International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 3 Issue 12, December-2014

Churn Prediction in Social Networks

Amit Chandavale
Dept. of Computer Engineering
MIT College of Engg
Pune, India

Reena Pagare(Assistant Professor)
Dept. of Computer Engineering
MIT College of Engg
Pune, India

***Abstract*— With the high speed development of online social networks, mobile devices and wireless technologies in the social network systems are increasingly available. Many people are integrated with social networking sites such as Facebook, Twitter and LinkedIn in their daily lives. Social networks have become major source of news, along with the traditional information propagation mediums such as television and newspapers. In social networks influence takes place in the form of "word-of-mouth". Customer churn take place due to the termination of contract or the customer move to the other service providers. It is fundamental issue to find a subset of most influential nodes (i.e. the customers) that are going to churn. By concentrating on these influential nodes initially we can restrict the churning of customers early. These influential nodes are nothing but existing customers which are using the service provided by the company. The customers are terminating the contract due to unsatisfactory service or they are moving to other service providers due to influence from others.**

Keywords*— Churn prediction, influence, Social Networks, Customer Churn, and Rumor

I. Introduction

Today there is large number of social networking sites and these sites are new platform for business organizations.

A social network is social structure consists of independent (or corporation) called "nodes", which are bounded(connected) by one or more distinct types of inter-association, such as friendship, common topic of interest, business ties, despise, or relationships of opinion, mastery or reputation.[9]

In a simple form, a social network is a map of specified ties or relationships, such as friendship, between the nodes being studied. The social contacts of an individual are the nodes to which an individual is connected. The network can also used to measure the social capital of an individual that is nothing but the value of an individual that he gets from social matrix. These concepts are often displayed in a social matrix diagram, where nodes are the points and ties are the lines. e.g. - Facebook, Twitter, LinkedIn etc. Today almost everyone is using social networks. People frequently communicate with each other through social networks. The decisions made by one can influence others decision on social groups in social networks. If one user of the social group terminates the existing service there is possibility that the other members also terminate the service. It is always better to retain the existing customers than acquiring new ones. It is better to identify these users in the social network before they churn. The churn prediction methods are useful to resolve this problem of identifying the churning customers.

Customer attrition, also known as customer churn, customer turnover, or customer defection, is the loss of customers. Customer churn is nothing but change and turn in which customer terminates the existing service and joins service provided by other service providers. By analyzing the social networks and activities of customers on social networks we can predict the churning customers. By preventing the churning early we can avoid the loss to an organization. There are various methods of churn prediction which helps to predict the churn early and avoid the losses. The churn prediction also plays a vital role in making business strategies.

In social networks people's decision influences other people's decisions. Influence in social networks takes place in the form of "word of mouth". The influence from one node to the other node is decided on the basis of many factors such as degree of influence, communication weight between two nodes in the social network. The most influential nodes in the social network will be influencing large number of nodes in the social network. If this influential node is going to churn then there is possibility of other nodes will also churn due to the influence from that node. By identifying these influential nodes as top churning customers we can apply retention policy to this set of customers before they decide to churn. Early detection of top churning customers will help in designing retention policies and can prevent the influence of these customers to other customers.

II. Churn prediction

A. Churn Prediction

- ***Definition*** - Churn prediction is the way to predict the customers which will be going to churn in the near future for which the behavior of existing customers is compared with the customers churned in the past. Churning is the termination of existing contract or leaving the service due to some reason.

- ***Customer Churn*** - Churn is the word derived from change and turn. It is nothing but discontinuation of contract. The customer leaves existing service and take service from other service providers is customer churn.

B. Types of Churn

There are three types of churn:

- ***Active***– The customer decides to terminate the existing service by making decision of switching to other service provider. These are the several reasons for this: the customer is not satisfied with the service quality (e.g. services are not up to the mark as given in service agreement), expensive service costs, no optional price plans, no rewards for customer loyalty, poor understanding of service scheme, bad support, no information about reasons and predicted resolution time for service problems, discontinuity or fault resolution, issues about the privacy.

- ***Rotational*** –The customer terminates the existing service by making decision of not switching to other service provider. There are several reasons for this, changes in the circumstances which are preventing the customer from further requirement of service, e.g. poor financial condition, it is impossible for customer to pay bill; or the customer changes its geographical location to the place where company services are not available.
- ***Passive*** – The Company terminates the contract by itself.

