International Journal of Computer and Communication Technology

Volume 7 | Issue 4 Article 17

October 2016

SOCIAL NETWORK ANALYSIS FOR CHURN PREDICTION IN TELECOM DATA

PUSHPA PUSHPA
Computer Science and Engineering Department, R.V. College of Engineering, Bangalore-560059, India, flowersunpr@yahoo.co.in

Dr. Shobha G
Professor, RVCE, Bangalore, India, shobhag@rvce.edu.in

Recommended Citation
PUSHPA, PUSHPA and G, Dr. Shobha (2016) "SOCIAL NETWORK ANALYSIS FOR CHURN PREDICTION IN TELECOM DATA," *International Journal of Computer and Communication Technology*: Vol. 7 : Iss. 4 , Article 17.
DOI: 10.47893/IJCCT.2016.1389
Available at: https://www.interscience.in/ijcct/vol7/iss4/17

SOCIAL NETWORK ANALYSIS FOR CHURN PREDICTION IN TELECOM DATA

PUSHPA [1]& G SHOBHA[2]

[1,2]Computer Science and Engineering Department, R.V. College of Engineering, Bangalore-560059, India
Email;flowersunpr@yahoo.co.in, shobhag@rvce.edu.in

Abstract— Social Network Analysis (SNA) is a set of research procedures for identifying group of people who share common structures in systems based on the relations among actors. Grounded in graph and system theories, this approach has proven to be powerful measures for studying networks in various industries like Telecommunication, banking, physics and social world, including on the web. Since Telecommunication industries deals with huge amount of data, manual analysis of data is very difficult. In this paper we explore the Social Network Analysis techniques for Churn Prediction in Telecom data. Typical work on social network analysis includes the construction of multi-relational telecom social network and centrality measures for prediction of churners in telecom social network.

Keywords*-Social Network Analysis (SNA), Churn Prediction, Centrality Measures.*

I. INTRODUCTION

A social network is a group of people who are connected through their use of communication services. The effectiveness of the social network analysis is concerned with the analysis of the influence of members in a community within a social network on a product purchase and service usage [1]. Mobile Social Network analysis (MSNA) is an upcoming and interesting area of concern in telecom industries since it not only helps in exploring the information regarding the social network of subscribers but also helps the operators' to focus on their business analytics. MSNA is being used to give a solution to some of the telecom problems such as to improve churn prediction, expedite the campaigns for up-selling and cross-selling activities, and overall for customer satisfaction and retention [2].The structure of customer communication networks provides a natural way to understand customers' relationships and to a certain extent the behavior of groups of highly connected customers. There are many related works which have been presented in Social Network Analysis (SNA) like finding of communities, alpha users within the communities, hub member and non-hub members' identification [3, 4].

From the data mining standpoint, a social network is a heterogeneous and multi-relational data set represented by a graph. The graph is typically very large, with nodes corresponding to objects and edges to relations or interactions between objects. In the real world, generally, the edges represent different types of relations and not just one relationship type. Several efforts have been made in order to analyze social networks [5, 6]. From a data mining perspective, the area that studies social networks is called link mining or link analysis [7]. One of the challenges for link analysis is group detection, which is the identifying of groups of objects that belong to the same group or cluster, based on their attributes as well as on their link structure.

In multi relational social network is extracted the different relations that exist in the network. Each relation can be modeled by a graph, depending on the customer data and call detail record information of telecommunication. [8]. Since Telecommunication companies generate and deal with a tremendous amount of data, including: call detail data, network data and customer data. Call detail data, which describes the calls that traverse the telecommunication networks, network data, which describes the state of the hardware and software components in the network, and customer data, which describes the telecommunication customers. The amount of data is so great that manual analysis of the data is very difficult, if not impossible. The proposed work describes the main applications of data mining in telecommunications: churn management by using Multi-relational Social Network.

