IJESRT
INTERNATIONAL JOURNAL OF ENGINEERING SCIENCES & RESEARCH TECHNOLOGY

TOWARD A COMPREHENSIVE TERRORIST PREDICITION IN SOCIAL NETWORK

Ahmad F. Al Musawi
Computer Science Department, College of Computer Science and Mathematics,
Thi Qar University, Iraq

ABSTRACT

Social network analysis can play significant role in detecting human personality. Consequently, special personality characteristics can be analyzed to predict potential terrorism actions. These features could be conducted by using different social representations such as athletics, social and regional characteristics and so on. Data can be mined to result set of people with common features and categories. Based on that, people can be clustered according to their affiliation clustering on scale of dangerous and peaceful ones. Herein, a social network partitioning and clustering model is implemented to detect how close is a citizen to terrorism or a terrorist.

KEYWORDS: Social Network Analysis, Terrorists Networks, Community Detection, Clustering.

INTRODUCTION

Predicting the possible activities and patterns of interactions among *individuals* or *groups of individuals* considered as an old and emerge field as well [1], [4], [5]. However, the results, roughly speaking, of such prediction can be extensively used to manage different fields of life as it will change the ways of implementing different life facilities such as curing, employing, leading, governing, trade marketing and much sophisticated approaches and techniques of strategic implementations. This prediction can play significantly in detecting suspects and terrorists in given society. Many academic and security institution and organization have focus their works on understanding the social connection among terrorists and their cover networks so that they determine the best strategy of defensing and attacking them [2], [6]. Predicting suspected people depends on the feature of each person and the social connectivity of that person. However, the features specification is very important. A terrorist is an ordinary person who have high weighted affiliation to specific features among others. Most researchers focus their attention on analyzing terrorists networks to searching for specific role terrorists among others[20],[21].

A person, from the informational perspective, is set of attributes that describe him/ her in any given time within his/ her environment. These attributes would have two different types: static and dynamic. Static attributes are these features that hardly be changed over time which can be his name, birthdate, eye color, hair color, figure print or any biometric feature, social security number and so on. Dynamic attributes are the personal features that simply be changed over time as a result of person interaction with his/ her environment, such as current position, health state, education, psychological mode, level of awareness, level of threading and so on. The categorical measurement specification of both the static and dynamic of human attributes may differ for many scholars and purposes. Computational social science deals with determination of these measurements.

Many researchers conducted to measure different human affiliation to one activity/ behavior or another. These researches can be found in [3]. However, most scientists are focus on one-time, self- reported data on relationship [1]. The problem specification of this article is basically on how can we utilize the static information about a community for predicting terrorists within it, i.e. given social network of set of people and their attribute matrix, how could we categorize them based on their common behavioral attributes and connection? The social network depicts the relationships among people and the attribute matrix shows the different scaled measurement features of each person.

Social network analysis [7] is the study of social interaction among people using any specified social relationship. Social network analysis SNA theories and algorithms have a wide range of contribution to other similar field such as disease modeling [8], [9], biological networks [10], [11], [12], [13], business management [14], [15], [16], [17], and more.

A social network is the mathematical representation of interactions among people. The mathematical field of study used for presenting networks and its algorithms is graph theory. A graph G, consist of a pair (V, E),where V refers to set of vertices or nodes and E refers to set of edges that presented as line connects one node with another. Individuals in SNA are presented as nodes (or vertices) and the relationships among them are presented as edges connecting them. The collection of relationships is presented as edges set, where each edge is a pair of individuals or their vertices. Better presentation of edges set is by using matrix representation. Let A be an adjacency matrix of P×P values of integer number.

$$A_{ij} = \begin{cases} 1 & if\ there\ is\ a\ relationship\ between\ person\ i\ and\ j \\ 0 & otherwise \end{cases}$$

A_{ij} would represent the relationship between the person in number i with the person in number j. A is undirected network as the orient of relationship is from both direction. This network may represent any social relationship as friendship, marriage, work collaboration, joining same club and so on. It is very required to mention as much relationships as possible. The different collection of relationships can be implemented using extra arrays.

