

AN EFFICIENT METHOD OF BUILDING THE TELECOM SOCIAL NETWORK FOR CHURN PREDICTION

PUSHPA¹ AND DR.G SHOBHA²

¹Associate Professor, Department of Computer Science & Engineering, AIT
Chickmagalur

flowersunpr@yahoo.co.in

²Professor , Department of Computer Science & Engineering,R.V.College of
Engineering,Bangalore

shobhag@rvce.edu.in

ABSTRACT

As the deregulation and great advance of new technologies, the competition in wireless telecommunication industry is getting severe. Hence, churn prediction and management have become of great concern to the mobile operators. Therefore they wish to retain their subscribers and satisfy their needs. Previous methods address the homogeneous social network analysis for churn prediction by considering the single relation. From the point of view of data mining, a social network is a dynamic, heterogeneous and multirelational in nature. Typical work on social network analysis includes the construction of multi-relational telecommunication social network and discovery of group of customers who share similar properties and classify the customers as churners and non-churners. In this paper we explore the various methods of representing the social networks. Considering the multi-relational data while constructing the telecom social network will increase the efficiency in prediction of customer churning.

KEYWORDS

Telecommunication, Social Network Analysis (SNA), Churn prediction, Call Detail Records (CDR).

1. Introduction

Telecommunication companies use the simple, traditional and advanced mathematical techniques to analyse large amount of data and deliver forecasts, explanations and predictions of how customers, systems, network and marketplace are likely to react to different situations. The telecommunications industry generates and stores a huge amount of data. These data are classified as 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 [1]. The amount of data, is so large that manual analysis of the data is difficult. The problem with this approach is that it is time consuming to obtain the knowledge from human experts and, in many cases the experts do not have the requisite knowledge. The need to handle such large volumes of data led to the construction of social network in which various attributes and relations are considered.

A network is a set of objects where each of them is connected with other objects. We can represent a network as a graph where the nodes, or vertexes, are related by edges. A social network reflects a social structure that can be represented by individuals or organizations, and their relations. In general the relations represent one or more types of interdependency (such as values, visions, idea, and religion) or more specific relationships (financial exchanges, friendship, communication, conflicts, and others). 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 [2, 3]. From a data mining perspective, the area that studies social networks is called link mining or link analysis [3]. 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 we are 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. [4].

The companies in the telecommunications industry face the problem of churning. Churning is the process of customer turnover. This is a major concern for the companies having many customers who can easily switch to other competitors. Data mining techniques such as Decision Trees, Nearest Neighbour, 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 [5]. 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.

2. 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.

2.1 Social Networks

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 behaviour. 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 behaviour or the behaviour 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. It is this view of the networks “a continuously evolving and self constituting system” [7].

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 [8, 9, and 10]. Social Network Analysis focus on relations and ties in studying actor's behaviour 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, behavioural 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.

2.2 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.

2.3 Types of Social Networks

There are two types of social networks: Homogeneous and Heterogeneous. Homogeneous social networks are those is only one kind of relationship between the customer 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 explains 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. [11]. However the strength of the ties is driven not only by the individuals involved in the ties, but also the network structure. Further, we believe that 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 [6]. Victor Stroele and Jonice Oliveira in there paper describes the method of mining and analyzing multi-relational social networks. There 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, co-

authored publications, Advisory works, and Technical Production. In which they consider the relationship of co-authors is one of the most important [13]. Hui-Ju Wu, I-Hsien Ting and Kai-Yu wang in their paper describes the method of combining social network analysis and web mining techniques to discover interest groups in the blogspace. The disadvantage of this method is that they consider the relationship value will be denoted as 1, if the value of measurement is larger than 3 else will be denoted as 0, 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 [12].

2.4 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 modelled 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 context, 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 we constructed a Weights Multi-relational Social Network, 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 homogeneous 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.