III. LITERATURE REVIEW

Churn prediction is an important area of focus for telecommunication providers. The newly emerging technique is the use of social network analysis to identify the potential churners. [1]

The Enhanced churn prediction method predicts the churning customers in telecommunication services by integrating SNA concepts. This method consists of three steps. These steps are Quantification of tie-strength, Influence propagation model, and Application of machine learning techniques to combine traditional and social predictor. In the Quantification of tie-strength step on the basis of calling attributes a call graph is constructed and the quantification of social ties is performed. The second step defines the model for churner influence propagation in call graph and the computation of overall influence at all nodes is performed.. The third step involves fed up information in the classification algorithm to predict the future churners such as service performance and usage metrics, billing, customer support call data and demographic information is combined with predictors which are socially relevant and social influence and this is aggregated. This method integrates the SNA concepts with traditional churn prediction methods. The approach is generic and applicable to any phenomenon that has influence diffusion. To target new services and applications the tie-strength and information diffusion model can be improved to detect social influencers. By linking the identity of user in the social media to subscriber identity of telecom domain the churn prediction can be improved. The decay of influence over time and distance (number of hops) is not considered in this method.

Pattern analysis framework for Churn Prediction [2] is the inductive customer churn analysis framework is designed with the main purpose of providing early suggestion to strategic planners before customers actually leaving the company.

Churn Prediction by Using Chat Graph [4] method for churn prediction two classification approaches are considered. The first is non-conventional in which the prediction is done independently for each instance. The second is to use iterative collective classification algorithm. The goal of this method is to predict the chat-activity churn, the construction of chat graph is considered. The nodes in the chat graph represent users and the directed edges between them indicate social ties between any two users. The edge is created between two users only when the chat initiated by one user is responded by other user and vice versa. The social tie strength is encoded as edge weights. The data driven approach in this method for churn prediction explores the underlying churn. The set of social features derived from graph theory and link analysis are not used in this method, these features can be used to capture the complex dependencies underlying churn.

Churn prediction by using clustering [5] method simply locates the churn users and then groups these users into different clusters on the basis of their online activities to deliver the appropriate retention solutions. This method consists of two steps prediction and clustering. The prediction step predicts the churn and non-churn users. The k-means clustering algorithm is used for classification of users. The churn users are analyzed and retention solutions are provided to prevent them from churning.

In Churn Prediction by Using Local Community Detection [6] method network is represented by an undirected and unweighted graph G = <V,E>, where V is the set of node and E is the set of edges. The general greedy scheme is used for community detection. Many quality functions are used for community identification. Local community-based attributes are relevant for churn prediction in real online social networks. The content and structure of the network is not considered.

Churn Prediction by Using Diffusion Process [7] discusses about identifying potential churners in an operator's network by exploiting social ties. The method starts with set of churners and their social relationships are captured in call graph. The above method concludes that social relationships play an influential role in affecting churn in operator's network. The graph theoretic in the network can be used to guide the diffusion process.

From this survey we come to know that only chat graphs and past churner's behavior taken into consideration that are not producing efficient results for churn prediction in online social networks. To get more efficient and accurate results we are taking temporal attributes of customers with influence maximization to predict the customers going to churn more effectively from the online social networks.

Influence Maximization: A Divide – and – Conquer Method [11] mines the most influential node from each community. This method divides the social network in different communities on the basis of their degree of influence and the speed of influence. In the first step it applies CGA algorithm for partition and combination process. In the second step it mines the most influential node from each community on the basis of degree of influence.

Multi – source – driven Asynchronous Diffusion Model [12] for video sharing in online social networks uses Multi – source influence to study the behavior of target user who has multiple neighbors will be influencing his/her decision. The diffusion model uses influence from multiple active sources and temporal information.

Rumor Restriction in Online Social Network [13] proposes two models for rumor restriction. LT model with γ - k rumor restriction use information threshold for each contaminated node to trust good information from decontaminated node and IC model with γ - k rumor restriction uses truth factor which indicates the probability that contaminated node becomes decontaminated after it is activated by decontaminated neighbor.

Thus from the above literature survey we come to know that we can take temporal attributes of most influential nodes into consideration for predicting the churning customer before they churn from the social groups. The use of influence maximization concept for churn prediction will give efficient results.

IV. PROPOSED SYSTEM

Fig.1 shows proposed system. Proposed system uses the combination of both the churn prediction method and the influence maximization techniques. The proposed system uses user's social networking data along with its call log details. The proposed scenario is shown in Fig 1 consists of different modules starts with the existing dataset.