LITERATURE REVIEW

The work on detecting terrorist groups within a community based on their social "interactions" is an old field. However, Social Networks Analysis SNA is an emerging field, specifically studying properties of interactions among individuals. Social networks commonly refers to recent web sites that connect individuals with different kind of relationships as friend requests, following posts, sharing posts, making comments and so on. Social networks can refer to any social interactions other than the web, as well. Famous social networks include but not limited to Facebook, Twitter, Instagram, LinkedIn… etc., or to ordinary social knowledge on people groups. Recent scientific researchers focus their analysis on nodes' (presented by individuals) centralities and links (relationships) statistics. However, extracting the social network itself considered as a complex mission due to complexity of collecting data. This complexity comes from :

1- Large number of population which is required to be analyzed.
2- Concern of violating people privacy.
3- The covert property of terrorists groups within society.
4- If such terrorists' group information collected, governmental agencies would consider these information as cofidential data.

Moreover, these data are static in term of they are not changing or hardly changing over time. These information are mainly extracted from the web using many social applications such as UCInet [27], Pajek [28], ORA [29], and others [3], or extracted by special agreement with resources provider for trading/military purposes as in mobile phone call places, emails of group of peoples, and etc. Knowledge extraction and preparation can be achieved by analyzing text messages sent by the group [20], counting the frequent of time being communicated among different peoples, or analyzing the friend request or likes and much more. Later, extracted entities are presented by a graph or a network and the relationships as links among them.

Mainly, there are two trends on studying social networks. The first [30], [31] who focuses on clustering algorithm, e.i. identifying set of objects with similar attributes only. The second [32], [33] focuses on community detection algorithms, e.i. specifying communities based on the network topology. Recently, many researches deal with implementing network structure with the node attributes [19], an implementation of combining the community detection with the use of node attributes. However, these models [20], [21] were not dedicated to discover/ predict covert or possible terrorist individuals or networks.

MATERIALS AND METHODS

Affiliation Based Individual Clustering

Individual's Affiliation Using Attribute Matrix

An attribute matrix (see matrix (1)) consists of two sets: people set (rows) and attribute set (columns). People are presented as nodes (refered later as nodes V) such that n represents number of nodes (individuals) within the model. Attributes, with c variables where each variable represents a measurement or an attribute of any individual such as age, gender, athnecity, address, political party and so on. Attribute matrix is presented as two dimension array where rows represent people and columns represent attributes' categories, (n people $\times$ c attribute variables):

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

$$\begin{bmatrix} x_{11} & \dots & x_{1i} & \dots & x_{1c} \\ \dots & \dots & \dots & \dots & \dots \\ x_{j1} & \dots & x_{ji} & \dots & x_{jc} \\ \dots & \dots & \dots & \dots & \dots \\ x_{n1} & \dots & x_{ni} & \dots & x_{nc} \end{bmatrix} \qquad (1)$$

Every individual would be presented with a value, β. This value will present the overall measurement that reflects the state of an individual based on his/her similarity with others affiliations as well as their connection properties. These *affiliation* and *connection properties* can include suspected/ terrorist features as well as closely existence to suspected/ terrorist individuals. β is the final evaluation for each person in our model.

Every person, however, has two types of attributes: static and dynamic. The focus of this article would be on static features only. Dynamic features would in turn be the study of dynamic multilayer networks, which is out of this research's scope.

In order to measure the contribution of a given person to one category of attribute it can be simply specified by a value between 0 and 1 for each category. Each of these categories represents social factors that a person has affiliation with or described by. For example, an educated person with Ph.D. would have 1 value, a person with high school would have 0.5 value, while an illiterate would have a 0 value (based on sequence of academic degrees level). Many social and personal features can be used to describe the participant, based on prespecified study. For example:

1- Education: 0 for no education and 1 for highest degree.
2- Mobility: 0 for no mobility and 1 for full mobility.
3- Wealth: 0 for no income at all and 1 for high income.
4- Responsibility: 0 for none and 1 for high level responsibility.
5- Political power: 0 for none, 1 for joining governing parties.
6- Tribe: 0 for no relationship to tribes, 1 for a relationship to powerful tribes.

And so on.

