which $A_{ij} = 1$ if i can reach j otherwise $A_{ij} = 0$, fragmentation measure F is defined as:

$$F = 1 - \frac{2\sum_i \sum_{j<i} A_{ij}}{n(n-1)} \tag{24.1}$$

In the fragmented components nodes are mutually reachable so F can be rewritten based on their sizes S_k as:

$$F = 1 - \frac{\sum_k S_k(S_k - 1)}{n(n-1)} \tag{24.2}$$

The above expression still does not include the shape and structure of the components, the solution is to measure sum of the reciprocals of distances, therefore re-writing (24.1) as;

$$D^F = 1 - \frac{2\sum_{i>j} \frac{1}{d_{ij}}}{n(n-1)} \tag{24.3}$$

the above equation is similar in respect as expression for F and in a similar way could achieve its maximum value 1 if all the nodes are disconnected.

In this paper we present a simulated study to investigate key player problem using Bayes probability theorem discussed later in Sect. 24.4, surprisingly, results are very similar to what Borgatti achieved through its distance fragmentation measure (we shall refer it as fragmentation index in the remaining text description) [1]. In our results presentation we have compared the two methods by removing each node individually and computing the distance fragmentation index, therefore, from these methods we can see both the probability of the individual node and the value of the fragmentation index. Our paper is organized in the following way; it provides a brief survey of SNA in Sect. 24.2, centrality measures and their mathematics used in the social network analysis is discussed in Sect. 24.3. This is necessary as we are using these measures in our proposed computational method. Bayes theory is briefly discussed in Sect. 24.4 in context of our implementation, Sect. 24.5 provides the proposed model, analysis and results, finally conclusion in Sect. 24.6.

24.2 SNA Survey

Social Network Analysis is a mathematical method for ‘connecting the dots’. SNA allows us to map and measure complex, and sometimes covert, human groups and organizations [4]. Given any network where the nodes/agents are individuals, groups, organizations etc., a number of network measures such as centrality or cutpoints are used to locate critical/important nodes/agents. Social network analysis is a multi-model multi-link problem so the challenges posed by such multi-dimensional task are enormous. The standard representation of a typical social network model is through a graph data structure. This type of model can be considered as an intellective simulation model, such types of models explain one particular aspect of the model abstracting other factors present in the model. The dynamics of larger social networks is so complex some time it becomes difficult to understand the various levels of interactions and dependencies just by mere representation through a graph. However, to overcome this limitation many analytical methods provide relationship dependencies, role of different nodes and their importance in the social networks. Insight visualization of any network typically focuses on the characteristics of the network structure. SNA measures indicate various roles of the nodes in a network, for example leaders, gatekeepers, role models etc. A node is central if it is strategically located on the communication route joining pairs of other nodes [5, 6]. Being central it can influence other nodes in the network, in other words potentially it can control the flow of information. The potential of control makes the centrality conceptual model for these nodes. The idea of centrality is not new it was first applied to human communication by Baveles in 1948 [5, 7]. In this study relationship between structural centrality and influence in group processes were hypothesized. Following Baveles it was concluded that centrality is related to group efficiency in problem-solving, perception of leadership and the personal satisfaction of participants [8–10]. In the fifties and sixties more research was conducted on these measures and it was concluded that centrality is relevant to the way groups get organized to solve problems. The following references provide a very deep and pioneering work on these measures [11–20]. The centrality concept is not exclusive to deal with group problem tasks, it has been used in other discipline as well [21, 22].

A number of centrality measures have been proposed over the past years. Most of the centrality measures are based on one of two quite different conceptual ideas and can be divided into two large classes [23]. The measures in the first class are based on the idea that the centrality of an individual in a network is related to how it is near to others. Second class of measures is based on the idea that central nodes stand between others on the path of communication [24–26]. A node being on the path of other nodes communication highway has the potential to control what passes through it. The simplest and most straightforward way to quantify the individual centrality is therefore the degree of the individual, i.e., the number of its immediate neighbors. In a graph if every node is reachable from any node in the graph it is called a connected graph also each path in the graph is associated with a distance equal to the number of edges in the path and the shortest path to reach a given pair of nodes is geodesic distance. Nieminen has provided a very systematic elaboration of the concept of degree [27]. Scott has extended the concept based on degree beyond immediate (first) neighbors by selecting the number of nodes an individual can reach at a distance two or three [28]. Similarly, Freeman produced a global measure based on the concept of closeness in terms of the distances among the various nodes [25]. The simplest notion of closeness is obtained by the sum of the geodesic distances from an individual to all the other nodes in the graph [29]. These traditional social network measures and the information processing network measures can help in revealing importance and vulnerabilities of the nodes/agents in the network. Since the start of this century many terrorism events have occurred around the globe. These events have provided a new impetus for the analysis, investigation, studying the behavior and tracking terrorist networks (individuals).

Apart from centrality measures social scientist have also developed highly efficient techniques like data mining and decision making tree methods to process large amount of data. Data Mining technique extract particular kind of information from this huge data. Typically, once a particular information is located the data mining application alerts either system or the human operator which determines whether the application has provided the requested information. Data mining also allows to record the search process, so that patterns of objects and information can be visualized as graph. This visualization is quite useful for large amount of data information. In the beginning data mining methodology has been developed largely for businesses applications to help with marketing it also has applications in medical profession. However, more recently it has been used in law enforcement and intelligence operations [30]. The development and implementation of these systems require a cooperative effort on the part of those who develop and those who operate them. The importance of such systems is that they must provide complete information based on the input and typically sound alarm when targeted information is located however, the final action or judgment is still made by the user. On the other hand decision tree methodology can be used to make decisions. The core idea behind decision tree technique is to correctly locate and identify the choice options which are explicitly evaluated in terms of the importance of their outcome. The probability of that outcome is used in creating a sequence of decision map from start to end. The most positive aspect of this method is that decision is made explicit so that others can use the decision tree if faced with the similar questions. Similar to data mining techniques for application in law enforcement and intelligent operations decision trees can also be used to guide decisions. Both of these tools; data mining and decision tree have applications in the analysis of social networks.

Application of the above mentioned tools/concepts on the complex socio-technical systems like SNA is very demanding to squeeze out the required information. Most of the above mentioned measures and tools work best when the data is complete; i.e., when the information is inclusive about the interactions among the nodes. However, the difficulty is that large scale distributed, covert and terrorist networks typically have considerable missing data. Normally, a sampled snapshot data is available, some of the links may be intentionally hidden (hence missing data may not be randomly distributed). Also data is collected from multiple sources and at different time scales and granularity. In addition inclusive and correct information may be prohibitive because of secrecy. Obviously, there could be other difficulties but even these provide little guidance for what to expect when analyzing these complex socio-technical systems with the existing tools. Therefore, new concepts are really required in this area for better understanding/investigation of these nodes one such example is the fragmentation concept from Borgatti [1]. Our method could also be regarded as a new way to locate important key actors in a terrorist network analysis context, which combines the standard centrality measures into Bayes conditional probability method.

