Crime Detection and Prevention using Social Network Analysis

Shikhar Gupta
Department of Computer Science &
Engineering
KIET Group of Institutes,
Ghaziabad, India

Sanjeev Kumar
Department of Computer Science &
Engineering
Faculty of Engineering & Technology
KIET Group of Institutes,
Ghaziabad, India

ABSTRACT

A society is built of individuals and group of individuals make organization. These individuals or organizations are also called as nodes. A structure that consist these nodes and relation between these nodes is known as social networking. We all are surrounded by networks and plays an important role of individual unit in a network of different kinds of social relationships, biological systems. In current scenation social networking and blogs are the most popular kind of online activities. Usage of social networks are much more than personal emails. Facebook, Twitter & LinkedIn are some of the well-known examples of Social networking. It is the analysis of how social groups communicate & connect to each other & affect important features. Social network analysis is not just analysis of an individual, but it is the study of group of individual and relation between them. Social Network Analysis is current emerging area of importance in finance, politics, defense and security. Prediction of missing links & links that can occur in future between notes in social network is an attention holding topic. It is of interest that Fuzzy system analysis can make more significant & correct predictions.

There are many methods to depict knowledge in the field of soft computing. But the most common way to portray the human understanding is with the help of natural language expression also called fuzzy rule-based system. With the help of this rule based form if we aware of information, then we can drive another information called as conslusion. Fuzzy system is used to identify the trait of an individual. We take five different characteristics like economical status, family background, educational level, alcoholic &drug addict and criminal history. All of these Characteristics of an individual are mapped with the help of fuzzification & de-fuzzification techniques to obtain a de-fuzzification value. These values help to identify the criminal phycology of a person. Each value is then assigned a colour. Colours are in given order starting from highest criminal level to lowest: Dark Red, Red, Light Red, Orange, Yellow, White, and Green. When all the node gets its appropriate colour, we analyse the interaction between the nodes and reassign a colour to a node if its activities in not appropriate according to the characteristic values provided by the user or if a person is interaction with someone who is already placed in some other category. For this re-analysis we again use fuzzy system techniques.

Keywords
Social network analysis, Crime detection, Fuzzy rules, Network models, Synchronization process.

1. INTRODUCTION
Social Network is a network of collective interactions and personal associations. It's a social structure, which constitutes a set of social actors (nodes) and connection between these actors (Links). We all are surrounded by networks and we ourselves plays an important role of individual units in a network of different kinds of social relationships, biological systems [A. L. Barabasi et al.(2011)]. Networks can be an actual items in Euclidean space, for example an electrical power grid, internet, roads or subways and neural networks [S. Boccaletti et al.(2005)]. Social network analysis is a method to analyse the links between nodes [Wasserman et al. (1994)]. Social Network Analysis is an current emerging area of importance in finance, politics, defense and security.

In last few years, it has been seen that researchers gain interest in the study of complex networks. A complex network is a network having irregular structures, complex and dynamically changing network with time. The main focus of studying the complex network is to upgrade the knowledge of analysing of minor networks to a system of large networks with thousands or millions of nodes. Generally network study comes under graph theory [M. van Steen (2010)]. As in graph theory different types of problems such as maximum slop problem, graph colouring problem, can be solved easily and like the same this type of principle will be used in social network study. In social network study we analyse the relationship among the nodes and study their dynamics of behaviour and structural changes and its effects.

2. EVOLUTION OF SOCIAL NETWORK ANALYSIS
The evaluation in the field of complex networks is activated by two seminal papers, written by Watts and Strogatz on small world networks, appeared in Nature in 1998 and that by Barabasi and Albert on scale free networks. These studies include transportation networks, phone call networks, the internet and World Wide Web, and neural or genetic networks. In the late 18th century, Emile Durkheim gave his idea of social networking in his research. Later in 19th century, development in the field of Social network analysis reached to another level. In 20th century, Georg Simmel was one of the first researchers who thought in relatively explicit social network terms. He observed in what way third parties could disturb the connection between two entities—and he inspected how organizational assemblies or bureaucracies were needed to synchronize interactions in large groups.