The Model:

In order to count the final clustering value β for each person within the system, followed algorithm is implemented. However, two conditions are required initially:

1- People within the system should be presented as a social network where any two persons would have a connection (edge) if there is any social connection between them (such as friendship, marriage, working in the same place, etc.)
2- A connected database system such that it shows the characteristics of each participated person according to specified categories.

These two condition can be existed within any social networks such that many persons are members of some pages, or likes same issues and so on. Each of them has a profile within the system such that it contains his/ her information. However, governments should focus on getting more accurate details their people such that it must provide military, experience, education, political features to be included in the model other than normal information provided by social networks as in Twitter or Facebook or etc.

The algorithm below cluster individuals based to their affiliation and participation to specific communities:

1- Dividing the social network into finite number of subnetworks such that each subnetwork is specified for one affiliation/attribute categories. The existence of one person depends on his/her participation to that category. If a person does not participate within one affiliation category, then we simply ignore his/ her node and connections (edges) on that subnetwork. The result of this phase is set of subnetworks which number equals to number of affiliation categories.
2- For each subnetwork, we apply graph partitioning algorithms [3] to cluster the subnetwork into clusters or groups of people (nodes) where they internally share maximum number of connection and minimum number of connection with other groups.
3- Calculate the participation degree of each person within each cluster created in affiliation category. The main measurements that should be used here are degree, closeness and betweenness centralities.
4- Use a data mining clustering analysis algorithms to calculate the overall measure of a given person.
5- Measure the distance among nodes (two nodes) of the same category affiliation.

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

The result of above strategy will be classifying individuals according to their centrality representation within their own communities. The forth phase rank the individuals according to their similarities. Scientists can implement specific set of features (such as features of most terrorists) among others and determine the individuals with the most common features as a candidate for be a terrorist. Moreover, the model propose the distance between two nodes that belong to the same affiliation cluster. It is important to distinguish between graph partitioning (some time referred as community clusters) and data clustering, since the first belongs to graph theory field and the last belongs to data mining field. Graph partitioning is implemented in the second phase to group nodes with high topological connectivity, while data clustering is implemented in the forth phase of the model above, to group data tuples (herein, people's β)with high similarity.

Let $N = (V, E)$ be a social network, where V is set of people and E is set of edges among Vs. N would be partitioned into smaller subnetworks which number equals to number of affiliation categories: $\bigcup_{i=1}^{c} N_i = N$ Such that c refers to the number of affiliation categories. For each subnetwork N_i, number of nodes within it would be equal to $n_i = |N_i|$. The subnetwork N_i is constructed as follow: a, b ∈ V. a, b ϵ N_i if and only if a, b ϵ N and both a, b >0 and E(a, b) ϵ N. then E(a, b) ϵ N_i.

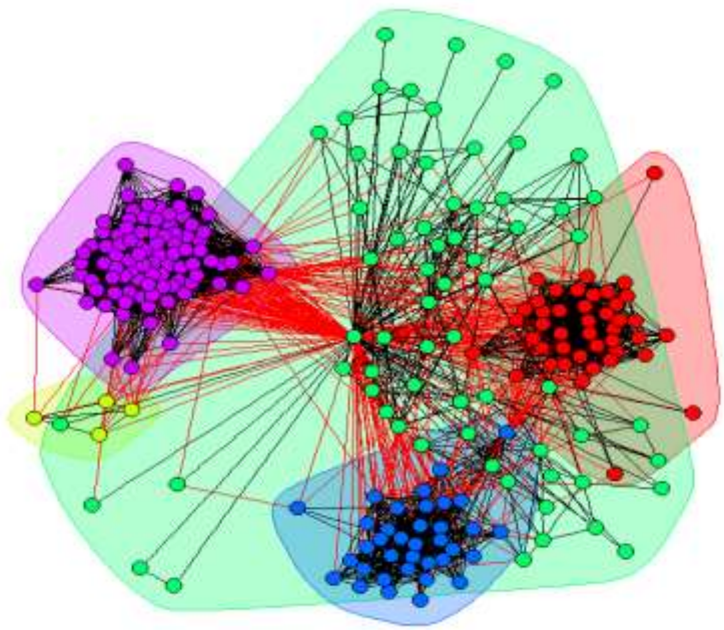

Figure (1) Sample of Communities in Social Network [18]. A Network is Clustered Using Community Partitioning Algorithm. Each Color Refers to A Group G_{ij}.