Typically, one has to identify the following characteristics in the context of SNA:

(a) Key players in the network
(b) Potential threat from the network
(c) Important individual, event, place or group
(d) Dependency of individual nodes
(e) Leader-Follower identification
(f) Bonding between nodes
(g) Vulnerabilities identification
(h) Efficiency of overall network

Kathleen Carley has also provided the following key characteristics for classification and distinctiveness of nodes [31].

(a) An individual or group that if given new information can propagate it rapidly
(b) An individual or group that has relatively more power and can be a possible source of trouble, potential dissidents, or potential innovators
(c) An individual or group where movement to a competing group or organization would ensure that the competing unit would learn all the core or critical information in the original group or organization (inevitable disclosure)
(d) An individual, group, or resource that provides redundancy in the network

24.3 SNA Measures

Social networks provides mapping and the social network analysis measure relationships and movement between people, groups, events, organizations or other information/knowledge processing entities. People, organization and groups are represented as nodes in the network while the links show relationships or movement between the nodes. SNA provides both visual and mathematical analysis of human relationships. This methodology could also be used by the management to perform Organizational Network Analysis [4]. There are many ways to determine important members of a network. The most straightforward technique is to compute member's **degree**; the number of direct connections to other members of the network apart from **degree** more well known measures are **betweenness** and the **closeness**. A node with relatively few direct connections could still be important if it lies between two or more large groups. On the other hand a member could also be important if it has direct and indirect links in such a way that it is placed closest to all other members of the group, in other words the node has to go through fewer intermediaries to reach other members than anyone else. It is important to note that terrorist cells have complex, dynamical and decentralized structures and these standard measures alone may not be enough to reveal information about important actors in the network. SNA has been used with other measures to highlight important nodes in terrorist cells [32–34], other applications like Googles PageRank systems is using the concept of network theory and centrality, in medical field network analysis has been used to track the spread of HIV, more recently a very interesting research has been carried out for the understanding of relationships from Enron's email records [35].

24.3.1 Degree

To comprehend networks and their participants, we evaluate the location of participants in the network. Degree provides the relative importance and the location of a particular node in the network. Typically, centrality means degree, with respect to communication a node with relatively high degree looks important. In a social network a node that is directly connected with many other nodes actually see itself and be seen by others in the network as indispensable. This means a node with low degree is isolated from direct involvement and see itself and by others not to be a stakeholder. A general measure of centrality $D_c(p_i)$ based on degree for a node p_i is given by [25];

$$D_c(p_i) = \sum_{j=1}^{n} d(p_j,\ p_i) \quad (\mathit{for\ all}\ j \neq i) \tag{24.4}$$

where

$$d(p_j,\ p_i) = \begin{cases} 1 & \text{if } p_j,\ p_i \text{ directly connected} \\ 0 & \text{otherwise} \end{cases}$$

A node can be connected with maximum of $(n - 1)$ number of nodes in a n size network. Therefore, the maximum degree value is $(n - 1)$, so to have a relationship which is proportion of other nodes that are directly connected to p_i can be written as follows.

$$D_c^{'}(p_i) = \frac{\sum_{j=1}^{n} d(p_j,\ p_i)}{(n\ -\ 1)} \tag{24.5}$$

24.3.2 Betweenness

Betweenness (also called load) measures to what extent a node can play the role of intermediary in the interaction between the other nodes. The most popular and simple betweenness measure based on geodesic path is proposed by Freeman and Anthonisse [24,26]. In many real scenarios however, communication does not travel exclusively through geodesic paths. For such situations two more betweenness measures are developed first based on all possible paths between couple of nodes [36] and second based on random paths [37]. Consider a graph $G = (V, E)$ with vertices V and edges E, a path from a source vertex to a target vertex is an alternating sequence of edges. The length of this path is the total number of edges from source to target and shortest path of these alternating routes is called the *geodesic*. Therefore, nodes located on many shortest paths (geodesics) between other nodes will have higher betweenness compared with others. For a graph $G = (V, E)$ with n vertices, the betweenness $B_c(k)$ for a vertex k is:

$$B_c(k) = \sum_{i \neq j, i \neq k} \frac{\sigma_{ij}(k)}{\sigma_{ij}} \tag{24.6}$$

where σ_{ij} is the number of shortest paths from i to j, and $\sigma_{ij}(k)$ is the number of shortest geodesic paths from i to j that pass through vertex k. It can be normalized by dividing through the number of pairs of vertices not including k, which is $(n-1)(n-2)$. Calculation of betweenness is quite complicated for networks when several geodesics connect a pair of nodes, which is the case in most real world networks. Also, $B_c(k)$ is dependent on the size of the network on which it is being calculated. Freeman [25] has provided relative centrality of any node in the network by the following relationship.

$$B_c^{'}(k) = \frac{B_c(k)}{(n^2 - 3n + 2)/2} \tag{24.7}$$

The idea is that maximum value of $B_c(k)$ is achieved by the central point of the star that is given by;

$$\frac{(n^2 - 3n + 2)}{2} \tag{24.8}$$

Therefore, the relative betweenness centrality is determined by the ratio given in (24.7) and is re-written as (24.9).

$$B_c^{'}(k) = \frac{2B_c(k)}{(n^2 - 3n + 2)} \tag{24.9}$$

24.3.3 Closeness

A more sophisticated centrality measure closeness based on geodesic distance can be defined, which is the mean geodesic (i.e., shortest path) distance between a node and all other nodes reachable from it. Closeness can be regarded as a measure of how long it will take information to spread from a given node to other nodes in the network. From retrospect closeness can provide the information about nodes independence. We are utilizing the closeness centrality in our implementation, so it is necessary to provide brief detail about closeness to complete the discussion on standard centrality measures typically used in SNA. The simplest mathematics for closeness centrality is provided by [29], which is determined by summing the geodesics from a node of interest to all other nodes in the network and taking its inverse. Closeness grows as the distance between node i and other nodes for example $(j....n)$ increases. The Closeness C_c for a node i is given by;

$$C_c(i) = \frac{1}{\sum_{j=1}^{n} d(p_j,\ p_i)} \tag{24.10}$$

where d is the geodesic distance between respective nodes, for all those nodes which are not connected the geodesic distance is infinity. The above expression is dependent on the size (number of nodes) of the network and it is appropriate to have an expression which is independent of this limitation. Beauchamp [38] suggested that relative Closeness (point centrality) for a node i is given by;

$$C_c^{'}(i) = \frac{(n\ -\ 1)}{\sum_{j=1}^{n} d(p_j,\ p_i)} \tag{24.11}$$

24.4 Bayes Probability Theorem

Bayes' Theorem is a simple mathematical formula used for calculating conditional probabilities. Bayes' Theorem originally stated by Thomas Bayes and it has been used in a wide variety of contexts, ranging from marine biology to the development of "Bayesian" Spam blockers for email systems. Through the use of Bayes' Theorem precise measures can be obtained by showing how the probability that a theory is correct is affected by new evidence [39, 40]. In a Bayesian framework the conditional and marginal probabilities for stochastic events for example A and B are computed through this relationship:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)} \tag{24.12}$$