Telecommunication social networks are social networks where two customers are considered connected if they have social relationships, in this contest, is based on the duration of voice calls, call frequency etc. that are exchanged during a certain period. These networks are more complex as their relationships involve different types of collaboration or interaction. Thus we can consider that Telecommunication Social Networks are a kind of Multi-Relational social Networks. In this work, we use attributes of call detail data and customer data as different relationship types to model our Multi-relational Telecommunication social network.

After the analysis of the available data constructed a Weights Multi-relational Social Network was constructed, in which each relation caries a different weight, representing how close two customers are with one another.

In the past, churn has been identified as an issue across most industry sectors. In its most general sense it refers to the rate of loss of customers from a company's customer base. There is a simple reason for the attention churn attracts: churning customers mean a loss of revenue. Emerging from business spaces like telecommunications (telecom) and broadcast providers, where churn is a major issue, it is also regarded as a crucial problem in many other businesses, such as online games creators, but also online social networks and discussion sites. Companies aim at identifying the risk of churn in its early stages, as it is usually much cheaper to retain a customer than to try to win him or her back. If this risk can be accurately predicted, marketing departments can target customers efficiently with tailored incentives to prevent them from leaving. In its most general sense churn refers to the full or partial defection of a customer. In the telecom industry, a subscriber is said to have churned when he leaves one carrier to move to another [9].

Churn rate is defined as the total gross number of subscribers who leave the service in the period divided by the average total customers in the period. The churn rate of a telecom company is a key measure of risk and uncertainty in the marketplace and will be quoted in the company annual report. Keaveney gives several factors influencing churn in the service industry such as pricing, inconvenience, core service failures, customer service failures and dissatisfaction with provider ethics [10]. From a global perspective, the average churn of mobile operators' in telecom industry was 2% and thus led them to loss about 100billion dollars per year [11]. Churning is costly processes for the company as it is much cheaper to retain a customer than to acquire a new one. Kotler [12] indicated that the cost for persuading a regular customer to stay will be 16 times more than that for developing a new customer, while the cost for attracting a new customer will be 5 or 6 times more than that for retaining a regular one. Reichheld and Sasser [13] indicated that, for the service industries, customer retention has more influence on promoting industry profit than other competitive factors such as the market share, unit costs, scale economy, etc. They further pointed out that an enterprise is expected to increase its profits by 25-95% while it has reduced its customer churn rate by 5%, which shows the significant impact of customer churn on business achievement. In order to reduce the losses caused by customer churn, telecom operators have to find those who are the most valuable customers and are incline to churn, and then carry out customer acquisition activities and customer retention policies for those customers.

Data mining techniques such as Decision Trees, Nearest Neighbor, and Artificial Neural Networks are used in churn analysis to perform two key tasks: Predict whether a particular customer will churn and reasons for particular customer churn [14]. The problem confronting wireless telecommunications' management is that it is very difficult to determine which subscribers leave the company and why. It is therefore more difficult to predict which customers are likely to leave the company, and devise cost effective incentives that will convince likely churners to stay.

II. LITERATURE SURVEY

In this section, related literature about social network analysis, components of social networks, Types of social Networks and importance of multi-relational social Network in Telecommunication industry will be reviewed and discussed.

A Social Network Analysis

Networks have been studied as graphs in mathematics, physics, sociology, engineering and computer science, biology and economics. Each field has its own theory of network and each field has its own way of aggregating collective behavior. In the past, the networks have been studied as objects of structure whose properties are static in time. Both these assumptions are far from truth. Real networks represent populations of individual components that are actually doing something involved in communication, generating power, sending data, or even making decisions. Here the structure of individual components is important because they affect their individual behavior or the behavior of the system as the whole. Also, the networks are dynamic objects, not because things happened in these systems, but because of the networks they are evolving and changing in time, with respect to activities or decision of the individual Components. Therefore, what happens and how it happens depend on the network, which in turn depend on what has happened previously [15].