Later, for each N_i, we apply the network partitioning algorithm to divide the subnetwork N_i into smaller communities / groups such that nodes in a community are densely connected with each other and have less connected with others communities' nodes. Herein, we would describe each small connected nodes (a community) with a label for later distinguishing. The purpose of this step is to specify the densely connected nodes in groups for measuring the centralities of each node within that group. The number of groups created from partitioning varied for each subnetwork according to the structure of that subnetwork. The number of (groups or communities or cluster) should be defined such that it separates highly connected nodes from others. Let k_i refers to number of resulted groups within N_i. For each subnetwork N_i, G_{ij} would refer to group number j which created within subnetwork N_i. Let G_{i1} be the first group created within N_i, G_{i2} be the second group and so on. The processing of phase 1 is depicted in figure (1). After the partitioning process, each group is labeled so that all nodes belong to that group is simply specified.

Once we have specified the groups of people in each affiliation category, we can measure the centralities of nodes within those *groups*. The most important centralities to use are degree, closeness and betweenness centralities.

1- *The degree centrality* refers to number of edges that connect current nodes within the network. Degree centrality of a node *v* equals to $\frac{D(v)}{|V|-1}$ such that $D(v) = \sum_{x=1}^{|v|} A_{vx}$. Note that the network here is undirected.

2- *Closeness centrality* [25] refers to how close is the current node to other nodes within the network. Closeness centrality definded as the sum of distances from a current node to all others within the network. The distance can be mean geodesic distance (shortest path between two nodes), or the sum of geodesic distances. Closeness centrality of a node *v* equals to $\frac{|v|-1}{\sum_{u \in V \setminus \{u\}} d_G(v,u)}$ such that d_G is the geodesic distance.

3- *Betweenness Centrality* [26] refers to how many times does a node comes in between two other nodes. Betweenness centrality of a node *v* equals to $\sum_{s \neq v \neq t} \frac{\sigma_{st}(v)}{\sigma_{st}}$ such that σ_{st} is the total number of shortest path from node *s* to node *t* and $\sigma_{st}(v)$is the number of those paths that pass through node *v*.

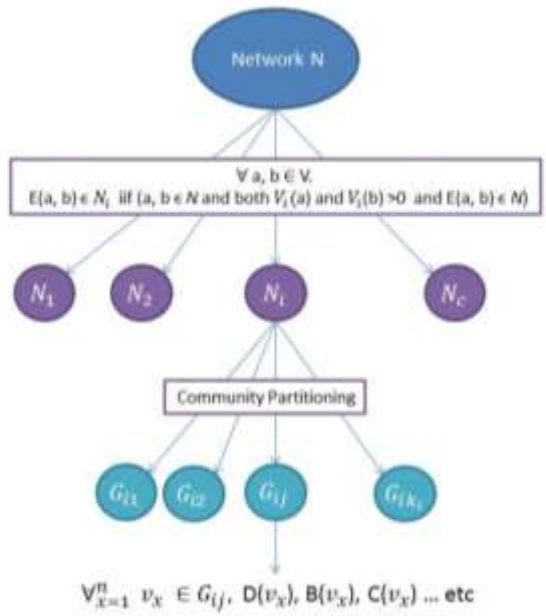

Figure (2) Network is divided into c subnetworks, then each subnetwork is partitioned, then calculats the different centralities over each community/ affiliation..

A new resulted matrix describes nodes' affiliation clustering (see matrix (2)). The matrix can be viewed in table (1). For each $v_x \in$ V, we shall calculate the degree, betweenness and closeness centralities for any affiliation does that person belong to. Later, a specific value should be associated to v_x equals to group number. Each node in network v_i ∈ V will have a *c* vector composed of four values referring to the contribution of v_x to the category of attribute *i*:

$$\begin{bmatrix} f_{11} & \cdots & f_{1i} & \cdots & f_{1c} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{j1} & \cdots & f_{ji} & \cdots & f_{jc} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ f_{n1} & \cdots & f_{ni} & \cdots & f_{nc} \end{bmatrix} \qquad (2)$$

Such that $f_{ji} = \{g_{ij}, D(v_j), B(v_j), C(v_j)\}$, where g_{ij} is the group number in attribute category *i*, $D(v_j), B(v_j), C(v_j)$ refers to the degree, betweenness and closeness centrality respectively.