$$P(A|B) \propto P(B|A)\ P(A) \tag{24.13}$$

where $P(A)$ is the prior probability or marginal probability of A, $P(A|B)$ is the conditional probability given B also called posterior probability. $P(B|A)$ is conditional probability given A, $P(B)$ is prior probability and considered as normalizing constant. $P(B|A)$ is in-fact equal to $L(A|B)$ which is the likelihood of A given fixed B, however, at times likelihood L can be multiplied by a factor so that it is proportional to, but not equal probability P. It should be noted that probability of an event A conditional on another event B is generally different from the probability of B conditional on event A, however, there is a unique relationship between the two which is provided by Bayes theorem. We can formulate the above relationship as:

$$posterior = \frac{likelihood\ \times\ prior}{normalizing\ constant} \tag{24.14}$$

We can re-write (24.12) as the ratio $P(B|A)/P(B)$ which is typically called as standardized likelihood or normalized likelihood so it can be written as:

$$posterior = normalized\ likelihood\ \times\ prior \tag{24.15}$$

Bayes conditional probability from above expression can be expanded as;

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B|A)\ P(A)\ +\ P(B|A')\ P(A')} \tag{24.16}$$

$P(A)$ is the probability regardless of any other information; $P(A')$ is the complimentary event of A; $P(B|A)$ is the conditional probability; $P(B|A')$ is the probability of B given $A^{'}$.

Bayesian approach have been used in dynamic SNA issues, statistical analysis and network measurement [40–43], our approach here is different, we are interested in evaluating the theory or hypothesis (24.12) for A based on B which is the new information (evidence) that can verify the hypothesis and $P(A)$ is our best estimate of the probability (known as the prior probability of A, prior to considering the new information). In other words we are interested to compute the probability that A is correct (true) with the assumption that the new information (evidence) is correct. The detailed implementation is discussed in the following Sect. 24.5.

24.5 Analysis & Results

We explain our model of computation by considering a network of 20 nodes as shown in Fig. 24.4, such a network is intentionally considered with some nodes having very large value of degree and vice versa, similar is the case for betweenness. We need four terms of the Bayes probability expression given by (24.16). $P(A)$ is the prior probability which is computed from the values of closeness and betweenness, the second term $P(A')$ which is complimentary mentioned earlier can be computed as $(1-P(A))$. Now the conditional probability $P(B|A)$ is taken as the betweenness over the entire network, similarly, $P(B|A')$ is taken as the degree over the entire network, all these values are given in Table 24.1. In order to see graphically which node has higher value of degree or betweenness, we have also shown the network through the nodes individual size comparison for degree and betweenness in Figs. 24.5 and 24.6 respectively.

There could be many situations of interest for the analysis of this network, let us consider two cases for this network; node 1 and 11 and assume that these are key player nodes and we eliminate them from the network as shown in Figs. 24.7 and 24.8 respectively.

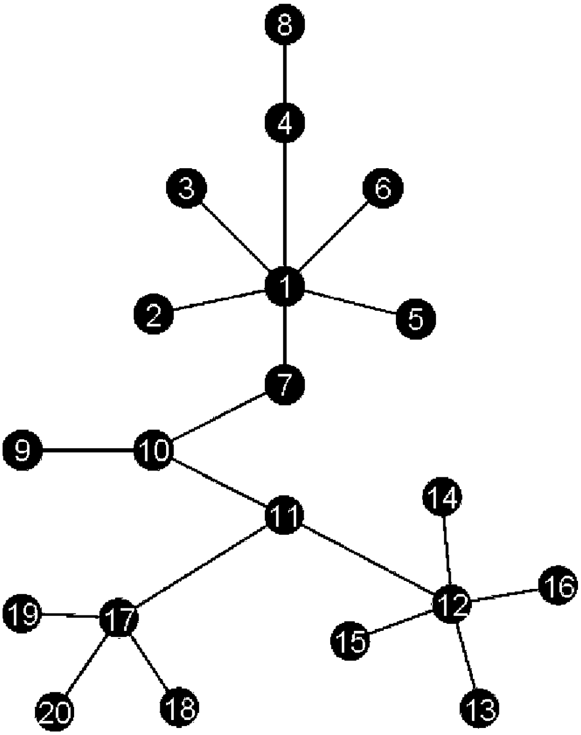
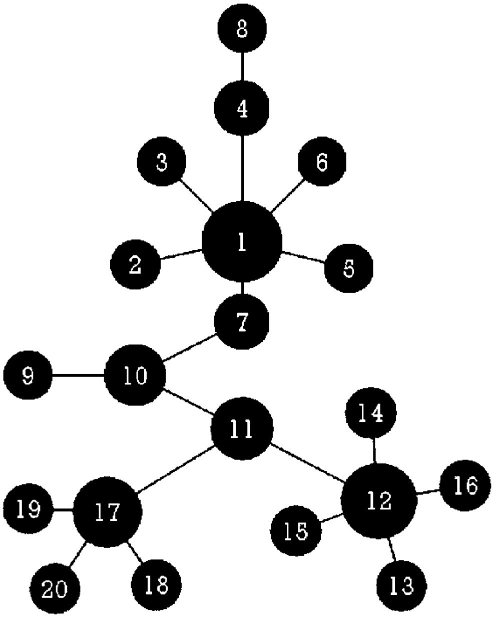


Fig. 24.4 Computational model

Table 24.1 Term values

Node numbers	$P(A)$	$(1 - P(A))$	$P(B\|A)$	$P(B\|A')$
1	0.409555	0.590445	0.497076	0.315789
2	0.123377	0.876623	0.000000	0.052632
3	0.123377	0.876623	0.000000	0.052632
4	0.158830	0.841170	0.064327	0.105263
5	0.123377	0.876623	0.000000	0.052632
6	0.123377	0.876623	0.000000	0.052632
7	0.404391	0.595609	0.450292	0.105263
8	0.102151	0.897849	0.000000	0.052632
9	0.141791	0.858209	0.000000	0.052632
10	0.459959	0.540041	0.532164	0.157895
11	0.495047	0.504953	0.602339	0.157895
12	0.342303	0.657697	0.362573	0.263158
13	0.123377	0.876623	0.000000	0.052632
14	0.123377	0.876623	0.000000	0.052632
15	0.123377	0.876623	0.000000	0.052632
16	0.123377	0.876623	0.000000	0.052632
17	0.296089	0.703911	0.280702	0.210526
18	0.120253	0.879747	0.000000	0.052632
19	0.120253	0.879747	0.000000	0.052632
20	0.120253	0.879747	0.000000	0.052632

