Group CRM: a New Telecom CRM Framework from Social Network Perspective

Bin Wu
Beijing University of Posts and Telecommunications
Beijing, China
wubin@bupt.edu.cn

Qi Ye
Beijing University of Posts and Telecommunications
Beijing, China
yeqibupt@gmail.com

Shengqi Yang
Beijing University of Posts and Telecommunications
Beijing, China
sheng_qi.yang@yahoo.com.cn

Bai Wang
Beijing University of Posts and Telecommunications
Beijing, China
wangbai@bupt.edu.cn

ABSTRACT

The structure of customer communication network provides us a natural way to understand customers' relationships. Traditional customer relationship management (CRM) methods focus on various customer profitability models, and they are short of ways to understand the social interactions. Graph mining and social network analysis provide ways to understand the relationships between customers, and there are already a few applications in CRM using these methods. To transform the traditional CRM methods from individuals to social groups, we propose a novel technical framework (GCRM) to manage the social groups in massive telecom call graphs. Our framework is based on a series of newly emerged methods for social network analysis, such as group detecting, group evolution tracking and group life-cycle modeling in telecom applications. We analyze the relationships between social groups and propose a method to find potential customers in these groups. To evaluate GCRM, we present a comprehensive study to explore the group evolutions in real-world massive telecom call graphs. Empirical results show that by taking this framework, analysts can gain deeper insights into the communication patterns of social groups and their evolutionary patterns which makes the management of these social groups much easier in real-world telecom applications.

Categories and Subject Descriptors

H.2.8 [**DATABASE MANAGEMENT**]: Database Applications—*Data mining*; H.4.2 [**INFORMATION SYSTEMS APPLICATIONS**]: Types of Systems—*Decision support(e.g.,MIS)*

General Terms

Management, Measurement

CNIKM'09, November 6, 2009, Hong Kong, China.

Keywords

CRM, Data Mining, Social Network Analysis, Call Graph, Visual Analytics

1. INTRODUCTION

In recent years, there has been an increased focus on Social Network Analysis (SNA) in the data mining community, and Customer Relationship Management (CRM) has been an active field of data mining applications. Integrating the techniques of social network analysis with CRM enables companies to offer better customer service. Telecom industry accumulates huge amounts of data which contain both of the usage behaviors of individuals and social interactions between customers. These available data and methods facilitate a new paradigm for telecom analytical CRM. Some studies have dealt with subjects of social CRM, however, most of them either put emphasis on the operational aspect rather analytical one or touch upon scattered analytical applications instead of a comprehensive framework. In this paper, a novel telecom CRM framework: Group CRM (GCRM) is proposed both from theoretical and practical perspective. In the framework GCRM, we try to manage the social groups (communities) based on a series of newly emerged methods for social network analysis, such as group detecting and group evolution tracking.

Customer segmentation is the practice of dividing customers into groups of individuals which are similar in specific ways relevant to marketing. Group detecting is the first step in GCRM by exploiting customer interaction data. Traditional customer segmentation methods focus on identifying customer groups only based on demographics and attributes, such as age, gender, interests and spending habits, etc. However, GCRM takes another natural way to segment customers into social groups by using community detection algorithms in SNA. Being in a group, customers are prone to attract new ones, retain old ones and accept a new product or service through leveraging mutual influences within a group. In the discipline of CRM, customer life-cycle is a primary term used to describe the steps a customer goes through. Similarly, group customer life-cycle modeling is a crucial step in GCRM. If a group life-cycle can be understood and predicted, a lot of other important questions for group management will be answered, including: how to compare the long-term effects on group customer value of different advertising approaches and product selections or pricing? who is a leader that has the most influence in a group? how a leader plays a role to

keep a group for a longer life? what is the lifetime value of a group compared with other groups and how to increase the group value effectively? In this paper, we not only propose a technical framework of GCRM, but also give case studies of GCRM in real-world telecom applications.

Our unique contributions are:

- We propose a novel telecom CRM framework—Group CRM (GCRM). By integrating the methods of graph mining and social network analysis, we will show the architecture, workflow, analysis methods of the framework.
- We propose the algorithms of GCRM in discipline of social network analysis in details. We also propose the methods of life-cycle modeling in telecom domain.
- Based on the framework, we analyze social groups in several massive telecom call graphs. We propose real-world applications to track the evolution of customer groups and try to explore the relationships between the group sizes and their stability in telecom call graphs. Based on the results of group analysis in the massive call graphs, we can find out the potential customers in other telecom companies.

This paper is organized as follows: Section 2 surveys the related work. The framework of GCRM is described in Section 3 and case studies and new algorithms in GCRM are presented in Section 4. Section 5 discusses our experiences. We conclude in Section 6 outlining the impact of GCRM.