There were many issues faced in the analysis of complex networks like the structures of these networks. Since the relative examination of networks taken from diverse fields generates a string of vivid outcomes. The research starts with the investigation of new ideas and describe the properties of real networks. The main task of research is to identify the unified principal and properties common in most of the real networks considered. An important property which is to be considered is the degree of a node that is the number of links a node have with other nodes directly. In real networks, the degree distribution P(k), is expressed as the possibility of choosing a vertex consistently at random has degree k. In general, a real network is characterized by finding the relation between the degree of nodes, by taking comparatively short tracks among any two nodes. Observations of these variables provides a revival of network modelling, since the concept of mathematical graph theory are different from the real needs. Scientists have to continuous searching for the new models to simulate the growth of a network and to propagate the structural properties perceive in real topologies. The topology of a actual network can be understand by the regular study of the forces, which forms it, and definitely affect the purpose of the system. Thus, this platform of research was motivated by the eagerness that understanding and demonstrating the structure of complex network would lead to a improved understanding of dynamical and functional behaviour, evolutionary mechanism. As expected it has been found that mixing architectural properties has noticeable significances over the functional properties of the network, its robustness and reactions to outward effects like random failure and targeted attacks.

Simultaneously, it appear as an outcrop for the first time the possibility of studying the dynamic behaviour of large assemblies of dynamical systems interacting via complex topologies, as the one observed empirically. This proves that the network topologies plays a crucial role in determining the emergence of collective dynamical behaviour like synchronization or in performing the main functions that takes place in complex networks, such as spreading of information and rumours.

3. ROLE OF SOCIAL NETWORK ANALYSIS IN CRIME DETECTION AND PREVENTION

Social Network Analysis getting more attention in last several years. It provides the knowledge about various network structures and their characteristics. There are various activities involved in crime like traffic violation, organizational frauds, kidnapping, murders, etc. The major challenge in front of authorities is how to effectively analyse the large amount of criminal data.

With the help of social network analysis we can investigate the interaction between a group of nodes. We can find out the closeness between the members of a group. Social Network Analysis provides a way to watch the interaction between those types of peoples, which are already involved in some kind of criminal cases. As we know that there is a leader in ever gang. We can a ranking list of people's ability of leadership. The leaders in the criminal are sure to come to surface.

4. TECHNIQUES OF SOCIAL NETWORK ANALYSIS

Social Network is random complex network, which has dynamic structure and random behaviour (functions). Only structure of Network is not sufficient for analysing the Social Network, while functions are also important. So the collections of functions and structural attributes are used to analyse the social network correctly. Social network is similar to neuronal system. So Social Network can be analysed as neuronal system analysed by using network topologies and dynamic behaviour. But the challenge is to synchronize the topologies and behaviour of social network, so there is a need to study of synchronization process.

4.1. Synchronization Process

Synchronization process [Yamir Moreno at el. (2009)] provides the technique of coupling the network structure and functions. Like in Neuronal system, synchronization is applied for coupling billions of neurons connected with each other and makes a complete network. Social Networking is also a complex network like neuronal system i.e. social network has complex structure, sharing small world and scale free features. Neuronal System is composed by two types of neurons, excitatory principal cells and inhibitory interneurons. Synchronization system used to understand the relation between complex network structure i.e. link structure and dynamic behavioral properties of complex networks. After studying many researchers work it can be calculated that synchronization is highly dependent on degree of the nodes in the network and independent network size [C. Zhou et al. (2006)].