Fig. 24.5 Degree comparison



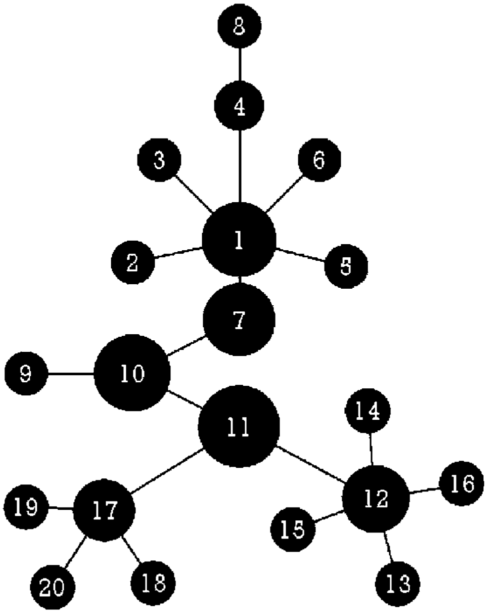


Fig. 24.6 Betweenness comparison

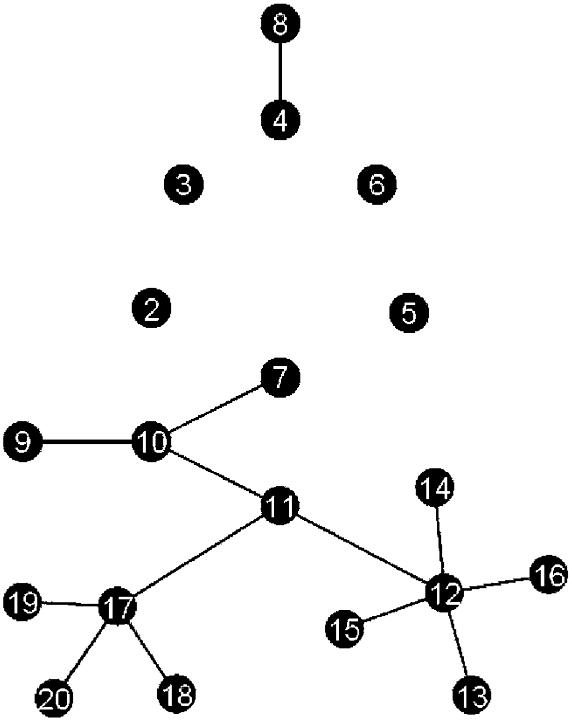


Fig. 24.7 Network structure with node 1 eliminated

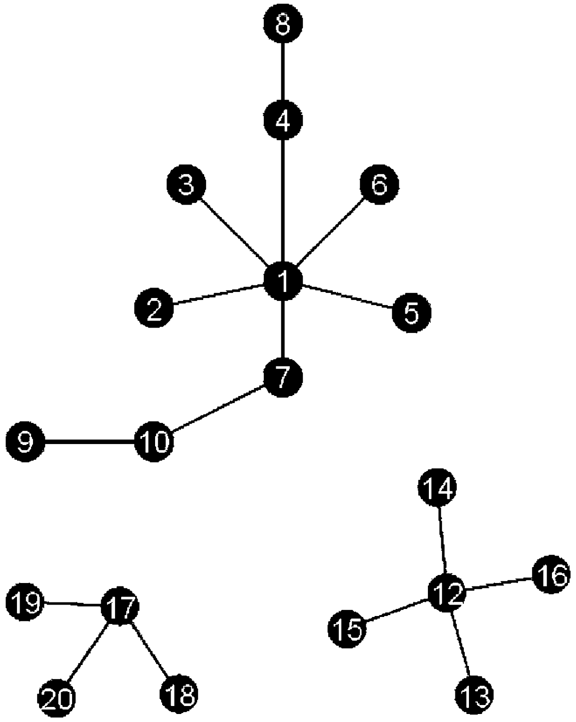


Fig. 24.8 Network structure with node 11 eliminated

Now if we analyze the structure of the network we can see that there are mainly three groups, one around node 1, second around node 12 and third around node 17, when node 1 is eliminated essentially first group is severely damaged but the other two groups are intact, the fragmentation index value is approximately 0.6 and our computed probability is approximately 50%, which is reasonable as it has effected nearly that percentage of the network. However, when we look at the case of node 11, it is quite evident that it has effected all three groups and the network is fragmented into three separate mini components, the fragmentation index is approximately 0.6 and the computed probability is much higher 80% indicating how important node 11 is in the network. The reason of higher probability value is because of our computational model, which is biased towards betweenness/closeness and as the betweenness for this particular node is quite high which caused induction and raising the probability value. Figure 24.9 has shown the probability values for all the nodes and it can be seen that nodes 1, 7, 10 and 11 are the key players in this network having over 50% probability.

The second network we have selected for our results is the kite network from David Krackhardt as shown in Fig. 24.1. In this network node 1 has the largest value in terms of degree and node 8 has the highest value of betweenness. As it is said earlier that it is also a unique network that even if we remove node 8 having largest betweenness the major portion of the network still remains intact and removal of

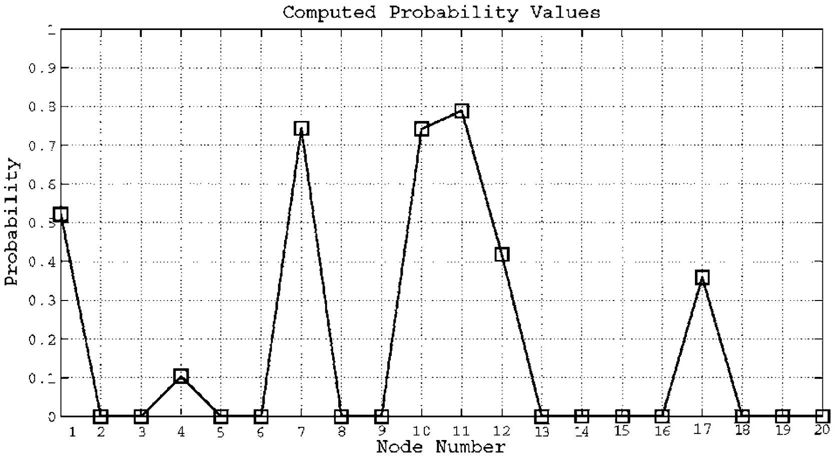


Fig. 24.9 Bayes conditional probability (random network)