2. RELATED WORK

Since the early 1980s, the concept of customer relationship management in marketing area has gained its importance. Many researches tend to develop comprehensive models of customer profitability to find 'Who are profitable customers?'. However, most of them focus on the cash flow derived from the past profit contribution. In order to attract and retain high-value customers, data mining and machine learning techniques have also been proposed in CRM for predicting survival possibility of customers [9, 19]. While the traditional Customer Lifetime Value methods focus on individuals, in our framework we will take a deep insight into the life-cycle of different customer groups in the call graphs. There are even some famous companies integrating the notion of social network analysis in their CRM tools, such as Oracle [1], IBM [8] and Xtract[2]. Xtract Social Links turns raw customer data into a vital marketing tool for mobile operators. By analyzing the social networks within large-scale mobile communication networks it identifies the underlying social network structures and the most influential people in the network, which Xtract calls Alpha Users. Processing such large-scale data and deriving business insights from it has become a very important issue in customer relationship management (CRM). However, unlike most of the current CRM tools which are based on various customer profitability models in different scenarios, GCRM focuses on analyzing the social networks retrieved from the daily communications records in telecom domain.

Recently, there has been considerable interest in analyzing real-world call graphs. Nanavati et al. [10, 11] study a broad set of structural properties in massive telecom call graphs obtained from call detail records. They report findings on various topological properties of these massive call graphs, including degree distributions, strongly connected components and bipartite cores. Onnela et al. [14] examine the communications patterns of millions

[1](http://sales.oracle.com/en-us)
[2](http://www.xtract.com)

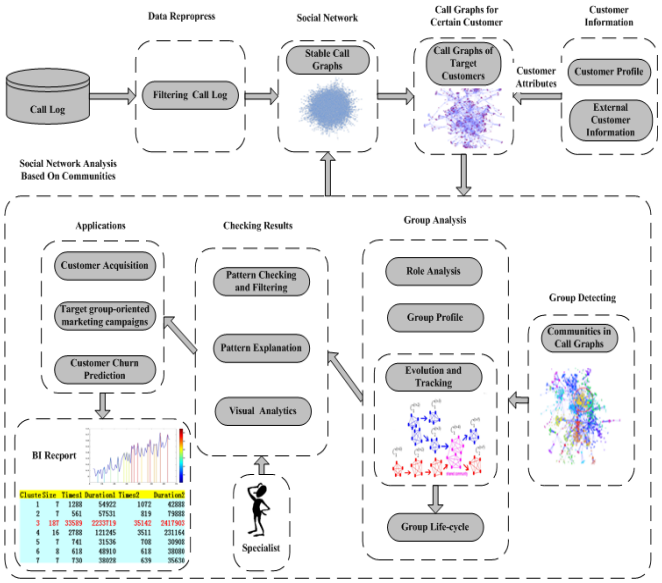


Figure 1: Architecture of Group CRM framework in telecom practices

of mobile phone users and observe a coupling between interaction strengths and network's local structure. They find that the mobile network is robust to the removal of the strong ties but fall apart after a phase transition if weak ties are removed. The goal of community detecting in social network analysis is to divide large data set of social networks into a number of sub-sets, called clusters or communities, and many community detecting algorithms have been proposed. Girvan and Newman [5] have introduced a divisive algorithm where the selection of the edges to be cut is based on the value of their "edge betweenness". They [12, 13] also propose a detection scheme based on global optimization of the modularity Q. However, the modularity-based community (group) detection algorithms have a resolution threshold such that small communities in a large network are invisible [4]. Palla et al. [16, 15] propose the CPM algorithm to find the overlapping communities in networks.

3. THE FRAMEWORK OF GCRM

In this section, we will show the architecture of this framework according to the typical analysis scenario in details. In Fig. 1, the architecture of GCRM follows a typical analysis scenario of group analysis in telecom domain. In the scenario, analysts get the raw data from the call log in the database. After that they can get the social networks from the raw data which may just contains the social network of certain target customers. Using different analytical algorithms in graph mining and social network analysis, analysts can get different communication patterns in social networks. Finally, the analysts and domain experts can check whether the patterns in social network are interesting and then choose these interesting patterns in the BI reports.

3.1 Social Network Extracting

The GCRM framework starts the analytical process from extracting the raw graph data structure in the call log. Retrieving information from call graphs (where people are nodes and calls are edges) obtained from the Call Detail Records (CDR) can provide major business insights to telecom providers for designing effective strate-

gies [10]. A call graph G is a pair $\langle V, E \rangle$, where V is a finite set of nodes (phone numbers), and E is a finite set of node-pairs from V(phone calls). So if user u calls user v, then an edge $\langle u, v \rangle$ is said to exist in E. The edges are undirected in this paper.

In some applications, we are only interested in the groups of some target customers, and it is convenient to explore the communities in the subgraph of certain target customers. By exploring the egocentric networks of a given set of target customers, we can get more clear knowledge on how these customers are linked to neighbors and how the neighbors communicate with each other. Egocentric network helps users to learn about how people correspond with their social networks [3]. By integrating these egocentric networks of target customers, we can explore the communication patterns of these target customers. To facilitate the following analysis, analysts can import other customer information to describe the nodes in the call network, such as names, addresses, subscribed value-added service information, etc.