4.1.1. Synchronization of Kuramoto Oscillators

Kuramoto Oscillator is the simplest and most efficient model for synchronization of complex network like social network. For applying this model, it assumes that each network component is an oscillator and the information exchange between them follows the rule of Kuramoto Model [Y. Kuramoto et al. (1975)].

$$\frac{d\theta_i}{dt} = \omega_i + \sum_j \sigma_{ij} a_{ij} \sin(\theta_j - \theta_i) \ (i = 1, \ldots, N).$$

$$r_{link} = \frac{1}{2N_l} \sum_i \sum_{j \in \Gamma_i} \left| \lim_{\Delta t \to \infty} \frac{1}{\Delta t} \int_{t_r}^{t_r + \Delta t} e^{i[\theta_i(t) - \theta_j(t)]} dt \right|$$

4.2. Network Models

4.2.1. Erdos-Renyi Random Graph Model

This Model is used for creating arbitrary graphs where links are fixed between vertices with equal probability. Two parameters are used to generate this model [R. Cohen et al. (2003)].

N-number of vertices in the network generated. p-probability that a link is created among any two vertices. Now we can derive a constant k with these parameters $k = \frac{2E}{N} = p(N - 1)$. Since there is no bias for a particular node in this model, the degree distribution is binomial

$$P(\deg(v) = d) = \binom{n-1}{d} p^k (1 - p)^{n-1-k}$$

4.2.2. Watts-Strogatz Small World Model

This model generates the graph with small world property. Each node is initially linked to its <k> closest neighbours [D.J. watts et al. (1998)]. Every link has a possibility p that it will be rewired to the graph as random edges. Estimated

number of rewired links is pE = pN < k >/2.This model have very high clustering coefficient [D.J. watts (1999)]].

4.2.3. Barabasi-Albert (BA) Preferential Attachment Model

This model shows "rich get richer" effect [A.L. Barabasi et al. (1999)]. A link is very probable to attach to vertiex with higher degrees. In BA model, a new vertex is attached to the network, one at a time. Each new vertex is linked to m existing vertices with possibility $P_i = \frac{k_i}{\sum_j k_j}$ where k_i is the degree of node i.

5. METHODOLOGY USED TO IMPLEMENT OUR MODEL

Zadeh introduced fuzzy set concept in 1965 to signify data and knowledge containing non-statistical uncertainties [T. J. Ross (2010)]. It was particularly intended to characterize uncertainties and imprecision mathematically, and provides standard methods for dealing with the imprecision intrinsic to many problems.

Fuzzy methods are appropriate for changeable or estimated reasoning, mainly for the systems having mathematical model that is difficult to drive.

Fuzzy Logic permits us to give an approximate value under incomplete or uncertain value.

The major change among classical set theory and fuzzy set theory is that in classical set theory a crisp, sharp and clear difference exist among member and a non member for some well expressed set of this theory while there is no well define boundaries exist between member of fuzzy set. In other words, in classical set when one asks the question "Is this entity a member of that set?" The response is either "yes" or "no.". This is correct for both the deterministic and the stochastic cases. In fussy set one asks the probability of an element to belong to a particular set. In probability and statistics, one may ask a question like "What is the probability of this entity being a member of that set?" In this case, although an answer could be like "The probability for this entity to be a member of that set is 90%," the final outcome (i.e., conclusion) is still either "it is" or "it is not" a member of the set. Zero is used to signify complete non-membership, the value one is used to represent complete membership, and values among zero and one is used to represent the level of membership in the set.

5.1. Fuzzy (rule-based) system

There are many methods to signify knowledge in the field of artificial intelligence. However, the best possible method representing the human knowledge is with the help of natural language expression of the type [S. Rajasekaran et al. (2009)]

IF hypothesis (ancestor), THEN inference (resultant).

This procedure is usually known as IF-THEN rule based method; this method is generally stated as inferential method. It states an inference such that if we know a detail (hypothesis, ancestor) then we can conclude, or drive, another element called as inference (resultant). This procedure is also known as shallow knowledge because it conveys human observed and heuristic knowledge in our respective language of communication.

The structure of a fuzzy rule based system	
Rule 1:	If antecedent C1, then
Rule 2:	conclusion R1
.	If antecedent C2, then
.	conclusion R2
.	
Rule p:	If antecedent Cr, then
	conclusion Rr

We can decompose any compound statement with the help of basic properties and operations of fuzzy set and reduce to simple canonical rules.

