Formation of Triads in Mobile Telecom Networks

N Naren Krishna
Ericsson India Private Limited
Ericsson R&D, Chennai, India
naren.krishna@ericsson.com

M Saravanan
Ericsson India Private Limited
Ericsson R&D, Chennai, India
saravanan.r.mohan@ericsson.com

Abstract— **In the present competitive telecom scenario, an intention of any operator is to increase the size of the network and establish more connectivity between its users. This can be achieved either by adding new customers to the network or by increasing number of links in the network. In this paper, we present a method called triad formation for increasing the connectivity in the network. Community detection has been done on the network before to set up the triads for efficient results. The communities formed are based on the modularity factor. The proposed method will resolve possible new combination of edges between nodes which are not connected earlier and it has a strong connection with a common node. The effectiveness of the triad formation is demonstrated on a huge telecom data and its importance is highlighted.**

Keywords: Call Detail Record (CDR); Modularity; Isolated Community; Triads; Mobile Social Network Analysis(MSNA).

I. INTRODUCTION

A social network is a compact structure which explores connected groups and it is used to predict the actions of individuals. These individuals in a network are called nodes and their communications are measured in the form of dependent edges. Dependency varies from friendship, common interests, beliefs and knowledge. One of the utmost interests of telecom operators is to increase the connectivity between customers to generate more profit. For this, they need to identify the existence of strong connections and influential members based on the mobile usage services. Mobile Social Network Analysis (MSNA) is an upcoming research area which indicates the importance of identifying the social groups in mobile networks [1]. It helps the operators in understanding and analysing the subscribers and increases the focus on their business. MSNA has proved to give extensive results in the areas like churn prediction, customer retention and campaign management [2].

The behaviour of highly connected customers and their relationship mainly depends on the social structure of the communities. With the help of social network analysis it is possible to find the potential users, influential members and weak users of a community [3]. The real challenge in telecom networks is it dimension. Processing millions of nodes and billions of edges is not only a tough task, but also a time taking process. One of the solutions to this challenge would be detecting communities and then performing analysis on individual communities which will derive efficient results. Community detection can be done based on common interests of the subscribers and their calling patterns. There are several interesting algorithms for community detection available which can segment the network based on modularity [3, 4, 5]. Modularity is a prime factor which defines the strength of a community. CNM algorithm is one of its kind which helps in finding communities using a hierarchical agglomeration algorithm for detecting community structure which is faster than many competing algorithms [4]. CNM algorithm works on undirected edges in a network. Hence, whenever it runs through a cycle in the network, the algorithm runs into an infinite loop and leads to a memory constraint and stops working. The algorithm has been slightly modified to solve this problem.

We propose a novel idea on new group formation called triads, which finds nodes which were not connected earlier and where there is a high probability of forming new edges between them because they have a strong common node. By doing this, the number of edges (number of calls) increase which in turn leads to increase in revenue for the operator. To the best of our knowledge, this is the first attempt to generate new edges in a telecom network at a community level. The main contribution of our research is based on the telecom data obtained from the Call Data Records (CDRs) which are used for community detection and formation of triads. The communities which are formed with higher modularity help the operator to understand its customers as a group and it is easy to analyse the behaviour of potential customers.

The first and foremost step is to clean the data to avoid the redundant and unnecessary attributes. The attributes of CDRs which are taken into consideration are Calling Number, Called Number and Duration of the call. This cleansed data will be used as input for the modified CNM algorithm which will detect communities. Once the communities with high modularity are detected, the task of finding nodes which have high probability in making new edges is easy which will lead to the formation of triadic closure [5] in mobile networks. A brief stepwise description of our research work is as follows

A. Data Pre-processing

First and foremost step is to clean data or in other terms filter out those set of attributes from CDRs which are not used in the next step of our research. In a telecom scenario for every call made, a CDR is generated. Records related to the calls made to toll free numbers and call centres should be filtered out. Records indicating voice calls are collected. Our study on triadic formation is based only on voice calls.