Social Network Analysis (SNA) is a set of research procedures for identifying structures in systems based on the relations among actors. Grounded in graph and system theories, this approach has proven to be a powerful tool for studying networks in physics and social world, including on the web [16, 17, and 18].Social Network Analysis focus on relations and ties in studying actor's behavior and attitudes. Thus the position of the actors within a network and strength of ties between them become critically important. Social position can be evaluated by finding the centrality of a node identified through a number of connections among network members. Such measures are used to characterize degrees of influence, prominence and importance of certain members. Ties strengths mostly involve closeness of bond.

There is general statement that strong ties contribute to intensive resource exchange and close communities, whereas weak ties provide integration of relatively separated social groups into larger social networks. The methods of social network analysis have attracted considerable interest from the social, behavioral and computer science community in recent decades. Much of this interest can be attributed to the appealing focus of social network analysis on relationships among social entities, and on the patterns and implications of these relationships. From the view of social Network Analysis, the presence of regular patterns in relationship, are referred as structure and the quantities that measure structure as structural variables. The focus on relations, the patterns of relations, and role played by each customer requires a set of methods and analytic concepts that are distinct from the methods of traditional statistics and data analysis.

B. Components of Social Networks.

A social network consists of a set of actors and one or more types of relations between them, Such as information exchange or economic relationship. An actor is a social entity. It could be a person or any other entity for which a relationship with another entity could be defined. The relationship between a pair of actors is called a tie, link or pair. Each link may be directed or undirected, binary (present or absent) or weighted (a set of values, usually with higher value implying stronger relationship).Links could also be of particular types, e.g., friendship, kinship. All links of the same type can be grouped together as a relation. A dyad consists of a pair or actors and the ties between them. A triad is a subset of three actors and the ties among them. Relationship among larger subsets of actors includes the subgroups or a group. Social network encompasses a set of actors and all the relations that could be defined on them. Usually depending on the number of actor type's n, a social network may be identified as being an n-mode network.

C. Types of Social Networks

There are two types of social networks: Homogeneous and Heterogeneous. Homogeneous social networks are those are only one kind of relationship between the customers of the telecommunication network for example the relationship may be friendship between the two customers are linked.

Koustuv Dasgupta and Rahul Singh in their paper explain the social ties and their relevance to churn in mobile telecom network and their analysis explores the propensity of a subscriber to churn out of a service provider's network depending on the number of ties (friends) that have already churned. [19]. However the strength of the ties is driven not only by the individuals involved in the ties, but also the network structure. The draw backs of this work are the local relationship between network topology and tie strength affects any global information diffusion process like churn. On the other hand, heterogeneous social networks represent several kinds of relationship between customers, and can be called as Multi-relational social networks [20].

Victor Stroele and Jonice Oliveira in their paper describe the method of mining and analyzing multi-relational social networks. Their work focuses on using data mining techniques to identify intra and inter organization groups of people with similar profiles that could have relationships amongst them. For this, they constructed a multi-relational scientific social network where researchers may have four different types of relations such as, project participation, coauthored publications, Advisory works, and Technical Production. In which they consider the relationship of co-authors is one of the most important [21]. Hui-JuWu,I-Hsien Ting and Kai-Yuwang in their paper describes the method of combining social network analysis and web mining techniques to discover interest groups in the blog space. The disadvantage of this method is that they consider the relationship value as 1 or 0 to represents the presence or absence of relationship between the members, so that accurate link value will not be measured and accuracy in discovery of social relationship between the blogs and in discovery of the association between members will be less appropriate [22].

D. Multi-Relational Social Networks

Indeed, most telecommunication social network mining methods consider only homogeneous social Networks. However, in the real world, almost all the social networks have several kinds of relationships between customers. A particular kind of problem about Multi-relational Social Network is in extracting the different relations that exist in the network. Each relation can be modeled by a graph. Depending on the information one wants to obtain, analyzing one of the relations will be more important than others.

Thus, for a better analysis of churn prediction in telecommunication network one needs to select the relations that have a positive effect on the social network. Telecommunication social networks are social networks where two customers are considered connected if they have social relationships, in this contest, is based on the duration of voice calls, call frequency etc. that are exchanged during a certain period. These networks are more complex as their relationships involve different types of collaboration or interaction. Thus the Telecommunication Social

Networks can be considered as a kind of Multi-Relational social Networks.