3.2 Group Evolutionary Pattern Detecting

3.2.1 Group Detecting

One of the most important properties of the social network is its community (group) structure which stands for a dense connected user groups in the nature of some family-like, colleague-like or interest-like relationships. There has been extensive research work on group detection by means of statistics and heuristics. However, they are generally blind to the following issues: how can we effectively extract the group structure with a user acceptable time costing from a graph of millions nodes and edges? Another important question comes to be although there are already many community detecting algorithms, many of them become impractical for large networks [7, 4]. How can we tell whether the communities detected by one method or another are truly robust? Clear answers to these questions are critical to the effective of our framework. From the industrial perspective, we implement serval prevalent clustering method and incorporate them into our framework. These methods include *GN* [5], *CNM* [2] and *CPM* [16], etc.

3.2.2 Group Analysis

Group Profile and Role Analysis.

In many cases, communities have been found to correspond to behavioral or functional units within networks [7]. This finding suggests that we may be able to gain deeper insights into call graphs by examining the common properties in these communities, and we call the common properties to describe these communities as the profiles of them. Although it is impractical to verify the real group structure in real-world massive call graphs, it is believed that the members within a group show the homogeneity of age, hobby (subscribed value-added service) or the mobility. Currently in our framework, we simply use the customer profiles to describe social groups. In the research field of social networks there is a long tradition of developing quantitative methods to get the importance of actors. A variety of measures have been proposed to determine the "centrality" of an actor in a social network [20, 1]. Based on the result of community detecting, there are some work try to classify nodes into different roles according to their intra- and inter-community connections [6].

Group Evolution Tracking and Life-Cycle Modeling.

Understanding the evolution of social network is helpful in inferring trends and patterns of social contracts in particular social context [17]. Group life-cycle also is a crucial step in GCRM to get the steps a group goes through. To understand the life-cycles of groups, we must first track the evolutionary processes of them and then get better understanding of their communication patterns. Group evolution is a critical issue to supply business insights for designing strategies, such as the early-warning of group churn and finding the potential groups. Recently there are some methods of group evolution tracking has already been proposed [15, 8]. By tracking the evolutionary processes of groups, people can get better understanding of the life-circle of different social groups and we can manage them more efficiently. So in this paper, we will focus on the characteristics of temporal social group and try to find out the evolutionary patterns. With the help of GCRM users could trace both the static communication relationships in communities and their communication evolution trends in the call graphs. To track the evolutionary groups, we need to solve the following problem: how to match the communities at different steps with the underlaying social groups? We propose a new method to identify the correlation of communities between different time steps. More details will be discussed in section 4.

One of the primary functions of CRM is to collect information about customers to track life-cycle revenues, costs, margins, and interactions between individual customers [18]. However, traditional methods or tools mainly focus on the profile of individuals that may be reflected in the business process. In GCRM framework, we try to propose a life-cycle model to illustrate the evolutionary processes of different communities. After tracking the evolutionary groups, we will first clustering the communities based on their evolutionary patterns. After clustering, we will zoom into the group clusters to explore their evolution patterns in more details and try to description their different evolutionary trends.

3.3 Pattern Checking

3.3.1 Pattern Checking and Interpretation

One greatest challenge is how to show the results of patterns in manner understandable by various specialists in telecom service. For illustration, some real comments from the specialists of some telecom companies are:

- How to measure the accuracy of patterns that we find in the call graphs. For example, the telecom domain exports many use the customers information to check whether the communities of target customers are the ones that they are interested.
- Whether the communities we detect are the social groups we are interested. The exports in the telecom company may just want to find out the social groups in some companies. They also want to get the descriptions of these communities in more detail from the customer profiles. Large communities in the call graphs who are mainly formed by our employees may not be the target customers who we are interested in.
- Specialists may not understand certain customers' communication patterns and want to investigate their communication patterns.

To overcome these challenges in real-world telecom applications, we must integrate the knowledge of domain experts. That is another reason we highlight the importance of interactivity in GCRM.

3.3.2 Visual Analytics

Visual data mining is an effective way to discover knowledge from huge amounts of data. To interpret the patterns to telecom domain exports, patterns at various granularity are expected through

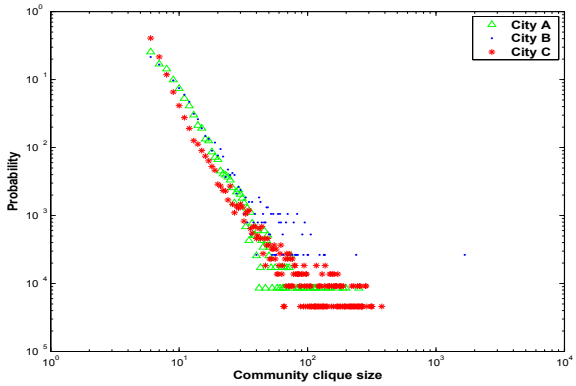


Figure 2: Distributions of communities which are detected by *FLIBer* algorithm in different call graphs.

showing the relationships between customers. To overcome the difficulties of showing the large-scale network, we adopt two ways. First, an egocentric network approach is taken to show the subgraph of certain important customer to know the customer's local call graph. Second, we use group graph in which nodes are groups to gain a coarse overview of massive call graphs. In this paper, we developed a tool called *TeleComVis* to analyze call graphs based on the visual analytical framework for analyzing real-world network called JSNVA [21].