6. IMPLEMENTATION OF PROPOSED MODEL

We first apply Fuzzy System techniques on our input data set and assign a colour to every node (person). Each colour defines about the criminal characteristics of a person. Then we again apply the Fuzzy System techniques over the nodes to determine the frequency of interaction between the nodes. We take five different characteristics, which are used to determine weather, a person posse's criminality or not. The characteristics are

- Economical Status: Here a person has to provide details about his/ her annual income.
- Family Background: Here a person has to provide details about his annual family income.
- Educational Level: In this section, a person shout declare his level of education
- Alcoholic and Drug Addict: Here a person has to tell about his addiction to drug, like he is heavily addicted, occasional or daily.
- Criminal History: In this section, a person has to declare about his criminal history, if he has a criminal background then up to which level.

Then we map all these five input data in fuzzy system, with the help of fuzzification and defuzzification techniques and a crisp defuzzified value, which is used to determine that weather a person can harm society in future or not. And if the person is harmful, then the level of criminal phycology is determined and we treat the mobile phone number of that person as a node and on the level of criminality he posses we assign a colour to that node. Then in future if a crime took place in a particular location, we analyse our cell phone data to see which are those nodes present at the location of crime and analyse the interaction between the nodes.

6.1. Fuzzy Rules (Canonical form)

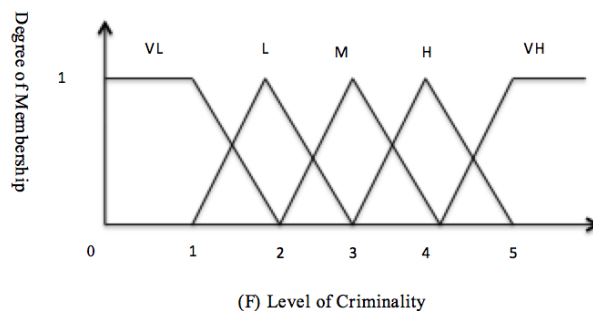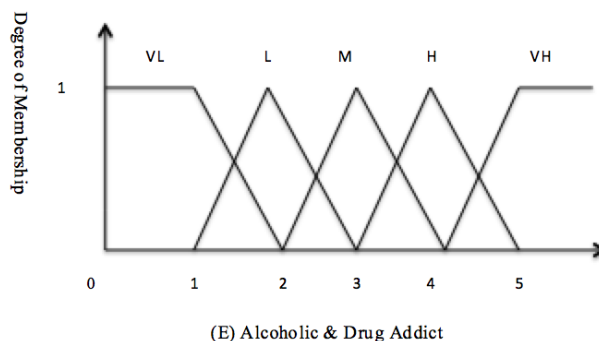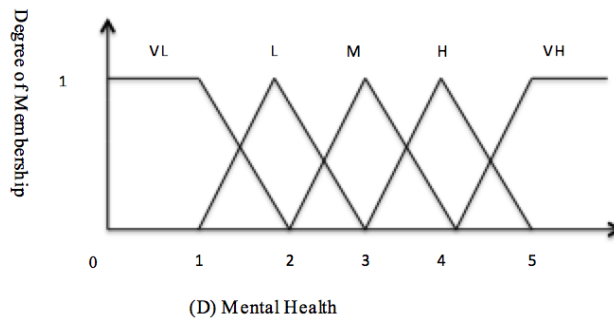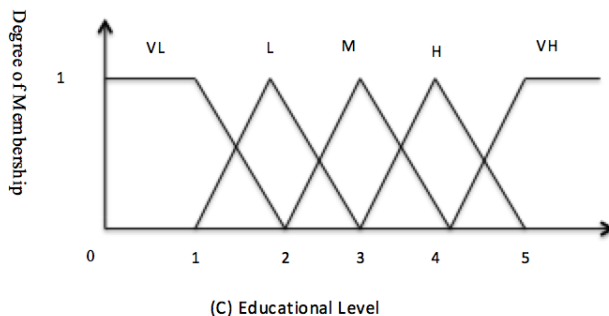A fuzzy rule base R governing the crime detection system are given below:

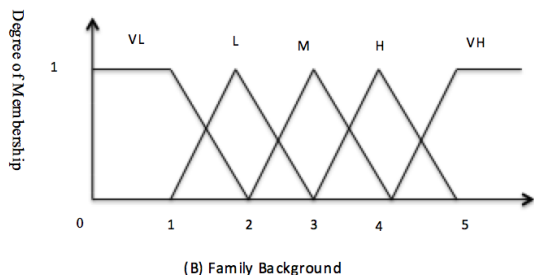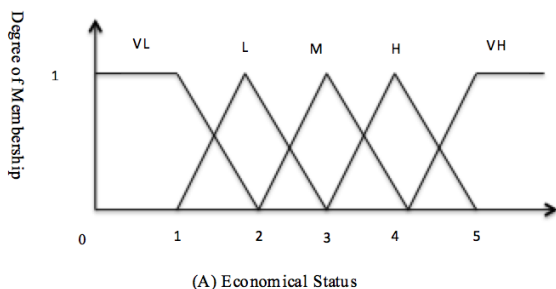| Rule 1:	If economic status is VL and family background is VL then chances of a person to be a criminal is VH.
Rule 2:	If economic status is VL and family background is L then chances of a person to be a criminal is VH.
Rule 3:	If economic status is VL and family background is M then chances of a person to be a criminal is H.
Rule 4:	If economic status is VL and family background is H then chances of a person to be a criminal is M.

Rule 5:	If economic status is VL and family background is H then chances of a person to be a criminal is L.

.
.
.

Rule 91:	If economic status is VH and drug addiction is VL then chances of a person to be a criminal is VL.
Rule 92:	If economic status is VH and drug addiction is L then chances of a person to be a criminal is VL.
Rule 93:	If economic status is VH and drug addiction is M then chances of a person to be a criminal is L.
Rule 94:	If economic status is VH and drug addiction is H then chances of a person to be a criminal is L.
Rule 95:	If economic status is VH and drug addiction is VH then chances of a person to be a criminal is M.

Here, in these rules the keys used are
VL- Very Low
L- Low
M- Medium
H- High
VH- Very High

6.2. Graphical representation of fuzzy rules (as described above)

The fuzzy sets, which characterize the input and output, are given below

(A) Economical Status

(B) Family Background

(C) Educational Level

(D) Mental Health

(E) Alcoholic & Drug Addict

(F) Level of Criminality

6.3. Colours Used

After getting the defuzzified values, we have to assign a specific colour to each node. Which shows the level of criminality possessed by that node. For this purpose we uses a variety of colours like Dark Red, Red, Light Red, Orange, Yellow, White, Green.

Dark Red Red   Light Red Orange Yellow   White Green

Highest   CRIMINALITY STSTUS   Lowest
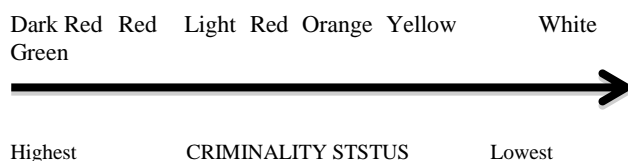
6.4. Interaction Between the Nodes

Now we again implement the Fuzzy System techniques called fuzzification and defuzzificztion over the output we got from the above technique and determine the interaction between the nodes on the basis of level of interaction between the nodes and the sensitivity of the area where the nodes are present. Then, if any node with lower criminal level is interacting frequently with some node which posses more criminality in a sensitive geographical location then we again change the colour of lower criminal level node accordingly.

Levels of interaction between any two nodes have been calculated on the basis of number of call made between the two nodes and the location where they are present.