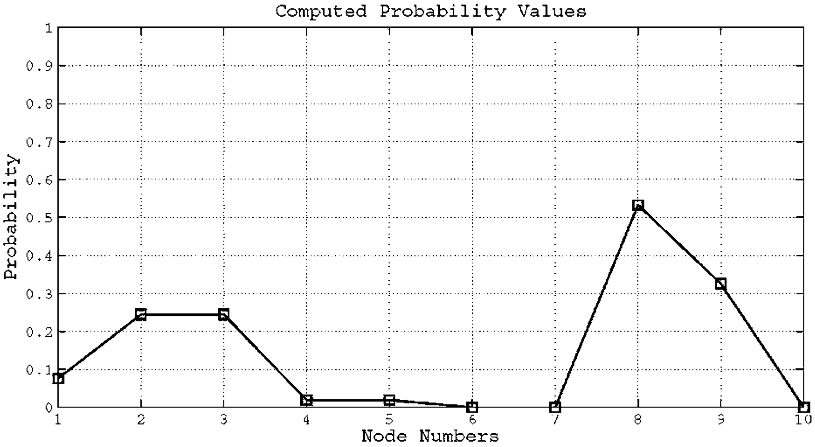


Fig. 24.10 Bayes conditional probability (Kite network)

node 1 cause no significance change and communication is still possible between most nodes but may be with a little extra overhead as the path ways are increased (structures are shown already in Figs. 24.2 and 24.3). The fragmentation index of node 1 if it is eliminated is very very low close to zero indicating that it not the key player and similarly probability computed is also low, less than 10%. The fragmentation index is little higher for the case of node 8 (approximately 0.2) and the computed probability is increased to about 50% , which in contrast to fragmentation index value is quite large the reason is again the same that the computed probability is biased towards betweenness/closeness which is the new evidence for correcting our earlier assumption in the Bayes conditional probability theorem. The conditional probability for each node of this network is shown in Fig. 24.10 for comparison, it

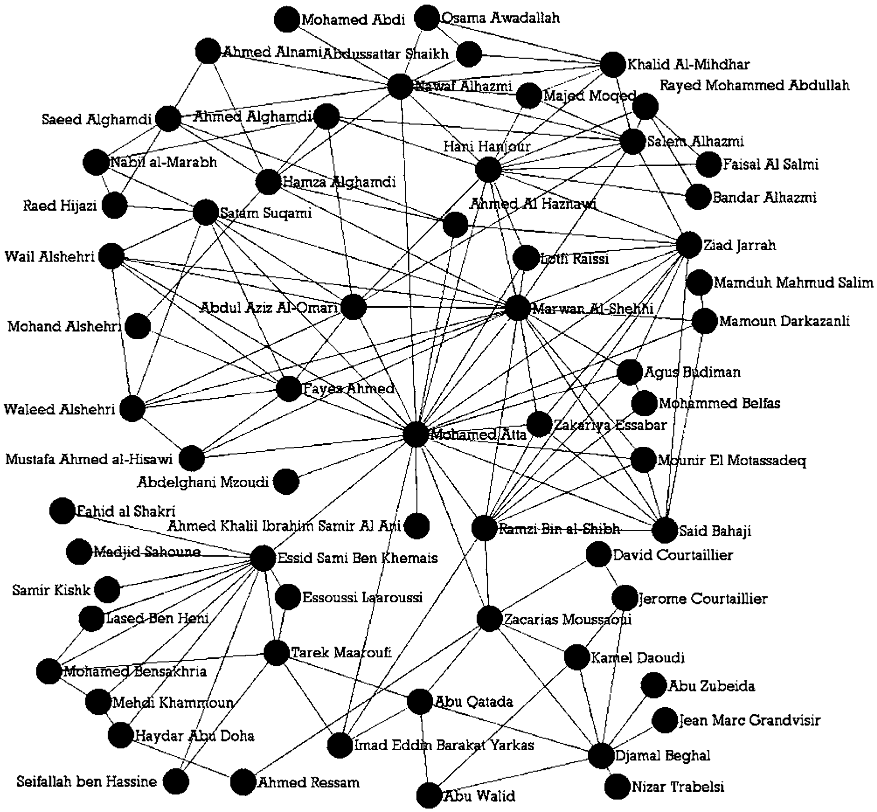


Fig. 24.11 9–11 Hijackers network (Valdis E. Krebs)

shows that node 2, 3 and 9 are the possible candidates for key player position apart from 9. It should be noted that the system has out rightly rejected node 1 to be a key player even though it has the largest value of degree centrality. The last network for the analysis is the 9–11 hijackers network [4] of 62 nodes as shown in Fig. 24.11 much has been said about this network our purpose of considering it here is to bench mark our results as most information about this network is complete. The corresponding node labels are given in Table 24.2 for convenience as we are more familiar with the names of these hijacker in contrast to their position/number in the network.

Table 24.3 shows the computed Bayes terms for the whole network. For this network we can consider many situations to see the important nodes, we all know that how important Mohamed Atta (node 33) was, so we will consider three situations node 20 (Essid Sami Ben Khemais), node 33 (Mohamed Atta) and node 39 (Marwan Al-Shehhi). Node 33 has both high degree and high betweenness measure, in contrast to this node 20 has high degree and reasonable betweenness on the other hand node 39 has low betweenness and very high degree measure. To visualize the effect

Table 24.2 Hijackers name and node numbers

Node numbers	Names
1	Abu Zubeida
2	Jean-Marc Grandvisir
3	Nizar Trabelsi
4	Abu Walid
5	Djamal Beghal
6	Ahmed Ressam
7	Kamal Daudi
8	Jerome Courtaillier
9	Haydar Abu Doha
10	Mehdi Khammoun
11	Abu Qatata
12	Zacarias Moussaoui
13	David Courtaillier
14	Essoussi Laaroussi
15	Mohamed Bensakhria
16	Tarek Maaroudfi
17	Lased Ben Heni
18	Imad Eddin Barakat Yarkas
19	Seifallah ben Hassine
20	Essid Sami Ben Khemais
21	Fahid al Shakri
22	Mohammed Belfas
23	Adelghani Mzoudi
24	Ramzi Bin al-Shibh
25	Madjid Sahoune
26	Agus Budiman
27	Mounir El Motassadeq
28	Ahmed Khalil Ibrahim Samir
29	Samir Kishk
30	Mustafa Ahmed al-Hisawi
31	Zakariya Essabar
32	Mamduh Mahmud Salim
33	Mohamed Atta
34	Mamoun Darkazanli
35	Said Bahaji
36	Fayez Ahmed
37	Ziad Jarrah
38	Wail Alshehri
39	Marwan Al- Shehhi
40	Waleed Alshehri
41	Abdul Aziz Al Omari
42	Lotfi Raissi
43	Bandar Alhazmi
44	Satam Suqami
45	Ahmed Al Haznawi

(continued)

Table 24.2 (continued)

Node numbers	Names
46	Hani Hanjour
47	Rayed Mohammed Abdullah
48	Mohand Alshehri
49	Salem Alhazmi
50	Ahmed Alghamdi
51	Faisal Al Salmi
52	Majed Moqed
53	Nabil al-Marabh
54	Hamza Alghamdi
55	Raed Hijazi
56	Nawaf Alhazmi
57	Saeed Alghamdi
58	Khalid Al-Mihdhar
59	Ahmed Alnami
60	Osama Awadallah
61	Aduussattar Shaikh
62	Mohamed Abdi