4. CHARACTERIZING GROUP EVOLUTION

In this section, we will propose a group life-cycle model to track the evolution of different social groups to demonstrate our framework by employing several real-world data sets. Our purpose is to provide prominent characteristics of massive call graphs while examining our framework.

4.1 Data sets

In this paper, we consider 3 data sets obtained from 3 different cities in China. Table 1 summarizes the basic information of these data sets.

CDR 1 comes from a major mobile operator in city A which containing 6 months full records. These records retained the basic information of each call interaction, such as call time, duration, customer identity and involved base stations.

CDR 2 is obtained from city B which spanning 10 months. These records only reserved the topological structure of the call graph.

CDR 3 is a 8 month-long data set from a fixed line telecom operator which is coming to supply mobile service in city C. Different from mobile network, fixed line call graph is a relatively stable network.

We find that mobile call network generally form relatively larger maximal cliques than the fixed line call graph even if the network is small (as CDR 1). This may respond to the stability of fixed line call graph while the un-stability of mobile graph.

4.2 Group Detection

For the sake of customer relationship management, telecom service providers may want to identify the family, friendship and fellow groups in the call graphs, so one may belong to different groups and overlapping community detecting algorithm is very useful in the telecom service. However, we find *CPM* is very slow in such large-scale call graphs and may yield too many "cliques chains" especially when the k is small. To avoid such deficiencies, we devised

Algorithm 1 *FLIBer*(C, t)

$\mathbb{U} \Leftarrow \phi$
for $C_i \in C$ **do**
  $\{K\} \Leftarrow C_j$ that is adjacent to C_i
  for $C_j \in \{K\}$ **do**
    if $\frac{C_i \cap C_j}{min(C_i, C_j)} > t$ **then**
      merge C_j into C_i
      $\{K\} - C_j$
      $\{C\} - C_j$
      update $\{K\}$ with the adjacent clique of C_j
    end if
  end for
  $\{C\} - C_i$
  $C_i \rightarrow \mathbb{U}$
end for
return $\mathbb{U}$

an algorithm called *FLIBer* based on the merging process of maximal cliques, which is presented as Algorithm 1. Our optimization promote the efficiency of the algorithm both in the space requirement and the time costing.

Our following analysis is all based on the communities obtained after this process. Fig. 2 shows the basic statistic of group size, and all the group sizes in different cities have similar "power-law" distributions. The last two columns in Table 1 represents the time costing of *CPM* and *FLIBer*. To make a comparison, we only count their time for clique merging with parameter $k = 4$ and $t = 0.8$ respectively on the Windows XP Operating System on a commodity PC with P4 3.0GHz, 2G RAM.

4.3 Group Evolution Tracking

To track the group evolution, we need to solve the following problem: how to match the groups at current step with the underlaying social groups? To solve this problem, we have to answer two questions. The first one is how to match an evolutionary group in consecutive time steps t and $t + 1$. The second problem is that the members in groups may have different calling-rate which may cause the structure of temporal groups challenging dramatically. To overcome these problems, we regard that the groups found in the cumulative network formed from time step t_1 to t_{max} as the underlying social groups. So we can define the node activity $A_n(C, t)$ of a group C to measure the fraction of active nodes in underlying social group C at time step t by the following equation:

$$A_n(C, t) = |V(C, t)|/|C| \tag{1}$$

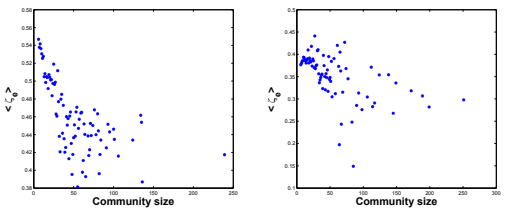
where $V(C, t)$ denotes the set of active nodes in group C at time step t. We define node v where $v \in C$ as an active node in C at time step t if there is at least a link between it and other nodes in C at time step t. We also define the edge activity $A_e(C, t)$ of a group C to measure the fraction of active edges in the underlying social group C at time step t:

$$A_e(C, t) = |E(C, t)|/|C| \tag{2}$$

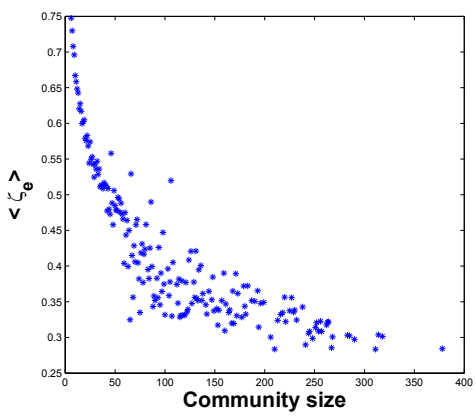
where $E(C, t)$ denotes the set of edges in group C appeared at time step t. Then we can get a vector $A_e(C) = (A_e(C, t_1), \cdots, A_e(C, t_{max}))$ in sequence from time step t_1 to t_{max} for group C. To quantify the stationary of each group evolution, we define the node stationary $\zeta_n(C)$ and the edge stationary $\zeta_e(C)$ of a social group C as the mean of node and edge activity of group C during the time steps:

Table 1: Data sets

name	months	nodes	edges	cliques(>= 3)	$N_{FLIBer}(k = 6, t = 0.8)$	T_{CPM}	T_{FLIBer}
CDR 1 (Mobile)	6	298k	1,659k	129,118	8,102	140s	5s
CDR 2 (Mobile)	10	1,632k	5,677k	1,320,700	51,845	1.5h	3.0m
CDR 3 (Fixed Line)	8	1,188k	9,251k	5,821,917	160,033	>24h	5.0m



(a) City A (b) City B



(c) City C

Figure 3: Group size and edge stability in temporal call graphs

$$\zeta_n(C) = \frac{\sum_{t=t_1}^{t_{max}} A_n(C,t)}{t_{max}} \quad (3)$$

$$\zeta_e(C) = \frac{\sum_{t=t_1}^{t_{max}} A_e(C,t)}{t_{max}} \quad (4)$$

By using these metrics, we first find most of the nodes in these social groups remain relatively stable that is, in each month, most of these group customers have some links with others in the same group. The average group node stationary $< \zeta_n(C) >$ of city A, B and C are 0.930, 0.826 and 0.974, respectively. This result shows the nodes in these communities are very stable which also indicates the correctness of our community detecting algorithm *FLIBer* from another respect. We also observe some interesting patterns in the temporal call graphs when investigating the relationship between the mean values of group edge stationary $< \zeta_e(C) >$ and the group size. As shown in Fig. 3, we find small groups are more stable than larger ones. We can find the phenomenon more obvious in fixed line call graphs than in mobile call graphs. This result may indicate that customer in larger groups are prone to communicate with each other by mobile phones than with fixed line phones.

4.4 Group Evolution Clustering and Life-Cycle Modeling

Be known as CRM with the factor of time, CLM (Customer Life-cycle Management) provides metrics of CRM. Traditional method model CLM into 5 parts, e.g. acquisition, build up, climax, decline and exit. Our focus on social network drives our interest into question whether the group structure has the same evolution pattern.

4.4.1 Group Evolution Clustering

Since its birth, one of the primary functions of CRM is to collect information about customers to track life-cycle revenues, costs, margins, and interactions between individual customers [18]. However, traditional methods or softwares mainly focus on the profile of individuals that may be reflected in the business process. To monitor and quantify the evolution of customer groups, we define following metrics at sequential time steps. In order to get an overview of the evolution of different groups, it is necessary to divide the evolution groups into clusters by their communication patterns—vectors $A_e(C)$. To find out the similarity between two vectors $A_e(C_i)$ and $A_e(C_j)$, the **Pearson's Correlation** coefficient between two vectors, is defined by the following equations:

$$\rho(A_e(C_i), A_e(C_j)) = \frac{cov(A_e(C_i), A_e(C_j))}{sd_{A_e(C_i)} sd_{A_e(C_j)}} \quad (5)$$

where $cov(A_e(C_i), A_e(C_j))$ is the covariance of $A_e(C_i)$ and $A_e(C_j)$; $sd_{A_e(C_i)}$ and $sd_{A_e(C_j)}$ is the standard deviation of $A_e(C_i)$ and $A_e(C_j)$, respectively. We then calculate the similarities between every pair of groups by using the equation 5. We set a similarity threshold S form a link between any two groups C_i and C_j if their are similarity that is $\rho(A_e(C_i), A_e(C_j)) >= S$. After the group evolution similarity graph has formed, we can use different graph clustering algorithms to find out clusters of groups that have similar evolutionary trends. In this case study, we set the similarity threshold $s = 0.98$ to form the similarity links between groups and use the GN algorithm to find groups clusters which have similarity evolution trends. The the maximal modularity value $Q = 0.92$, which indicates strong cluster structure in this similarity graph. Fig. 4 shows the largest 10 clusters of groups which have similarity evolution trends in the the call graph of city C.

4.4.2 Group Growth and Contraction

Due to the limit of the observed time length, it is difficult to quantify the whole life-cycle of groups. For simplicity, we focus on 2 patterns for each evolving group which are growth and contraction. To show temporal links, we define 3 types of edges: the persistent edges which appear both in previous and current time steps, the vanished edges which appear in previous time step but disappear in current time step and new born edges which do not appear in previous time step but appear in current time step. The grey edges indicate the persistent edges, and the vanished edges are in red, and the new born edges are in green. Fig. 5 shows the edge activity vectors A_e of clusters of groups which have same evolutionary patterns. Fig. 5(a) and Fig. 5(c) show the values of A_e of