3. METHODOLOGIES

Telecommunication social networks are social networks where two customers are considered connected if they have social relationships, in this context, 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 we constructed a Weights Multi-relational Social Network, in which each relation carries a different weight, representing how close two customers are with one another.

3.1 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, etc... 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 data set is in terms of Gigabytes 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.

3.2 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.

3.3 Construction of Customer 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.

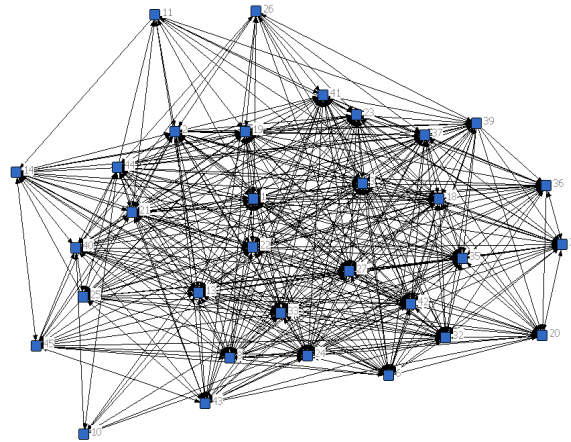


Figure 1: Multi-Relational Telecom Social Network

3.4 Social Network Analysis:

Social network analysis (SNA) has gained prominence due to its use in different applications such as product marketing, search engines and organizational dynamics. Recently there has been a rapid increase in interest regarding social network analysis in the data mining community. The basic motivation is the demand to exploit knowledge from copious amounts of data collected, pertaining to social behaviour of users in online environments. Data mining based techniques are proving to be useful for analysis of social network data, especially for large datasets that cannot be handled by traditional methods. Social network analysis is the mapping and measuring of relationships and flows between people, groups, Organizations, computers or other information/knowledge processing entities. (Valdis Krebs, 2002). Social Network Analysis (SNA) is a method for visualizing our people and connection power, leading us to identify how we can best interact to share knowledge.

3.4.1 Basic Demographics

The size of a network is defined by number of nodes or by number of edges of the network. In any network there are $(k * k-1)$ unique ordered pairs of nodes (that is AB is different from BA, and leaving aside self-ties), where k is the number of nodes. The density of a binary network is determined by the total number of ties divided by the total number of possible ties. For a valued network it is the total of all values divided by the number of possible ties. The density of a network is a measure of speed at which information diffuses among the nodes and it is used to measure the connectivity of the network.

The distance between two nodes is the minimum number of edges that takes to go from one to another. This is also known as the geodesic distance. The nodes that are nearer are able to exert more power than those that are more distant. If two actors are adjacent, the distance between them is 1 (i.e. it takes one step, or edge, to go from one to the other). If A links to B, and B links to C (and A does not link to C), then actors A and C are at a distance of 2.

3.4.2 Local Structure in Network

The smallest social structure is the dyad, that network consist of a pair of actors. For binary ties (present or absent), there are two possibilities for each pair of nodes in the network either they

have a tie, or they don't. If we are considering a directed relation, there are three kinds of dyads (no tie, one links to the other but not vice versa, or they both link to each other). A potentially interesting analysis is to study the extent to which a population is characterized by reciprocated ties; this may tell us about the degree of cohesion, trust, and social capital that is present. There are two different approaches to quantifying the degree of reciprocity in a population. One approach is to focus on the dyads, and calculate the proportion of pairs that have a reciprocated tie between them. A second (equivalent) approach consists in focusing on the relation, rather than on the actors. When we focus on the dyads and we consider all possible pairs.

Another form of social structure is a triad, which is formed by three actors. Triads allow for a much wider range of possible sets of relations (with directed data, there are 64 possible types of relations among 3 specific actors), including relationships that exhibit hierarchy, equality, and the formation of exclusive groups (e.g. where two actors connect, and exclude the third). Triad structure exhibit the property called transitive, that is, display a type of balance where, if A directs a tie to B, and B directs a tie to C, then A also directs a tie to C.