Table 24.3 Term values

Node numbers	$P(A)$	$(1 - P(A))$	$P(B\|A)$	$P(B\|A')$
1	0.125000	0.875000	0.000000	0.016393
2	0.125000	0.875000	0.000000	0.016393
3	0.125000	0.875000	0.000000	0.016393
4	0.132581	0.867419	0.001093	0.049180
5	0.219540	0.780460	0.107559	0.131148
6	0.165877	0.834123	0.007286	0.032787
7	0.162084	0.837916	0.008106	0.065574
8	0.158851	0.841149	0.001639	0.065574
9	0.164033	0.835967	0.007013	0.049180
10	0.154314	0.845686	0.000546	0.049180
11	0.190337	0.809663	0.037978	0.081967
12	0.331109	0.668891	0.226503	0.131148
13	0.153266	0.846734	0.000000	0.032787
14	0.158031	0.841969	0.000000	0.032787
15	0.161826	0.838174	0.004281	0.065574
16	0.187560	0.812440	0.032423	0.098361
17	0.152500	0.847500	0.000000	0.032787
18	0.221774	0.778226	0.034153	0.065574
19	0.158031	0.841969	0.000000	0.032787
20	0.342359	0.657641	0.252095	0.180328
21	0.151741	0.848259	0.000000	0.016393

(continued)

Table 24.3 (continued)

Node numbers	$P(A)$	$(1 - P(A))$	$P(B\|A)$	$P(B\|A')$
22	0.163978	0.836022	0.000000	0.032787
23	0.185976	0.814024	0.000000	0.016393
24	0.246314	0.753686	0.053780	0.163934
25	0.151741	0.848259	0.000000	0.016393
26	0.201023	0.798977	0.011020	0.081967
27	0.195513	0.804487	0.000000	0.065574
28	0.185976	0.814024	0.000000	0.016393
29	0.151741	0.848259	0.000000	0.016393
30	0.196970	0.803030	0.002914	0.065574
31	0.196774	0.803226	0.000000	0.065574
32	0.140553	0.859447	0.000000	0.016393
33	0.582919	0.417081	0.579299	0.360656
34	0.201508	0.798492	0.016940	0.065574
35	0.200599	0.799401	0.002505	0.114754
36	0.217631	0.782369	0.025865	0.131148
37	0.223314	0.776686	0.020055	0.163934
38	0.203444	0.796556	0.002914	0.098361
39	0.276376	0.723624	0.087104	0.295082
40	0.167992	0.832008	0.000820	0.098361
41	0.224958	0.775042	0.023342	0.147541
42	0.209481	0.790519	0.012295	0.081967
43	0.156410	0.843590	0.000000	0.032787
44	0.230922	0.769078	0.049681	0.131148
45	0.207088	0.792912	0.015483	0.065574
46	0.286164	0.713836	0.123798	0.213115
47	0.158441	0.841559	0.000820	0.065574
48	0.156706	0.843294	0.000592	0.032787
49	0.190775	0.809225	0.014080	0.131148
50	0.173888	0.826112	0.006995	0.081967
51	0.156410	0.843590	0.000000	0.032787
52	0.164865	0.835135	0.000000	0.065574
53	0.163696	0.836304	0.002923	0.065574
54	0.192955	0.807045	0.024964	0.114754
55	0.157676	0.842324	0.000920	0.049180
56	0.298492	0.701508	0.154954	0.180328
57	0.172905	0.827095	0.012477	0.098361
58	0.170219	0.829781	0.007104	0.098361
59	0.162234	0.837766	0.000000	0.049180
60	0.155612	0.844388	0.000000	0.049180
61	0.155612	0.844388	0.000000	0.049180
62	0.154040	0.845960	0.000000	0.016393

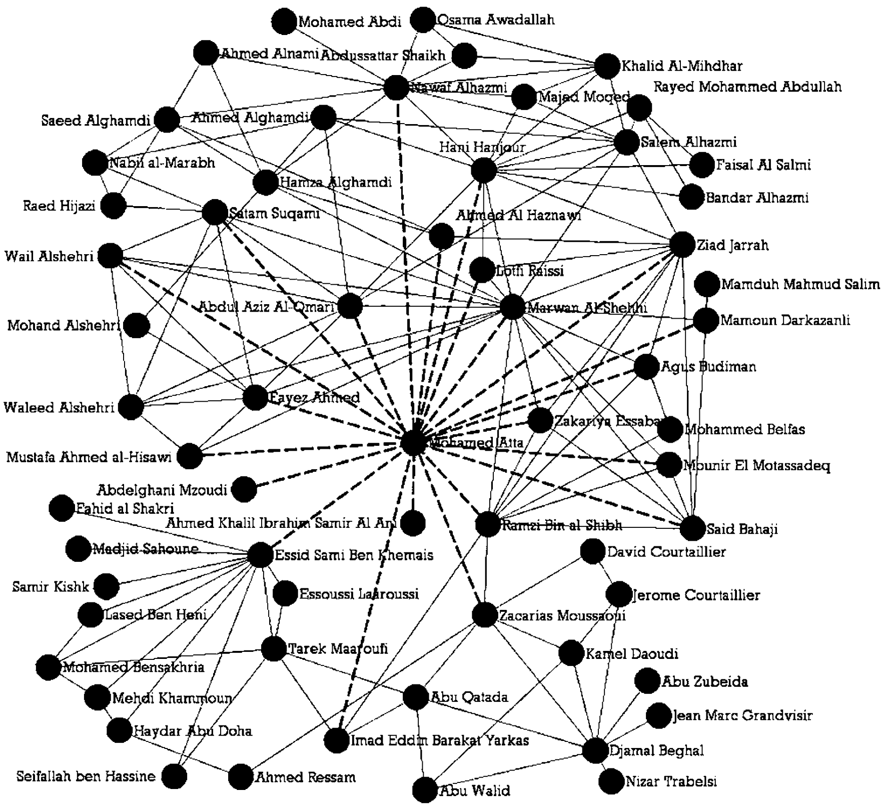


Fig. 24.12 Mohamed Atta eliminated

Figs. 24.12–24.14 shows the structures of the networks once these nodes are eliminated from the network, the removal is manifested through dotted lines on these figures rather than wiping off these connections because removal of one node out of 62 nodes visually does not make a significant impact. It is quite evident from the figures that node 33 and node 20 are the key actor nodes having a probability of 70 and 42% respectively. Whereas the probability value for node 39 is quite low about 10% which is quite right because this node does not have a central role in the network compared with other nodes. Also, there is redundancy because many nodes after its removal are still reachable from via other nodes in the network. The fragmentation indices are approximately 0.36, 0.31 and 0.25 for nodes 33, 20 and 39 respectively. Figure 24.15 shows the computed probability values for the whole network to see potentially possible key players in the network, which is quite in agreement with earlier results for this network.