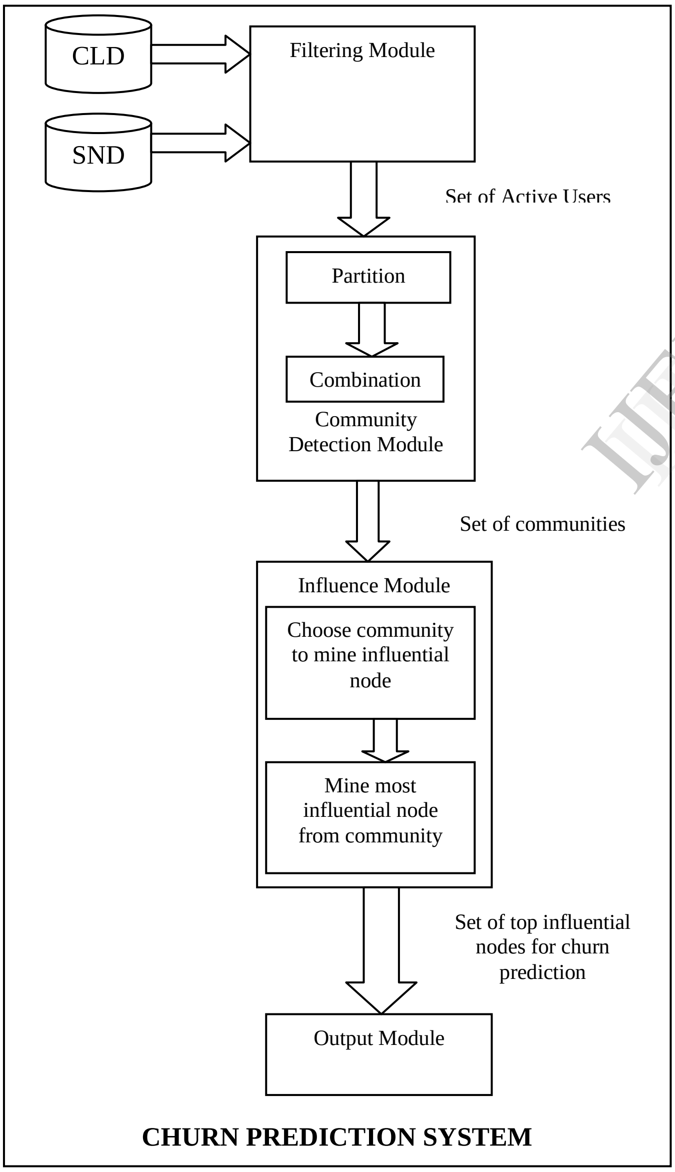

Fig. 1. *Proposed System*

A. *Filtering Module*

The system takes set of user's as input with user's social network data and call log details. The users are filtered on the basis of its activity of past six months. If the difference between the current date and the last call or message date of user is less than six months then that user will be active one and that user will be added to active set of users. If the difference is greater than six months then the user will be not active one. The users other than these users will be active one. The set of active users is the output of filtering module. The set of active users is passed to the community detection module.

B. *Community Detection Module*

This module consists of two sub modules partition module and combination module. This module takes set of active users as input. This module calculates the communication weight between all users. Then the users are partitioned into different communities. The combination step combines the two different communities having some users in common. The output of this step is set of communities.

- **Partition**
  The partition step takes set of active users as input. The communication weight is calculated on the basis of topic of interest of users from social network data and call duration and message duration from call log details of user. The formula to calculate communication weight is given below in (1).

$$W = \sum_{a_i \quad A} (TOI)*\alpha + \beta*(CLD) \qquad (1)$$

  Where W is the communication weight, TOI is the weight of topic of interest, CLD is weight of call log and α, β are tuning parameters.
  The system partitions set of users into different communities on the basis of community label assigned to each user. The speed of influence and the degree of influence from one node to other node is taken into consideration On the basis of community label assigned to each user the user will be added to their respective community. The formula for community partition is given below in (2).

$$a.C^z = \text{argmax}\{1- \prod_{1\le i\le y \quad b_j \in N.C_i^{z-1}} (1 - qab_j)\} \qquad (2)$$

  Where N is the set of neighbors and N.C is the set of neighbor communities, y is the number of communities and qabj is the influence speed of node a to node bj.