The proposed work, considers the attributes of call detail data and customer data as different relationship types to model our Multi-relational Telecommunication social network. After the analysis of the available data Weights Multi-relational Social Network was constructed, in which each relation carries a different weight, representing how close two customers are with one another. Typical work on social network analysis includes the discovery of group of customers who shares similar properties. This is known as community mining . Most algorithms for community mining assume that there is only one social network, representing a relatively homogenous relationship. In reality there exist multiple, heterogeneous social networks, representing various relationships. A new challenge is the mining of hidden communities on such heterogeneous social networks, to group the customers as churners and non-churners in Telecommunication social network.

III. METHEDOLOGY

Telecommunication social networks are social networks where two customers are considered as connected if they have social relationships. The social relationship is based on the duration of voice calls; call frequency and age that are exchanged during a certain period. These networks are more complex as their relationships involve different types of collaboration or interaction. Thus Telecommunication Social Networks can be a kind of Multi-Relational social Networks. The attributes of call detail data and customer data are considered as different relationship types to model Multi-relational Telecommunication social network. After the analysis of the available data Weighted Multi-relational Social Network, has been constructed, in which each relation caries a different weight, representing how close two customers are with one another.

A. Data Collection

Telecom data consist of Customer data and Call Detail Data (CDR).Customer data consist of attributes like customer name, Address, age, Customer ID, sex, and Call Details Records (CDR), is a data record produced by a telephone exchange ,it contains the fields such as Calling number , Called number , Call type (Outgoing / Incoming),Service used (Voice/SMS/Etc.),Call start time, Call end time, Duration , Cell ID , Call sequence Number ,Switch number or ID. The 32,000 records are considered while analysis and contains detailed information about voice calls, SMS, value added calls of users. During pre-processing we excluded the service numbers, e.g. the operator's customer service number, the number similar to 1-800…so that number of nodes in Telecom Social Networks can be reduced. The attributes that are considered while construction of multi-relational telecom Social Networks are customer name, age from customer data and calling number, called number ,duration from CDR.

B. Representing Social Network Data

Conventional data consist of rectangular array of measurements. The rows of arrays are the cases, subjects, or observations. The columns consist of scores (quantitative or quality) on attributes, variables, or measures. Social network data consist of a square array of measurements. The rows of array are the cases, subjects, or observations. The column of the array is the same set of cases, subjects, or observations. Each cell of the array describes the relationship between actors. The most common method of representing the social network data is by using matrices. The common form of matrix in social network analysis is a square matrix. The most common matrix is binary in which cells of a matrix is filled with either 0 or 1.The relationship used here is friendship, that is if the two customer is connected then the cell is filled with 1 , otherwise it is filled with 0.This method of representing the social data is called adjacency matrix. The adjacency matrix is extremely useful to conduct various formal analysis of network, in particular it tells us how many paths of length 1 there are from each actor to each other actor. In general, it can be shown that the powers of the adjacency matrix give the number of walks of length n from each actor to each other actor.

IV. RESULTS

A. Construction of weighted Multi relational Telecom Social Network Model

The term network has different meaning in different disciplines. In the data mining disciplines social network is defined as a set of actors (or agents, or nodes, or points, or vertices) and one or more relationships (or links, or edges, or ties) between pairs of actors. Network that represent a single type of relation among the actors are called simplex, while those that represent more than one kind of relation are called multiplex.

Multiplex relations are analyzed using different networks, one for each relation type. Each tie or relation may be directed, or undirected that represents co-occurrence, co-presence, or a bonded-tie between the pair of actors. Directed ties are represented with arrows, and bonded tie relations are represented with line segments. Directed ties may be reciprocated (A links to B and B links to A) such ties can be represented with a double-headed arrow. The ties may have different values, that is binary values representing presence or absence of a tie, signed values representing a negative tie, a positive tie, or no tie, ordinal values representing whether the tie is strongest, next strongest, and numerically valued measured on an interval or ratio scale. The Multi-Relational Social Network constructed for Telecom data by considering the multiple relations as friendship, age and frequency of call, volume of call as ties between the customers is given below figure 1.