6.5. Fuzzy Rules base governing the crime detection (Canonical form)

A fuzzy rule base R governing the crime detection system are given below:

1. If number of calls is VL and sensitivity of area is VL then chances of a person to be a criminal is M.
2. If number of calls is L and sensitivity of area is L then chances of a person to be a criminal is M.
3. If number of calls is M and sensitivity of area is L then chances of a person to be a criminal is H.
4. If number of calls is M and sensitivity of area is M then chances of a person to be a criminal is M.
5. If number of calls is M and sensitivity of area is H then chances of a person to be a criminal is H.
6. If number of calls is M and sensitivity of area is VH then chances of a person to be a criminal is VH.
7. If number of calls is H and sensitivity of area is L then chances of a person to be a criminal is H.
8. If number of calls is H and sensitivity of area is M then chances of a person to be a criminal is H.
9. If number of calls is H and sensitivity of area is H then chances of a person to be a criminal is VH.
10. If number of calls is H and sensitivity of area is VH then chances of a person to be a criminal is VH.
11. If number of calls is VH and sensitivity of area is L then chances of a person to be a criminal is VH.
12. If number of calls is VH and sensitivity of area is M then chances of a person to be a criminal is H.
13. If number of calls is VH and sensitivity of area is H then chances of a person to be a criminal is VH.
14. *If number of calls is VH and sensitivity of area is VH then chances of a person to be a criminal is VH.*

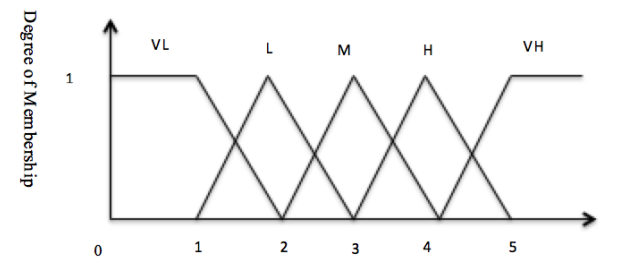## 6.6. Graphical Representation of rule base governing the crime detection

The fuzzy sets, which characterizes input and output in this section are:
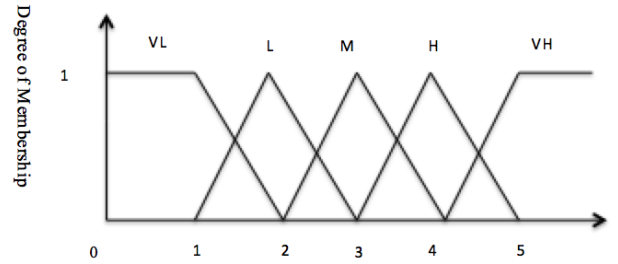
(G) Number of Phone Calls
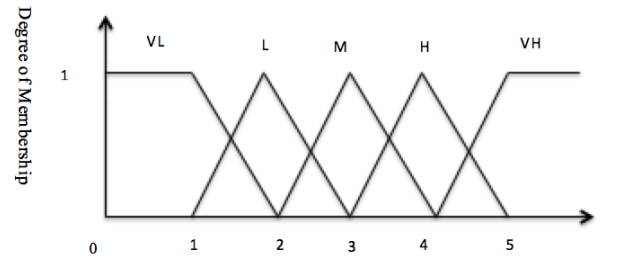
(H) Sensitivity of Geographical Location

(I) Final Output

7. RESULT

TABLE 7.1: Color assigned to a node

Phone Number	UID Key	First Parameter	Value Assigned	Second Parameter	Value Assigned	Color
88xxxx xx23	Sh883	Economical Status	0.0	Mental Health	0.0	Red
77xxxx xx92	Vi777	Economical Status	4.1	Mental Health	4.5	Light Red
98xxxx xx52	An989	Economical Status	3.2	Family Background	3.8	Orange
27xxxx xx34	Ab273	Economical Status	4.5	Educational Level	4.5	Orange
19xxxx xx34	Sh198	Economical Status	4.5	Family Background	4.0	White