- **Combination**
  After partitioning the system into different communities the combination decay between communities is calculated. If the combination decay of communities is greater than the threshold then two communities are combined. If the combination decay is less than the threshold value then the communities will not combined. The formula for combination decay is given below in (3).

$$\mathrm{CoDecay}(CD^{f}_{g}) = \max_{a \in Cg,\, b \in L[Cg],\, b \in Cf} \qquad (3)$$

Where CDfg is the combination decay of community Cf to Cg. L[Cg] is the set of live nodes of community Cg. Pg({b}) is the influence degree increment of node a and Pg ({a}) is the influence degree increment of node a in its community Cg.

C. Influence Module

The influence module mines the most influential node from each community. In this module the communities are chosen to mine the influential node. The set of influential nodes is generated as output. This set of top churning nodes is passed as input to the next module.

- **Calculate Maximal increase in influence degree**
  The maximal increase in degree of influence for each community is calculated. Formula to calculate maximal increase in influence degree is given below in (4).

  $$\Delta Pg = \max\{Pg(Ik\text{-}1 \cup \{aj\}) - Pg(Ik\text{-}1) \mid aj \in Cg\} \qquad (4)$$

  Where ΔPg is the maximal increase in influence degree and I_{k-1} is the set of influential nodes in previous k-1 steps

- **Choose the community**
  To mine the most influential nodes choose the community with maximal increase in degree of influence among all communities. The formula is given below in (5).

  $$P[g,k] = \max\{P[g-1,k],\ P[G,k-1] + \Delta Pg\}$$
  $$P[g,0] = 0,\ P[0,k] = 0 \qquad (5)$$

  Where P[g,k] (g ∈ [1,G] and k ∈ [1,K]) is the influence degree of mining kth influential in the first m communities.The community selected to mine the influential node is represented by the sign function given below in (6):

  $$r[g,k] = \begin{cases} r[g-1,k], & P[g-1,k] \geq P[G,k-1] + \Delta Pg \\ m, & P[g-1,k] < P[G,k-1] + \Delta Pg \end{cases} \qquad (6)$$
  $$r[0,k] = 0$$

D. Output Module

The set of influential nodes mined from each community are generated as output. The set of communities will also be generated as output. The set of most influential nodes from each community are mined as output for churn prediction.

V. Mathematical Model

A. Set Theory

Let S be our churn prediction system which is defined in the following manner in (7)

$$S = \{U, SND, CLD, A, C, I, TCN\} \qquad (7)$$

Where, U is the set of users. SND is the social networking data of user. CLD is the call log data of user. A is the set of active users. C is set of communities. I is a set of influential nodes. TCN is a set of most influential nodes for churn prediction and N is the set of neighbors. SND and CLD belong to U. The sets used in mathematical model are given below in (8). Where, n is the number of users, i is the number of active users, z is the number of communities, k is the number of influential nodes, l is number of most influential nodes for churn prediction and t is number of neighbor nodes. Input sets are shown below in (9). The output sets are shown below in(10).

- **Set Theory**
  $U = \{U_1, U_2, U_3 \ldots\ldots\ldots\ldots U_n\}$
  SND = {user, relation, topic of interest}
  CLD = {from_user, to_user, time, date_of_call, call_duration, date_of_msg, msg_duration}
  $SND = \{SND_1, SND_2, SND_3 \ldots\ldots\ldots SND_n\}$
  $CLD = \{CLD_1, CLD_2, CLD_3 \ldots\ldots\ldots CLD_n\}$ (8)
  $A = \{A_1, A_2, A_3 \ldots\ldots\ldots A_j\}$
  $C = \{C_1, C_2, C_3 \ldots\ldots\ldots C_z\}$
  $I = \{I_1, I_2, I_3 \ldots\ldots\ldots I_k\}$
  $TCN = \{TCN_1, TCN_2, TCN_3 \ldots\ldots\ldots TCN_l\}$
  $N = \{b_1, b_2, \ldots, b_t\}$

- **Input**
  U = {SND, CLD}
  SND = {user, relation, topic of interest}
  CLD = {from_user, to_user, time, date_of_call, call_duration, date_of_msg, msg_duration}
  $SND = \{SND_1, SND_2, SND_3 \ldots\ldots\ldots SND_n\}$
  $CLD = \{CLD_1, CLD_2, CLD_3 \ldots\ldots\ldots CLD_n\}$ (9)

- **Output**
  $C = \{C_1, C_2, C_3 \ldots\ldots\ldots C_z\}$
  $TCN = \{TCN_1, TCN_2, TCN_3 \ldots\ldots\ldots TCN_l\}$ (10)

B. State Transition Diagram

Fig.2 shows in state S0 the set of users is filtered to find the set of active users If the user is active then it is added to the active user set and transferred to state S1.If the user id not active it is not added t active user set and transferred to state S2. If Ui is active user then add user to active user set and pass to state S3. If the user not active one then directly exits from the system from S2 to S18.