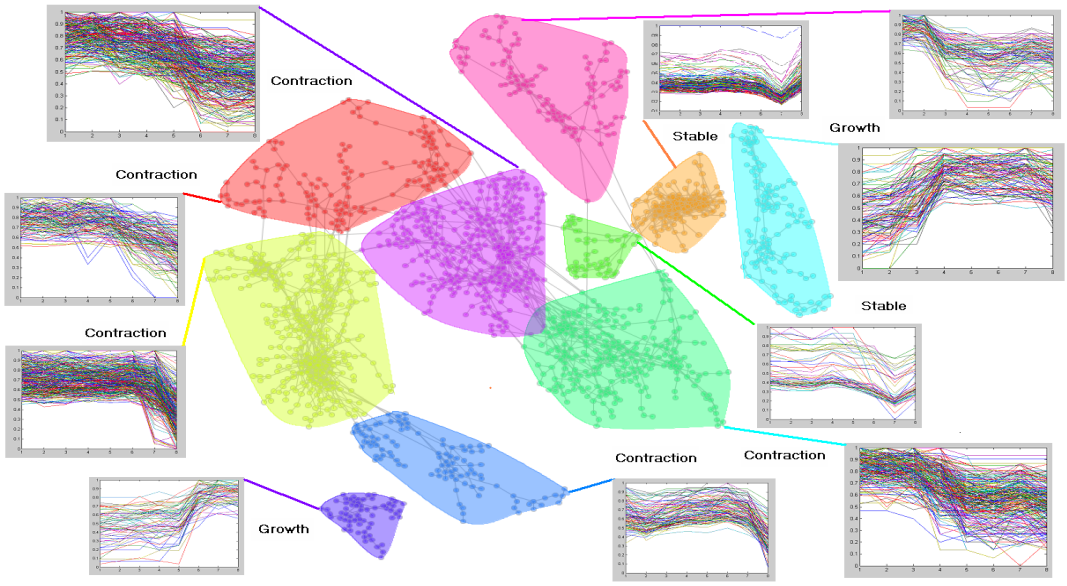
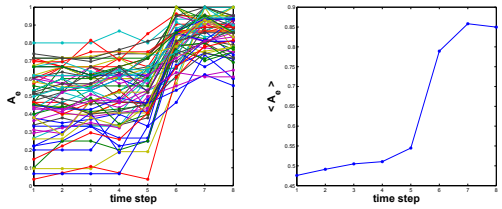
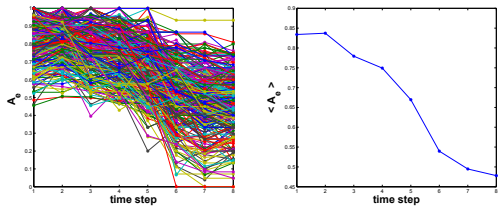


Figure 4: The largest 10 clusters of groups which have similar evolutionary trends in the temporal call graph of City C.



(a) The values of A_e of a growth group cluster

(b) The mean values of A_e of a growth group cluster



(c) The values of A_e of a contraction group cluster

(d) The mean values of A_e of a contraction group cluster

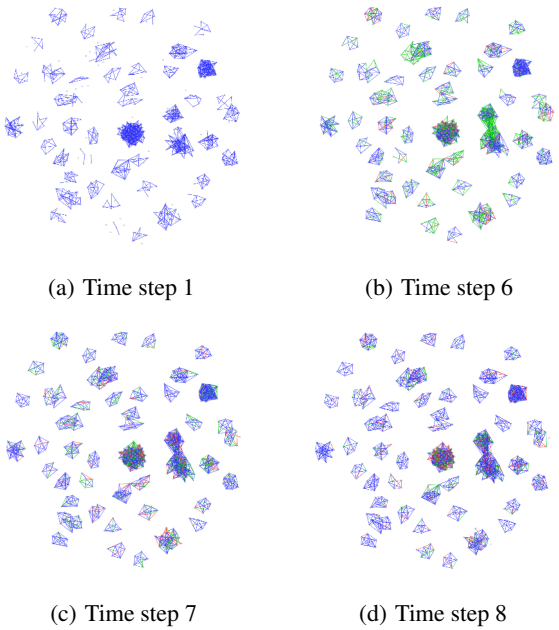
Figure 5: The edge stability A_e of evolution group clusters.

a cluster of growing groups and a cluster of contracting groups, respectively. To show their growing pattern more clearly, we get the mean value of A_e which are shown in Fig. 5(b) and Fig. 5(d), respectively. To measure the accuracy of patterns, we will show the evolution process of the growing groups and the contracting groups visually. Fig. 6 shows the growing processes of the cluster of groups shown in Fig. 5(a). From Fig. 5(b) we can find there are lots of new born edges which are green in the group cluster at time step 6, and we can check this pattern in Fig. 6. To make the group contracting process more clearly, we focus on the evolution process of a certain group in contracting group cluster in Fig. 5(c). Fig. 7 shows the contracting process of this group in the consecutive 8 time steps. Telecom service providers should pay more attention to this kind of groups and try to prevent them to leave the call graph.