B. Classification of users

To enrich the results and to achieve a better accuracy, it is suggested to classify users into weak and potential users. It is a challenging task for an operator to find out its active or potential users who add weight to their network. Such classification rules would solve the problem and does make

 ISBN:978-1-61208-115-1

it easy to identify related users and in order helps to understand their behaviour and also the behaviour of the community the user is related to. It is at the ease of operator to customize the rules and classify the users based on his need.

C. Community Detection

Communities are formed based on the increase or gain in the modularity when nodes merge [6]. This is a recursive step and stops when there is no more gain in the modularity. Unlike other greedy algorithms, data structures have been used to handle sparse adjacency matrices.

D. Identifying Isolated Communities

Communities of a network which do not have any edge or connection to other communities are called isolated communities. The users of these communities do not have any edges or connections to other community members. The total number of inter-community calls is zero. Communities which have more than one edge or connection to other communities are called non-isolated communities. The need for identifying isolated communities drills down the problem of understanding the customers, as the behaviour of an isolated community depends on the behaviour of the alpha user of that community. This is the major advantage of isolated communities. An alpha user is a highly potential node who influences the community [1] and the community behaviour can be concluded based on the alpha user's behaviour and the time taken to spread the information to the network from this point is minimal. Another major advantage of an isolated community is that the community does not get disturbed by external factors.

E. Generation of New Edges

This is the final step of our research work. Generating new edges between nodes which are not connected and which have a strong common node.

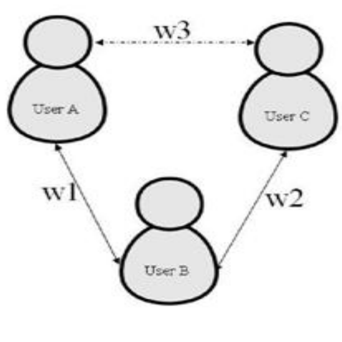

Figure 1 Triad formation

This step is called triad formation which is represented in Fig. 1. It shows three users A, B and C where User B is the common node. The edge AB whose edge weight is represented by w1 and edge BC whose edge weight is represented by w2 are the existing edges and the dotted edge AC denote the newly generated edge. The edge weight w3 of edge AC is generated.

The rest of the paper is organized as follows. In Section II, we review the related literatures and in Section III, we describe the procedure of experiment at different stages. Section IV gives our experiment results followed by discussion in Section V and concluding remarks and future work in Section VI.

II. RELATED WORK

Understanding the customers and making them feel comfortable without creating any disruption in the network is considered to be the top most priority of any network provider. Mobile Social Network Analysis (MSNA) is an evolving research area where many challenges like finding influential members in network or community, detecting community of high modularity, understanding the behavior of a community, churn prediction are yet to find perfect solutions. In this section we describe some important studies which laid roots to community detection and triadic formation in mobile networks.

A. Stanley six degree experiment

Stanley Milgram is famously known for his "Small World Phenomenon" [7]. He routed messages in a network based on the familiarity of the first name and deduced to an interesting result which states that any person can be reached by any other person by not more than six hops. In simple terms, distant people can be connected by short paths in which every edge connects two people who know each other quite well. The data considered in this experiment was not as complex as the telecom data. This experiment has provoked us to experiment with mobile networks to explore the existence of knitted communities. Recent work has suggested that this phenomenon has effect in networks arising from nature and technology, and is a fundamental ingredient in the structural evolution of the World Wide Web. Such experiments on telecom network would give us an idea about the structure of the network. The number of hops or the time taken for information or message to exchange between any two random nodes can be deduced.

B. Pareto's 80/20 rule

It states that 20% of landowners own 80% of the land, 20% of the workers do 80% of the work, 20% of criminals carry out 80% of the crime and 20% of websites get 80% of the traffic. This rule when tuned to the telecom network it will give us an idea that 20% of our customers make 80% of our business [8]. Identifying these set of customers is one of the challenges as mentioned previously and these customers are called alpha users in many studies.