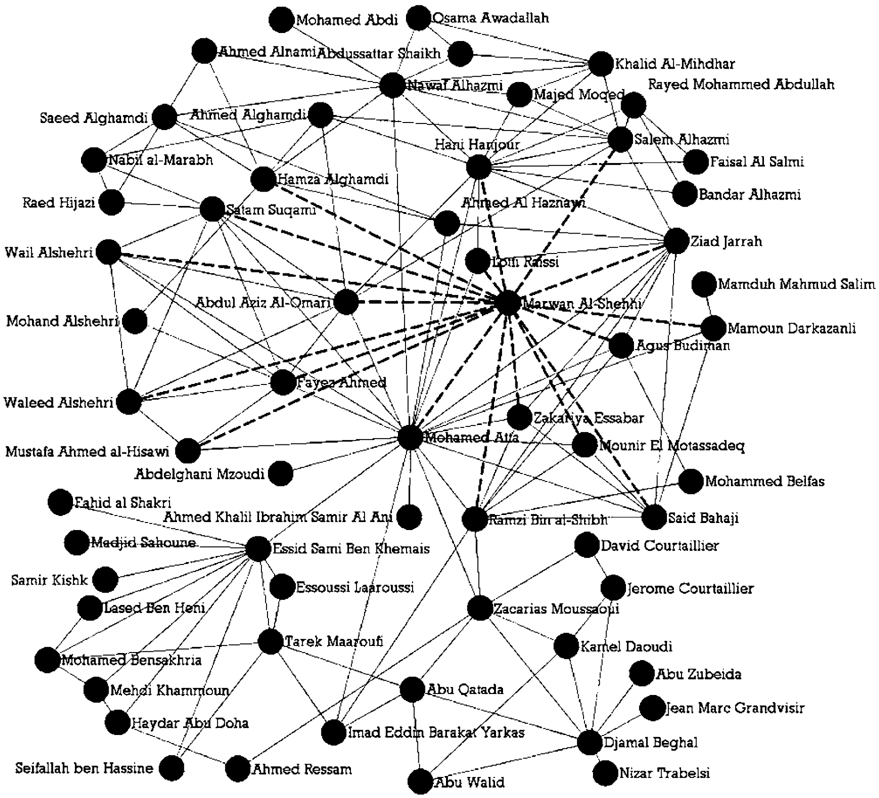


Fig. 24.13 Marwan eliminated

24.6 Conclusion

SNA has been performed by researchers in various contexts for example in Covert/Terrorist organizations to study the behavior of individual nodes. Although, typically SNA has mainly dealt with small quite bonded networks, the relationship between the nodes which are normally people is often very simple for example "friendship". However, there has been consensus among researcher that SNA can be applied to terrorist/covert networks [31,44] although they posses different structures compared with typical hierarchical organizations, they are more like cellular and distributed. The idea of applying typical SNA measures is that we can establish relationship between nodes of these networks and try to identify/isolate the important actors in the network by locating their position through centrality measures.

The study presented here is based on the similar assumption, we have considered that data is perfect that is there is no missing links etc when the sample is made. The results are provided with three different scenarios, two of these networks are well known and the information about them is easily accessible through literature.

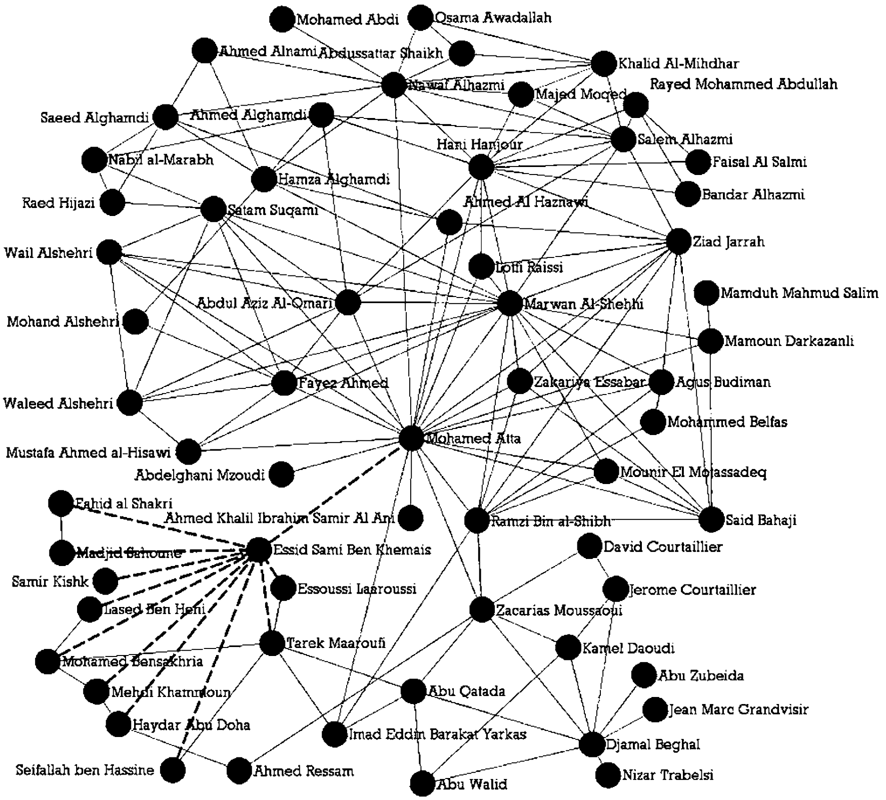


Fig. 24.14 Essid Sami Ben Khemais eliminated

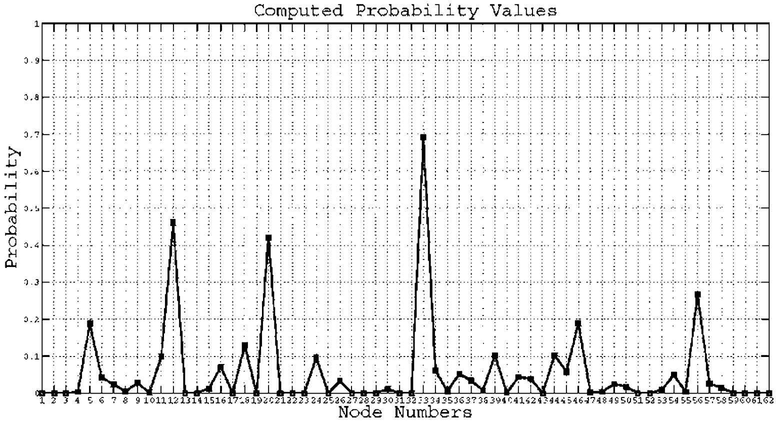


Fig. 24.15 Bayes conditional probability

The first network was intentionally created with some nodes having sort of low degree, betweenness and closeness and some having the inverse of these values. It has been shown through simulations that standard centrality measures can be combined and adopted to be used in Bayes conditional probability theorem. The results obtained through their utilization achieved a robust computational mechanism which can evaluate the likelihood probability of a node to be a key actor or otherwise in the network. The results as mentioned earlier are biased towards nodes having betweenness/closeness, which is a reasonable assumption due to two reasons one is that the most important centrality measure in SNA is betweenness and the second is the new evidence we consider is based on betweenness. Finally, the results are compared with the fragmentation index provided by Borgatti and it shows that the obtained results are consistent with his idea of distance fragmentation measure.