The set of active users is passed to state S4. Calculate communication weight of each active user and pass it to state S5.The communication weight is calculated on the basis of call and message duration of user from CLD and topic of interest from SND of that particular active user in S5. In S6 for each user of active user set apply partition users will be partitioned into different communities according to their community label. Community label is assigned on the basis of degree of influence and speed of influence. The user's which does not belong to any community after partition passed to S7 from S6 and from S7 system directly exits to state S18. Each user Ai is added to particular community on the basis of community label assigned to that user after partition and passed to state S8. The set of communities is generated from state S8 to state S9.

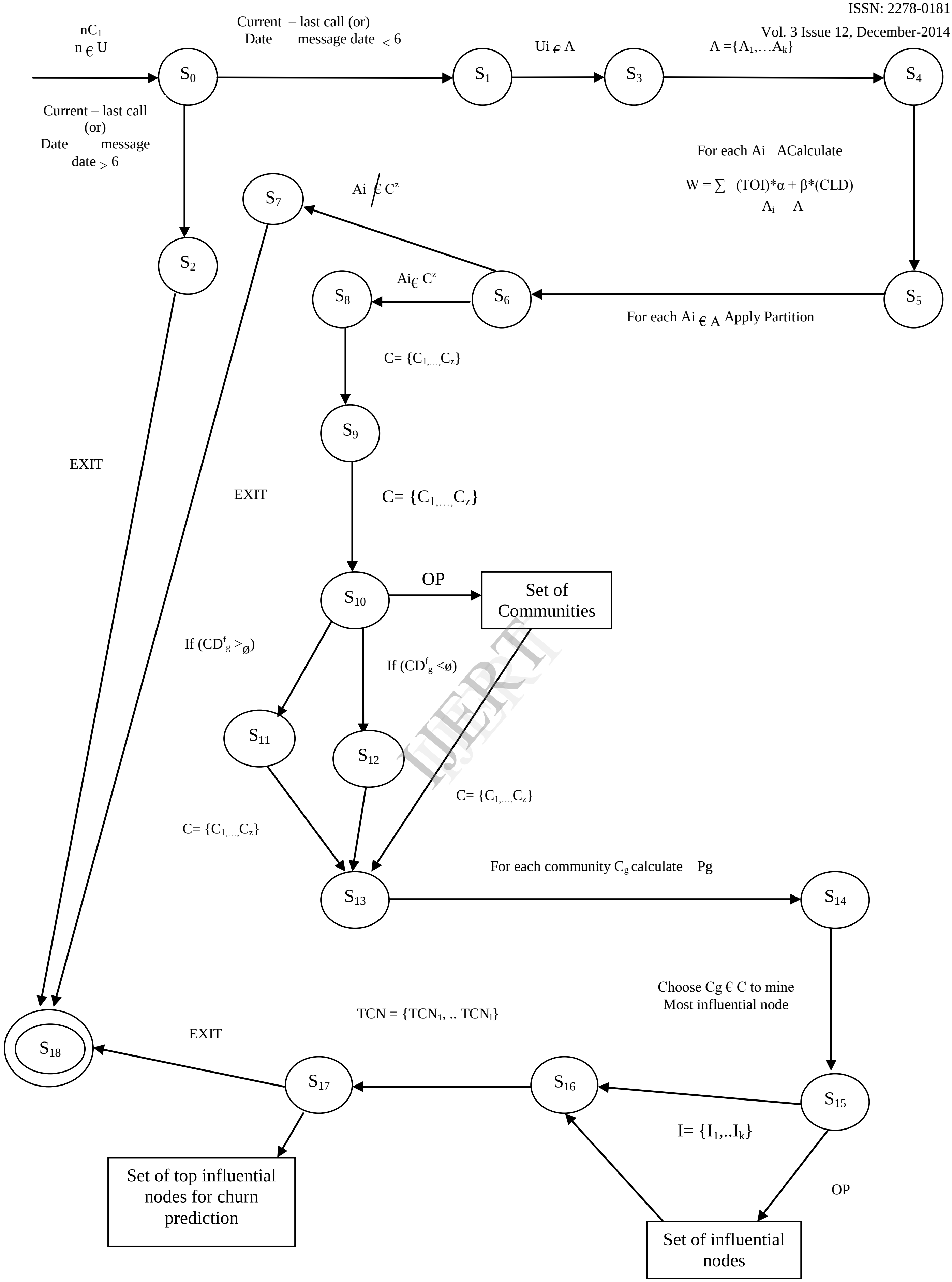

Fig. 2.State Transition Diagram

The set of communities generated passed as input to state S10. If Combination entropy between two groups is more than threshold then the groups are combined and set of communities is passed to state S_{11}.If Combination entropy between two groups is less than threshold ø then the groups are not combined and set of communities is passed as it is to state S_{12}. The set of communities are passed as input to the state S13. In S14 for each community calculate maximal increase in degree of influence. In state S15 influential node is mined from each chosen community and set of influential nodes is generated. The set of influential node is passed as input to the next state S16. Set of top influential nodes for churn prediction is mined and generated as output. Then the set of top influential nodes for churn prediction is generated and the system enters the final state S18.

VI. CONCLUSION AND FUTURE WORK

The proposed system combines the social networking details and call log details of users for predicting the churning customers. The early prediction of churn helps the organizations to design the retention policies. The concept of social network analysis plays an important role for business applications in predicting the churning customers.

In future user's geographical data can be used for churn prediction.

REFERENCES

[1] Chitra Phadke, Huseyin Uzunalioglu, Veena B. Mendiratta, Dan Kushir, and Derek Doran, "Prediction of Subscriber Churn Using Social Network Analysis", Bell Labs Technical Journal, Alcatel-Lucent, Vol. 17, No. 4, 2013, pp 63 – 76