5. DISCUSSION

In this section, we share our experience of designing, building and developing a framework to analyzing social groups in the telecom applications. Groups are the basis of Group CRM and the quality of community detecting will greatly influence the following analysis results. Nowadays, there are many community detecting algorithms. In such large-scale call graphs, it is unnecessary to divide each customer into a group and we only interest to find the social groups that are statistically significant. To get the significant social groups, we propose our *FLIBer* algorithm to connect these highly overlapping cliques to find the cohesive groups in massive call graphs. *FLIBer* can produce similar results as *CPM* but cost a significantly less time. In the telecom domain, group of different sizes many response different relationships such as family relationship groups, friendship groups, leisure-based relationship groups, service-based relationship groups and working relationship groups. How to identify these different types of groups is still a challenge in the future. In our application, we find large groups are prone to be working relationship groups by checking the profiles of the customers. How to get the small groups precisely and provide group-oriented marketing strategies for these groups is still a challenge. In order to understand the life-cycle of different social groups, we propose the algorithm for tracking group evolution. To track group evolution, based on the metric of edge activity A_e, we can find the clusters of growing or contracting groups. There is another important question raised: how can we use the metric $A_e(C)$ to predict the future of group C? How to predict the future of certain group and get an early warning of the churn of groups is still a challenge.

We highlight the importance of interactivity in real-world tele-



(a) Time step 1 (b) Time step 6

(c) Time step 7 (d) Time step 8

Figure 6: The growing process of a cluster of groups in time step 1, 6, 7 and 8 in the call graph of city C.

com applications. Visualization can give us more hints of other underlying patterns in the mobile call networks and can help us eliminate unrealistic or useless patterns. As our understanding of real-world networks improves, patterns may be meaning distinguished based on as-yet-undiscovered properties of real-world networks. To attract and retain high-value customers, analysts need to know the social groups of certain high-value customers. In following part, we will show how to analyze the social groups in call graph given a list of target customer names. We will analyze the groups of the textile industry in city C using the CDR 3 and the public data—the Yellow Page of the city. We want to detect the social groups of target phone numbers and find out potential customers in other companies.

5.1 Communities in Target Call Graph

In the CDR 3, we can get call records between the customers in the telecom company. The data set contains both of the intra- and inter-net call records not only the records between the customers in this company. To find out the social groups in the textile industry in the city, we first get the phone numbers of the textile industry from the Yellow Page in the city. By extracting the egocentric networks of these target customers, we can find communication relationships of the target phones. We integrate these egocentric networks together to form a call graph of the textile industry. There is about thousands of nodes and tens of thousands of edges in the target call graph.

5.2 Group based Visualization

Even the target call graph is much smaller than the raw one, we find it is very dense and it is hard to find patterns just through visualization. We then use *CPM* algorithm to find the stable dense 5-clique communities in the textile industry call graph. In Fig. 8, the nodes are customer groups, and the edges are the calls between the customers in these groups. The size of nodes indicates the size

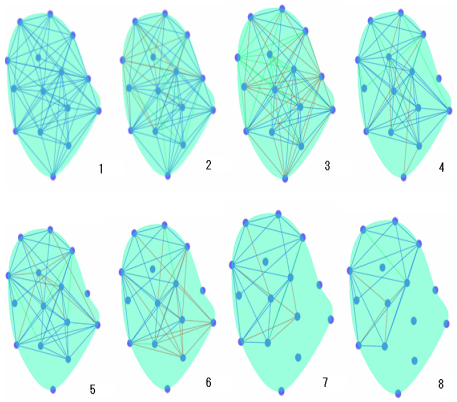


Figure 7: The contraction process of a group from time step 1 to time step 8 in the call graph of city C.

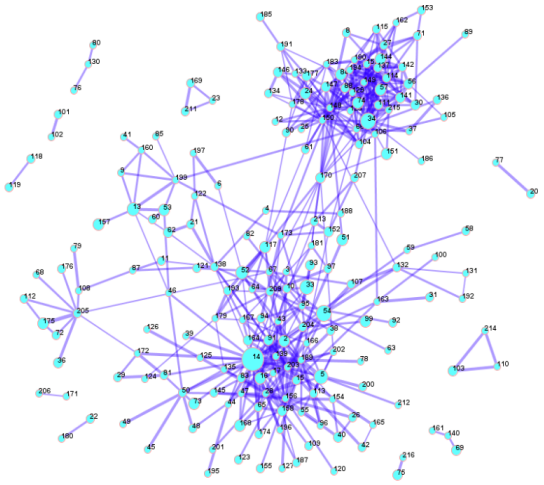


Figure 8: Links between groups of the textile industry call graph of city C.

of groups and the thickness of edges indicates the number of calls between them. As shown in Fig. 8, we can find there are two clusters of groups. By checking the profiles of customers in these social groups, we find one cluster of groups is in the domain of garment industry and the other cluster of groups is related in the domain of wool textile industry. There are groups of other related occupations around these two clusters of groups such as department stores, equipment firms and accounting firms. We also use the company names in the Yellow Page to describe the social groups. By combining the customer profiles with these company names, we find most of the social groups should be working relationship groups. We regard the customers in other telecom companies as the potential customers in these social groups. Based on the profiles of different groups, analysts can infer what kind of people the customer communicate with and what kind of service the potential customer need and which makes the customer acquisition much easier. Using the customer profiles, analysts can visualize the groups with evidence. Checking the patterns interactively will develop analysts' confidence when they use these patterns.