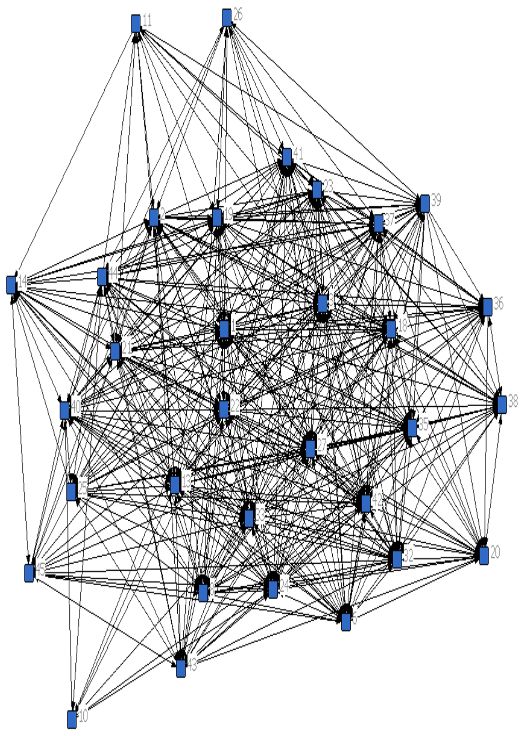

Figure 1: Multi-relational Telecom Social Network

B. Centrality Measures:

The centrality of a node in a network is a measure of its structural importance that is, how important a person is within a social network. There are three approaches to calculate the centrality of a node: based on degree, on closeness, and on betweenness. Degree approaches are based on the idea that having more ties means being more important. Finally, when using betweenness approaches, it is being in between many other actors what makes an actor central. These three approaches describe the locations of nodes in terms of how close they are to the “centre” of the action in a network.

A central actor, presumably, has a stronger influence on other network members (i.e. central positions tend to be powerful positions). Thus, measures of centrality are often interpreted as measures of power. Actors who have many ties with other actors may be in an advantageous position. Having many ties may mean having alternative ways of satisfying needs, it may mean having access to more resources, and it may also mean acting frequently as a third-party or deal maker in exchanges among others. With directed data, however, it is important to distinguish between in-degree centrality and out-degree centrality. That is, many other actors seek to direct ties towards them, and this may be an indicator of importance. Actors with unusually high out-degree may be able to influence many others, or make many others aware of their views. Thus, actors with high out-degree centrality are often called influential actors.

Closeness centrality using geodesic distance is the reciprocal of the sum of geodesic distances to all other vertices in the graph. These scores can be normalized dividing by the maximum value. Another way of thinking about how close an actor A is to all others is to calculate the proportion of other actors that A can reach in one step, two steps, three steps, etc (or, alternatively, the proportion of nodes that reach A in *n* steps).

We calculated a single index for each node by summing up the proportion of other nodes reached (for the first time) at a given distance, appropriately weighted (e.g. 1 for nodes at distance 1, ½ for nodes at distance 2…). These scores can be then normalized dividing by the maximum value, if this is considered appropriate. The idea behind betweenness centrality is that the customer who is between the two customers is considered as powerful customer as he is able to control the flow of information, resources, between them. Nodes with high betweenness are often called key-players. Betweenness centrality using only shortest paths is the nodes that occur on many shortest paths between other nodes have higher betweenness than those that do not. The betweenness of a node A is the fraction of all the possible shortest paths between any two nodes in the network (excluding A) that pass through A. Centrality measures for the nodes are given below Table 1.