3.4.3 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). One can then calculate 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, $\frac{1}{2}$ 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 being in between actors makes you powerful because you may be able to control the flow of e.g. information, resources, gossip... 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

International Journal of Data Mining & Knowledge Management Process (IJDKP) Vol.2, No.3, May 2012
 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: Centrality measures for the nodes 11, 18,2,14.

Key	Value for node 11	Value for node 18	Value for node 2	Value for node 14
ID	11	18	2	14
Degree	12	27	27	20
Betweenness	0.259	3.917	4.745	0.588
Closeness	50.000	35.000	35.000	42.000
Harmonic Closeness	21.500	29.000	29.000	25.500
Eigenvector	0.090	0.194	0.191	0.130
In degree	7.000	23.000	25.000	19.000
Out degree	8.000	24.000	24.000	11.000

4. CONCLUSION AND FUTURE WORKS

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 to extract the hidden communities of the churners and non-churners. Social network analysis (SNA) has gained importance due to its use in different applications such as product marketing, search engines and organizational dynamics. The basic motivation is the demand to exploit knowledge from copious amounts of data collected, pertaining to social behaviour of users in online environments. Data mining based techniques are proving to be useful for analysis of social network data, especially for large datasets that cannot be handled by traditional methods.

RERERENCES.

- [1] S. Sumathi and S.N. Sivanandam: “Data Mining in Telecommunications and Control”, Studies in Computational Intelligence (SCI) 29, 615–627 (2006).
- [2] S. Wasserman and K.Faust. “Social Network Analysis: Methods and Applications”. Cambridge University Press, Cambridge, UK, 1994.
- [3] L. Freeman. “Centrality in social networks: Conceptual clarifications”. Social Networks, 1:215-239, 1979.
- [4] D. Liben-Nowell, J. Kleinberg. “The Link Prediction problem for social networks”, In Proceedings of the twelfth international conference on Information and Knowledge management, pages 556-.559, 2003
- [5] Wei, C., & Chiu, I (2002).” Turning telecommunications call details to churn prediction: A data mining approach”. Expert Systems with Applications, 23(2), 103-112.

- [6] 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.
- [7] Duncan J.Watts.Six Degree:"Ths science of a connected Age".W.W.Norton & Company, February 2003.
- [8] S.D. Berkowitz. "An Introduction to Structural Analysis": The Network Approach to Social Research. To ronto:Butterwoths,1982.
- [9] Lada A.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.
- [10] Reka Albert and Albert-Laszlo Barabasi."Statistical mechanics of complex networks". Reviews of Modern Physics, 745:47-97, 2002.
- [11] 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.
- [12] Hui-Ju Wu, I-Hsien Ting, Kai-Yu Wang, Combining Social Network Analysis and Web Mining Techniques to Discover Interest Groups in the Blogspace, 2009 Fourth International Conference on Innovative Computing, Information and Control.
- [13] Victor Stroele,Jonice Oliveira,"Mining and Analyzing Multirelational Social Networks",in International Conference on Compputational Science and Engineering,2009.pp.711-716.

Authors

Mrs. Pushpa, Associate Professor in the Department of Computer Science and Engineering, AIT, Chickmagalur and Research Scholar at R.V.College of engineering, Bangalore. She had 12 years of teaching experience. She has completed her Master of Technology from Visvesvaraya Technological University. She had published research papers in international journals.Her research interests are in the area of data mining and Computer Networks.



Dr G Shobha, Dean Circuit branches and Professor in the department of Computer Science and Engineering. She had 18 years of teaching experience and 5 years of Research experience. She had completed her PhD from Mangalore University; in 2006.She had completed her M.S from BITS, Rajasthan in 1997. She has published many research papers in the national and international conferences and journals. She has got many Research projects to her credit funded by different agencies. Her research interests are, data mining, neural networks, Image processing and Computer Networks.