6. CONCLUSIONS

Graph mining, link analysis and social network analysis are useful to capture the topological and communication patterns of social data sets in telecom domain. To transform the traditional CRM methods from individuals to social groups, we propose a technical framework (GCRM) to manage the communities (groups) in massive telecom call graphs. We show its architecture, analysis methods in detail. Our framework is based on a series of newly emerged methods for social network analysis, such as group detecting, group evolution tracking and group life-cycle modeling, etc. We highlight the importance of interactivity in this framework and analysts can check the patterns found visually. In the case studies of real-world telecom applications, we also propose the methods to track the evolution of different social groups. We also apply the technique of visual analytics to analyze the relationships between social groups in target call graph to improve the interactivity of GCRM. Analysts can use the methods in GCRM to manage social groups and check their results visually in the telecom applications.

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the supports of the National Natural Science Foundation of China under Grant 60402011 and the National Key Technology R&D Program of China under Grant No.2006BAH03B05.

8. REFERENCES

[1] S. Brin and L. Page. The anatomy of large-scale hypertextual web search engine. In *Computer Networks and ISDN Systems*, volume 30, pages 107–117, 1998.

[2] A. Clauset, M. E. J. Newman, and C. Moore. Finding community structure in very large networks. *Physical Review E*, 70(6):066111, December 2004.

[3] D. Fisher. Using egocentric networks to understand communication. *IEEE Internet Computing*, 9(5):20–28, 2005.

[4] S. Fortunato and M. Barthélemy. Resolution limit in community detection. *Proc. Natl. Acad. Sci.*, 104:36–41, 2007.

[5] M. Girvan and M. E. J. Newman. Community structure in social and biological networks. *Proc. Natl. Acad. Sci.*, (12):7821–7826, June 2002.

[6] R. Guimerà and L. A. N. Amaral. Functional cartography of complex metabolic networks. *Nature*, 433(7028):895–900, February 2005.

[7] B. Karrer, E. Levina, and M. E. J. Newman. Robustness of community structure in networks. *Phys. Rev. E*, 77:046119, 2008.

[8] Y. Li, B. Liu, S. Sarawagi, and S. Mehta. A visual-analytic toolkit for dynamic interaction graphs. In *Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pages 1016–1024, 2008.

[9] D. R. Mani, J. Drew, A. Betz, and P. Datta. Statistics and data mining techniques for lifetime value modeling. In *ACM SIGKDD international conference on Knowledge discovery in data mining*, pages 94–103, 1999.

[10] A. A. Nanavati, S. Gurumurthy, and et al. On the structural properties of massive telecom call graphs: Finding and implication. In *Proceedings of ACM Conference on Knowledge and Information Management (CIKM)*, pages 435–444, 2006.

[11] A. A. Nanavati, R. Singh, and et al. Analyzing the structure and evolution of massive telecom graphs. *IEEE Transactions on Knowledge and Data Engineering*, 20(5):703–718, May 2008.

[12] M. E. J. Newman. Modularity and community structure in networks. *pnas*, (103):8577–8582, 2006.

[13] M. E. J. Newman and M. Girvan. Finding and evaluating community structure in networks. *Phys. Rev. E*, 69:026113, 2004.

[14] J.-P. Onnela, J. Saramäki, and et al. Structure and tie strengths in mobile communication networks. *PNAS*, 104(18):7332–7336, 2007.

[15] G. Palla, A.-L. Barabasi, and T. Vicsek. Quantifying social group evolution. *Nature*, 446(7136):664–667, April 2007.

[16] G. Palla, I. Derényi, I. Farkas, and T. Vicsek. Uncovering the overlapping community structure of complex networks in nature and society. *Nature*, 435(7043):814–818, June 2005.

[17] V. Pandit, N. Modani, and et al. Extract dnese communities from telecom call graphs. In *Internation Conference on Communication Systems Software and Middleware and Workshops*, pages 82–89, 2008.

[18] J. S. T. Robert C. Blattberg, Gary Getz. *Customer Equity: Building and Managing Relationships as Valuable Assets*. Harvard Business Press, Watertown, MA, 2001.

[19] X. Song and C.-Y. L. ant et al. Modeling and predicting personal information dissemination behavior. In *SIGKDD*, pages 479–488, 2005.

[20] S. White and P. Smyth. Algorithms for estimating relative importance in networks. In *Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 266–275. ACM Press, 2003.

[21] Q. Ye, T. Zhu, D. Hu, B. Wu, and N. Du. Cell phone mini challenge award: Social network accuracy—exploring temporal communication in mobile call graphs. In *IEEE International Symposium on Visual Analytics Science and Technology*, pages 207–208, Columbus, USA, 2008.