Data Clustering Analysis and Terrorists Detection:

The node affiliation matrix is process to cluster the nodes into k clusters. k is specified by the user, k represents number of people's categories s.t. 1 ≤ *k* ≤ *n*, *n*: number of nodes within network N. For example, if k=10 then we are partitioning the people to 10 different clusters where each cluster contains peoples with highly similar features. Clustering can be achieved by using data mining partitioning algorithms that divide the data into k partitions such as k- means algorithm [22], [23] and k – medoids algorithm [24]. Notice that each cluster must contain at least one node and each node must belong to one cluster only. β_{v_x} would be equal to the cluster of node v_x.

This model can be implemented over diverse and sophisticated attrubited community. However, most researches on terrorist networks consider only terrorists network, which means terrorists are already known for authorities and the model. The scope of this paper is to point out suspected people who share the most common features, within a dataset contains terrorists and ordinary people. In one hand, authorities have to specify determined features for clustering and further analysis. In another hand, however, the final data partitioning (clustering) can be oriented to another direction with the existence of terrorsits attributes such that another data mining approach can be achieved such as classification and futher predicting.

In general, by having the cluster type of each node, analysis can go further and measure the shortest distance [3] from current node to other nodes that share the same clustering values. The final $\bar{A}$ $(n \times n)$ matrix would reflect the possible next suspected individuals who share the same attribiutes. $\bar{A}_{ij} = shortest\ path(v_i, v_j)$ if and only if $\beta_{v_i} = \beta_{v_j}$.

CONCLUSION AND FURTHER STUDIES

In this research, a mathematical model is presented to detect the suspected people who share the same features with others. Authorities has to set out the common features required to analyze given set of people. The model will show the shortest distance from one node to another same affiliated node as a measurement of possiblility to contribute to the same actions. The disadvantages of this research is the absence of empirical analysis due to absence of complete dataset. However, this is a novel model as it provides an appropriate methodology to estimate the future terrorist within given community.

More studies can be achieved on predicting the future terrorist based on behavioral (dynamic) attributes of known terrorist as well as implementing heuristical analysis to measure the contribution of (nodes on path) to the relationship between a terrorist and his/ her recruited future terrorist.

REFERENCES

[1] D. Lazer, A. Pentland, L. Adamic, S. Aral , A. L. Barabási, D. Brewer, N. Christakis, N. Contractor, J. Fowler, M. Gutmann, T. Jebara, G. King, M. Macy, D. Roy, M. V. Alstyne, "Social Computational Science", Science, Vol 323, page 721 – 723, Feb 6th, 2009.
[2] Moreno, J. L., "Who Shall Survive?", Beacon House, Beacon, NY, 1934.
[3] M. E. J. Newman, "Networks: an Introduction", Oxford University Press, New York, 2010.
[4] Linton Freeman, "The Development of Social Network Analysis: A study in the Sociology of Science", Vancouver, Empirical Press, 2004.
[5] Stanley Milgram, "The Small World Problem", Psychology Today, May 1967:60 67.
[6] Steve Ressler, "Social Network Analysis as an Approach to Combat Terrorism: Past, Present, and Further research", Homeland Security Affairs, Vol. II, No. 2, July 2006.
[7] Albert-Lazio Barabasi, " Linked: The New Science of Networks", New York, Preseus, 2002.
[8] S. Eubank, H. Guclu, V. Kumar, M. Marathe, A. Srinivasan , Z. Toroczkai, and N. Wang. Modelling disease outbreaks in realistic urban social networks. Nature , 429:429:180–184., Nov 2004. Supplement material.
[9] M. Kretzschmar and M. Morris. Measures of concurrency in networks and the spread of infectious disease. Math. Biosci., 133:165–195, 1996.
[10] J. J. Faith, B. Hayete, J. T. Thaden, I. Mogno, J. Wierzbowski, G. Cottarel, S. Kasif, J. J. Collins, and T. S. Gardner, “Large-scale mapping and validation of escherichia coli transcriptional regulation from a compendium of expression profiles,” PLoS biology, vol. 5, no. 1, p. e8, 2007.
[11] A. A. Margolin, I. Nemenman, K. Basso, C. Wiggins, G. Stolovitzky, R. D. Favera, and A. Califano, “Aracne: an algorithm for the reconstruction of gene regulatory networks in a mammalian cellular context,” BMC bioinformatics, vol. 7, no. Suppl 1, p. S7, 2006.
[12] Pellegrini Matteo, Haynor David, Johnson JM: "Protein interaction network" , Expert Rev Proteomics 2004, 1(2).
[13] Jeong, H., B. Tombor, R. Albert, Z. N. Oltvai and A. –L. Barabasi: "The large scale organization of metabolic networks", 2000, Nature (London) 407, 651.
[14] S. Bernstein, A. Clearwater, S. Hill, C. Perlich, and F. Provost. "Discovering knowledge from relational data extracted from business news". Proc. Workshop on Multi-Relational Data Mining, 2002