TABLE 7.2: Re-calculation of color based on the interaction between the nodes

Color of 1st Node	Color of 2nd Node	Number of Phone calls	Sensitivity of Location	Color assigned to 1st Node	Color assigned to 2nd Node
Dark Red	Green	4.5	0.1	Dark Red	Green
Green	Yellow	2.0	1.0	Green	Green
White	White	2.1	0.0	White	White
Orange	White	3.1	1.0	Orange	Orange
White	Green	4.1	0.0	Green	Green

8. CONCLUSION

We have proposed techniques of future, where we will be assumed presence of nodes, which keeps all essential data of its citizens on most of things are mapped that is Adhar card with account, health, education and unique mobile number too. As present government is very eager to map individuals with Adharcard, even those are leading to the pure e-governance. So we assume in future we'll have one existing system where every individual will be mapped with this Adharcard. There will be one communication point like one India, one contact number and identity would also be reviled with biometric parameters that is given to that person.

Every person in our application is treated as a node without revelling the privacy of that person. We are treating a particular node with a key number. All private information are linked with the national agencies. This application will help our national agencies to present future events that are possible events as well as some other unpleasant situations.

Every person (node) has its attributes like level of education, economical status, mental health and these attributes are measured on different scales. We've created fuzzy rules to formulate possible events as well as to assign a colour to a node, representing a level of attributes.

Colour shows the level of attributes possessed by that node. For this purpose, here we have used different colours like dark red, red, light red, orange, yellow, white and green. Level of criminality is highest for dark red & white for non-criminality.

After colours are assigned to all nodes, mapping is done between these nodes on the basis of two factors. First, the number of phone calls between any two nodes and second level of sensitivity of the geographical location where the nodes visiting. On the basis of above two factors, we again implement fuzzy techniques to map possible events.

This bonding may re-determine the colour of the node of low criminality between the two compared nodes. Meaning, level of Criminality is identified again based on the bonding between two nodes.

All the Attribute values & colours assigned to nodes are being saved in an excel file. This file can be used to retrieve data for further analysis.

This is multivariable dynamic problem so attributes and other parameters would get changed with time, space and other dependent variables.

In future, as we would include nondeterministic multivariable parameters so stochastic modelling and analysis of problem would be required to deal with uncertainty among multidimensional attributes.

9. REFERENCES

[1] A.L. Barabasi, N. Gulbahce, J. Loscalzo, "Network medicine: a network-based approach to human disease", Nature Reviews Genetics, 12(1), 56–68, 2011.

[2] A.L. Barabasi, R. Albert, "Emergence of Scaling in Random Networks", Science 286 (1999) 509, 1999.

[3] C. Zhou, A. E. Motter, and J. Kurths, "University in the Synchronization of Weighted Random Networks". Phys. Rev. Lett. 96, 034101(2006).

[4] D.J. Watts, S.H. Strogatz, Nature 393 (1998) 440.

[5] D.J. Watts, Small Worlds: The Dynamics of Networks between Order and Randomness, Princeton University Press, Princeton, NJ, 1999.

[6] Maarten van Steen, "An Introduction to Graph Theory and Complex Network", 2010.

[7] R. Cohen, S. Halvin, "Scall-Free networks are ultrasmall". Phys. Rev. Lett90 (5): 058701, 2003.

[8] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, D.U. Hwang, "Complex networks: Structure and dynamics", journal of ELSEVIER, 2005.

[9] S. Rajasekaran, G. A. Vijayalakshmi Pai, 2009. "Neural Network, Fuzzy Logic, and Genetic Algorithms Synthesis and Applications". PHI Learning private Ltd. , Twelfth Edition.

[10] T.J. Ross, "Fuzzy Logic with Engineering Applications", WILEY publication, Third edition, 2010.

[11] Wasserman, Stanley and Katherine Faust, "Social Network Analysis: Methods and Applications", Cambridge University Press, 1994.

[12] Y. Kuramoto, "Self-entrainment of a population of coupled nonlinear oscillators", Lect. Notes Phys, 30, 420 (1975).

[13] Yamir Moreno, "Complex Network Modeling: A New Approach to Neurosciences"Springer Series in Computational Neuroscience Volume 2, 2009, pp 241-263.