Id	Degree Centrality	Betweenness Centrality	Closeness Centrality	Harmonic Closeness	Total Centrality	Endogenous	Exogenous
1	31	9.553	31.000	31.000	62.000	31.000	31.000
2	27	4.745	35.000	29.000	49.000	24.000	25.000
3	28	4.644	34.000	29.500	49.000	23.000	26.000
6	25	2.328	37.000	28.000	47.000	23.000	24.000
8	26	2.515	36.000	28.500	45.000	22.000	23.000
10	9	0.000	53.000	20.000	11.000	9.000	2.000
11	12	0.259	50.000	21.000	15.000	8.000	7.o000
14	19	1.554	43.000	25.000	50.000	22.000	28.000
19	28	5.045	34.000	29.500	30.000	11.000	19.000
20	19	0.520	43.000	25.000	47.000	24.000	23.000
21	25	3.090	37.000	28.000	48.000	23.000	25.000
22	30	7.378	32.000	30.500	29.000	14.000	15.000
23	28	4.764	34.000	29.500	38.000	25.000	13.000
24	29	5.895	33.000	30.000	58.000	28.000	30.000
26	14	0.270	48.000	22.500	47.000	23.000	24.000
27	29	6.438	33.000	30.000	46.000	18.000	28.000
32	26	2.578	36.000	28.500	36.000	19.000	17.000
33	29	5.895	33.000	30.000	21.000	14.000	7.000
36	21	0.567	41.000	26.000	50.000	21.000	29.000
37	26	2.903	36.000	28.500	48.000	22.000	26.000
38	20	0.354	42.000	25.500	54.000	25.000	29.000
39	25	3.143	37.000	28.000	44.000	18.000	26.000
41	26	3.824	36.000	28.500	33.000	16.000	17.000
43	20	1.940	42.000	25.500	46.0000	23.000	23.000
44	27	6.891	35.000	29.000	37.000	19.000	18.000
45	16	0.442	46.000	23.500	34.000	25.000	9.000

Table 1: Centrality Measures for Telecom Social Network.

IV. DISCUSSION

The structural property of the Telecom Social Network of size 32000 nodes has been realized in the research work. For illustration purpose 26 nodes has taken. Figure 2 shows distribution of degree centrality, betweenness centrality, closeness centrality, harmonic centrality and total centrality of the nodes. It has the values range between 9-31, 0.000-9.553., 31.000 – 53.000, 20.000-31.000 and 11.000-62.000 respectively. Endogenous refers to out-degree and exogenous refers to in-degree of the nodes. Hence showing they are closely connected to each other. These measures show the intensity of the influence of the customer to churn other customers in the network. By this measures we can found out the more influenced customers in the network and by providing the more services for those customers, we can retain them.

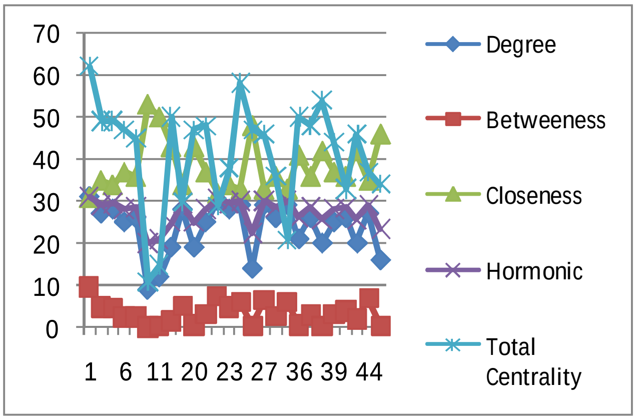

Figure 2: Centrality distribution

VI. CONCLUSION

In reality the social networks are heterogeneous in nature, the accuracy of the churning model can be increased by considering the multiple relationship between the customer while construction of the telecom social network. In this work, we use attributes of call detail data and customer data as different relationship types to model our Multi-relational Telecommunication social network. After the analysis of the available data we constructed a Weights Multi-relational Social Network, in which each relation caries a different weight, representing how close two customers are with one another. The centrality measures depicts the intensity of the customer closeness, hence we can found the customer who influence the other customer to churn.