[2] Nittaya Kerdprasop, Phaichayon Kongchai and Kittisak Kerdprasop, "Constraint Mining in Business Intelligence: A Case Study of Customer Churn Prediction", International Journal of Multimedia and Ubiquitous Engineering Vol. 8, No. 3, May, 2013

[3] Richard J. Oentaryo, Ee-Peng Lim, David Lo, Feida Zhu, and Philips K. Prasetyo, "Collective Churn Prediction in Social Network", *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012, pp 210 – 214

[4] Xi Longy, Wenjing Yin, Le An, Haiying Ni, Lixian Huang, Qi Luo, and Yan Chen, "Churn Analysis of Online Social Network Users Using Data Mining Techniques", *Proceedings of the International Multi Conference of Engineers and Computer Scientists 2012 Vol I*, 2012

[5] Blaise Ngonmang, Emmanuel Viennet, and Maurice Tchuente, "Churn prediction in a real online social network using local community analysis", *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*, 2012, pp 282 – 288

[6] Koustuv Dasgupta, Rahul Singh, Balaji Viswanathan, Dipanjan Chakraborty, Sougata Mukhereja, Amit A. Nanavati, and Anupam Joshi, "Social Ties and their Relevance to Churn in Mobile Telecom Networks", EDBT '08 Proceedings of the 11th international conference on Extending database technology: Advances in database technology, 2008, pp 668-677

[7] Francesco Bonchi, Carlos Castillo, Aristides Gionis, and Alejandro Jaimes, "Social Network Analysis and Mining for Business Applications", ACM Transactions on Intelligent Systems and Technology, Vol. 2, No. 3Article 22, 2011, pp22.1-22.37

[8] Thesis, Long Pan, "Effective and Efficient Methodologies for Social Network Analysis", 2007

[9] Caroline Haythornthwaite, "Social Network Analysis: An Approach and Technique for the Study of Information Exchange", 1996

[10] Alex Kosorukoff, Alxeedo, Billlion "Social Network Analysis Theory and Applications", 2011

[11] Guojie Song, Xiabing Zhou, Yu Wang and Kunqing Xie, "Influence Maximization on Large-Scale Mobile Social Network: A Divide-and-Conquer Method", *IEEE Transactions on Parallel and Distributed Systems*, No. 1, *2014, pp 1*

[12] Guolin Niu, *IEEE*, Xiaoguang, Victor O.K. Li, Yi Long, and Kuang Xu, "Multi-source-driven Asynchronous Diffusion Model for Video-Sharing in Online Social Networks", *IEEE Transactions on Multimedia*, Vol.16, No. 7, *2014*, pp 2025 – 2037

[13] Songsong Li, Yuqing Zhu, Deying Li, Donghyun Kim, and Hejiao Huang, "Rumor Restriction in Online Social Networks", IEEE/ *Performance Computing and Communications Conference*, 2013, pp 1-10