ISSN: 2277-9655
(I2OR), Publication Impact Factor: 3.785

[15] K. Carley and M. Prietula, eds. "Computational Organization Theory" . Lawrence Erlbaum ass., Hillsdale, NJ, 2001.
[16] C. Papadimitriou. "Computational aspects of organization theory". Lecture Notes in Computer Science , 1997.
[17] C. Papadimitriou and E. Servan-Schreiber. "The origins of the deadline: Optimizing communication in organizations". Complexity in Economics. , 1999.
[18] Francesco Pochetti, "Community Detection in Social Networks", "http://francescopochetti.com/community-detection-social-networks/".
[19] J. Yang, J. McAuley, J. Leskovec. "Community Detection in Networks with Node Attributes". IEEE International Conference on Data Mining (ICDM), 2013.
[20] Carley, Kathleen M., "Dynamic Network Analysis for Counter - Terrorism". Unpublished manuscript.
[21] Ala Berzinji, Lisa Kaati, Ahmed Rezine, "Detecting Key Players in Terrorist Networks", *EISIC*, 2012, European Intelligence and Security Informatics Conference, European Intelligence and Security Informatics Conference 2012, pp. 297-302, doi:10.1109/EISIC.2012.13
[22] S. P. Lloyd. "Least Squares Quantization in PCM". IEEE Trans. Information Theory, 28:128– 137, 1982, (original version: Technical Report, Bell Labs, 1957.)
[23] J.MacQueen. "Some methods for classification and analysis ofmultivariate observations". Proc. 5th Berkeley Symp. Math. Statist. Prob., 1:281–297, 1967.
[24] L. Kaufman and P. J. Rousseeuw. "Finding Groups in Data: An Introduction to Cluster Analysis". JohnWiley & Sons, 1990.
[25] Gert, S., "The centrality index of a graph" . Psychometrika 31(4) , December 1966 , 581- 603.
[26] Everett, M., Borgatti, S.P., "Ego network betweenness" . Social Networks 27(1), January 2005, 31-38.
[27] (http:/ / www. analytictech. com/ ucinet/ ucinet. htm).
[28] (http:/ / pajek. imfm. si/ doku. php).
[29] (http:/ / www. casos. cs. cmu. edu/ projects/ ora/)
[30] D. M. Blei, A. Ng, and M. Jordan. "Latent dirichlet allocation". JMLR, 3:993–1022, 2003.
[31] S. Johnson. "Hierarchical clustering schemes". Psychometrika, 1967.
[32] S. Fortunato. "Community detection in graphs". Physics Reports, 2010.
[33] J. Xie, S. Kelley, and B. K. Szymanski. "Overlapping community detection in networks: the state of the art and comparative study". ACM Computing Surveys, 2013.

AUTHOR BIBLIOGRAPHY

Ahmad F. Al Musawi
Faculty member in CS Department, Thi Qar University. M.Sc. in CS from School of Engineering, Virginia Commonwealth University, USA. B. Sc. in CS from College of Science, Thi Qar University, Iraq.