REFERENCES

[1] Shaun Doyle, "Social network analysis in the Telco sector Marketing applications", Journal of Database Marketing & Customer Strategy Management (2008) 15, pp130-134.doi: 10.1057/dbm.2008.8

[2] Lei Tang and Huan Liu, "Graph Mining Applications to

Social Network Analysis", Managing and Mining Graph Data, Springer US, PP 487- 513, 2010.

[3] Kleinberg, J. "Authoritative sources in a hyperlinked environment," Journal of the ACM (JACM), vol. 46, no. 5, pp.604–632, 1999.

[4] Sen Wu; Lingyu Wu; Haiyan Jin; and ShujuanGu," Customer Ranking Authority-Hub Algorithm for Mobile Communications", in China, Fourth International Conference on Fuzzy Systems and Knowledge Discovery, 2007.FSKD 2007. pp. 559 – 563.

[5] S. Sumathi and S.N. Sivanandam: "Data Mining in Telecommunications and Control",Studies in Computational Intelligence (SCI) 29, 615–627 (2006).

[6] S.Wasserman and K.Faust. "Social Network Analysis: Methods and Applications". Cambridge University Press, Cambridge, UK, 1994.

[7] L. Freeman. "Centrality in social networks: Conceptual clarifications". Social Networks, 1:215- 239, 1979.

[8] D. Liben-Nowell, J. Kleinberg. "The Link Prediction problem for social networks", in roceedings of the twelfth international conference on Information and Knowledge management, pages 556-.559, 2003.

[9] H. Hwang, T. Jung, and E. Suh, "An ltv model and customer segmentation based on customer value: a case study on the wireless telecommunication industry," Expert Syst. Appl., vol. 26, no. 2, pp. 181–188, 2004.

[10] S. M. Keaveney, "Customer switching behavior in service industries: An exploratory study," The Journal of Marketing, vol. 59, no. 2, pp.71–82, 1995. [Online]. Available: http://www.jstor.org/stable/1252074.

[11] A. Berson, S. Smith, and K. Thearling, "Building data ining applications for CRM," New York: McGraw-Hill, 2000.

[12] P. Kotler, "Marketing Management," Prentice-Hall International, 1994.

[13] F. F. Reichheld, W. E. Jr. Sasser, "Zero defections: Quality comes to services", Harvard Business Review, Vol.68, 1990, pp. 105-111.

[14] D. Cai, Z. Shao, X. He, X. Yan, and J. Han, Community mining from multi-relational Networks ", in Proceedings of the 2005 European Conference on Principles and Practice of Knowledge Discovery in Databases (PKDD'05), Porto, Portugal, 2005.

[15] DuncanJ.Watts.Six Degree:"Ths science of a connected Age", W.W.Norton& Company,February 2003.

[16] S.D. Berkowitz. "An Introduction to Structural Analysis", The Network Approach to Social Research. Toronto: Butterwoths, 1982.

[17] LadaA.Adamic. "The small world web",In proceedings of the third European Conference on Research and Advanced Technology for Digital Libraries, ECDL, number 1696, pages 443-452. Springer-verlag, 1999.

[18] Reka Albert and Albert-Laszlo Barabasi, "Statistical mechanics of complex networks", Reviews of Modern Physics, 745:47-97, 2002.

[19] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi, "Social ties and their relevance to churn in mobile telecom networks," in EDBT '08: Proceedings of the 11th international conference on Extending database technology, (New York, NY, USA), pp. 668–677, ACM, 2008.

[20] Wei, C., & Chiu, I (2002)."Turning telecommunications call details to churn prediction: A data mining approach". Expert Systems with Applications, 23(2), 103-112.

[21] Victor Stroele, JoniceOliveira,"Mining and Analyzing Multi-relational Social Networks", In International Conference on Computational Science and Engineering, 2009. pp.711-716.

[22] Hui-Ju Wu, I-Hsien Ting, Kai-Yu Wang, "Combining Social Network Analysis and Web Mining Techniques to Discover Interest Groups in the Blog space", 2009 Fourth International Conference on Innovative Computing, Information and Control.