References

1. Borgatti S. P:. Identifying sets of key players in a network. Computational, Mathematical and Organizational Theory 12(1), pages 21–34, 2006.
2. Borgatti S. P:. The key player problem. In R. Breiger, K. Carley, and P. Pattison, (eds.), Dynamic social network modeling and analysis, workshop summary and papers. National Academy of Sciences Press, pages 241–252, 2003.
3. Borgatti S. P. and M. G. Everett:. The centrality of groups and classes. Journal of Mathematical Sociology 23(3), pages 181–201, 1999.
4. Valdis Krebs:. Connecting the dots, tracking two identified terrorists, 2002.
5. Bavelas A:. A mathematical model for group structures. Human Organization 7, pages 16–30, 1948.
6. Shaw M. E:. Group structure and the behaviour of individuals in small groups. Journal of Psychology 38, pages 139–149, 1954.
7. Bavelas A:. Communication patterns in task oriented groups. Journal of the Acoustical Society of America 22, pages 271–282, 1950.
8. Leavitt Harold J:. Some effects of communication patterns on group performance. Journal of Abnormal and Social Psychology 46, pages 38–50, 1951.
9. Smith Sidney L:. Communication pattern and the adaptability of task-oriented groups: an experimental study. Cambridge, MA, Group networks laboratory, research laboratory of electronics, Massachusetts Institute of Technology, 1950.
10. Bavelas A. and D. Barrett:. An expermental approach to organizational communication. Personnel 27, pages 366–371, 1951.
11. Glanzer M. and R. Glaser:. Techniques for the study of team structure and behaviour. Part II: Empirical studies of the effects of structure. Technical report, Pittsburgh, American Institute, 1957.
12. Glanzer M. and R. Glaser:. Techniques for the study of group structure and behaviour. Part II: Empirical studies of the effects of structure in small groups. Psychological Bulletin 58, pages l–27, 1961.
13. Cohen A. M:. Communication networks in research and training. Personnel Administration 27, pages 18–24, 1964.
14. Shaw M. E:. Communication networks. In L. Berkowitz (ed.), Advances in experimental social psychology, vol. vi, pages 111–147, New York, Academic Press, 1964.
15. Stephenson K. A. and M. Zelen:. Rethinking centrality: methods and examples. Social Networks 11, pages 1–37, 1989.

16. Flament C:. Applications of graph theory to group structure. Englewood Cliffs, NJ, Prentice Hall, 1963.
17. Burgess R. L:. Communication networks and behavioural consequences. Human Relations 22, pages 137–l59, 1968.
18. Snadowski A:. Communication network research: an examination of controversies. Human Relations 25, pages 283–306, 1972.
19. Rogers D. L:. Socio-metric analysis of inter-organizational relations: application of theory and measurement. Rural Socioeonv 39, pages 487–503, 1974.
20. Rogers D. L:. Communication networks in organizations. Communication in organizations, pages 108–148, New York, Free Press, 1976.
21. Cohn B. S. and M. Marriott:. Networks and centres of integration in indian civilization. Journal of Social Research I, pages 1–9, 1958.
22. Pitts F. R:. A graph theoretic approach to historical geography. The Professional Geographer 17, pages 15–20, 1965.
23. Latora V. and M. Marchiori:. A measure of centrality based on network efficiency, arxiv.org preprint cond-mat/0402050, 2004.
24. Freeman Linton C:. A set of measures of centrality based on betweenness. Sociometry 40, pages 35–41, 1971.
25. Freeman Linton C:. Centrality in social networks: conceptual clarification. Social Networks 1, page 215–239, 1979.
26. Anthonisse J. M:. The rush in a graph, University of Amsterdam Mathematical Centre, Amsterdam, 1971.
27. Nieminen J:. On centrality in a graph. Scandinavian Journal of Psychology 15, pages 322–336, 1974.
28. Scott J:. Social networks analysis, 2nd edition, London, Sage Publications, 2003.
29. Sabidussi G:. The centrality index of a graph. Psychometrika 31, pages 581–603, 1966.
30. Shaw M. J., C. Subramaniam, G. W. Tan and M. E. Welge:. Knowledge management and data mining for marketing. Decision Support Systems 31(1), pages 127–137, 2001.
31. Carley K. M., J.-S. Lee and D. Krackhardt:. Destabilizing networks, Dept. of Social and Decision Sciences, Carnegie Mellon University, Pittsburgh, PA 15143, November 2001.
32. Akbar Hussain D. M:. Destabilization of terrorist networks through argument driven hypothesis model. Journal of Software 2(6), pages 22–29, 2007.
33. Akbar Hussain D. M. and D. Ortiz-Arroy:. Locating key actors in social networks using bayes posterior probability framework, lecture notes in computer science. In Intelligence and Security Informatics, vol. 5376/2008.
34. Ortiz-Arroy D. and D. M. Akbar Hussain:. An information theory approach to identify sets of key players. In Intelligence and Security Informatics, vol. 5376/2008.
35. Adibi J. and J. Shetty:. Discovering important nodes through graph entropy the case of enron email database. In Linkkdd 2005: Proceedings of the 3rd international workshop on link discovery, pages 74–81, ACM, New York, 2005.
36. Freeman L. C., S. P. Borgatti and D. R. White:. Centrality in valued graphs, a measure of betweenness based on network flow. Social Networks 13, pages 141–154, 1991.
37. Newman M. E. J:. A measure of betweenness centrality based on random walks, cond-mat/0309045, 2003.
38. Beauchamp M. A:. An improved index of centrality. Behavioral Science 10, pages 161–163, 1965.
39. Oliver C. Ibe:. Fundamentals of applied probability and random processes, Elsevier Academics Press, ISBN 0-12-088508-5, 2005.
40. Montgomery D. C. and G. C. Runger:. Applied statistics abd probability for engineers, 4th edition, ISBN 978-0-471-74589-1, Wiley, 2006.
41. Koskinen J. H. and T. A. B. Snijders:. Bayesian inference for dynamic social network data. Journal of Statistical Planning and Inference 137, pages 3930–3938, 2007.

42. Siddarth K., H. Daning and C. Hsinchum:. Dynamic social network analysis of a dark network: identifying significant facilitators. In Proceedings of IEEE international conference on intelligence and security informatics, ISI 2007, New Brunswick, NJ, USA, May 23–24, 2007.
43. C. J. Rhodes and E. M. J. Keefe:. Social network topology: a bayesian approach. Journal of the Operational Research Society 58, pages 1605–1611, 2007.
44. Sparrow M.:. The application of netwrok analysis to criminal intelligence: an assessment of the prospects. Social Networks 13, pages 251–274, 1991.