C. Community Detection

There have been many community detection algorithms seen in recent times. CNM algorithm [4] is one of the best of its kind. It is a hierarchical agglomerative algorithm which iteratively merges nodes into communities based on the gain in modularity. Its running time on a network with n vertices and m edges is O(m d log n) where d is the height

ISBN:978-1-61208-115-1

of the dendogram. Fast unfolding [9] is another community detection algorithm based on modularity. It is divided into two phases where in the first phase every node has its community number. Then the modularity is calculated with respect to all its neighbors. If there is positive gain in the modularity then the nodes are merged.

The CNM algorithm at this point has to be modified accordingly in order to meet few exceptions. The algorithm is purely based on graph theory where it considers every user as a node. It does not work under the condition where a loop exists between two nodes. A loop in telecom network can be understood as a bi-directional edge between two users. In order to solve this problem the algorithm has been slightly modified to consider the directional edges. Whenever such situations are raised the source and destination are swapped in order to overcome the loop as the bidirectional edge is made into two unidirectional edges.

D. Triadic closure

Triadic closure is a concept in social network theory, first suggested by German sociologist Georg Simmel in the early 1990s. It is a property between 3 nodes A, B and C. Suppose there exists a strong tie between A-B and B-C, then there will be a weak or strong tie between A-C. In [5], Mark Granvotter has synthesized the theory called “Cognitive Balance” which refers to the tendency of two individuals who feel the same way about an object. If the triad between three nodes is not closed, then the node connected to both the nodes has the ability to close this triad in order to complete the closure in the relationship network.

III. PROCEDURE

A stepwise presentation of our research work has been presented in this section. The unnecessary records from CDRs are filtered out and for better results the users are classified into active and potential users. After the users are classified into the mentioned categories, they are used as input data for the community detection algorithm which results to isolated and non-isolated communities. The last step which is the major part of our research is the generation of new edges between nodes which are not connected and have a strong common node. The new edges which are generated complete the triad formation.

A. Data processing

The telecom data used here had 1.2 million CDRs from African Telecom operator. This data includes the CDRs related to voice, SMS, GRPS. The data used here was over duration of week. Our research is carried only on voice calls and hence the first step is to filter out all the call records related only to voice calls. The next step of data cleaning is to identify the called numbers which have an outdegree 0 or indegree 0. Outdegree 0 can be defined as the caller ID which in all cases is the destination of call and not a source. Indegree 0 can be defined as the caller ID which in all cases is the source of call and never a destination.

- Example of Outdegree 0: Toll Free numbers and call centers
- Example of Indegree 0: IVRS generated advertisement.

As the caller ID’s are nearly 10 digits in length, processing them at every step is yet another time consuming process. Unique ID’s were given to the users to overcome this problem.

B. Classification of users

The filtered data from the processing step is used for this step. The attributes used are Calling Number, Called Number and Time Duration. The data has 1,82,000 distinct users.

Users have been classified into weak and potential users based on their calling patterns and their behavior in the network. Thresholds have been set on Indegree (number of calls received) and Outdegree (number of calls made). Certain rules have been followed in order to classify the users accordingly. These rules are user specific and depend on the data used. The rules designed by us are on the basis of the 7 days data on which we carried our experiments. The following are the rules laid down by us which classify a user as a potential user:

1) User has to be active on any 4 days of the week ,
2) Indegree of the user should be greater than or equal to 10,
3) Outdegree of the user should be greater than or equal to 10,
4) Sum of indegree and outdegree should be greater than or equal to 20.

The basis for the rules is as follows:

Rule 1) User has to be active on any 4 days of the week:

We found an interesting pattern on observing the data over the duration of the data. Some of the users were active only on the first day of the week, some are active on the last day of the week and most of the users were active for an average of 4 days of the week. This has been calculated by noting down the availability (A) and non-availability (NA) of the user over a week. The presence of a user j on a day i is defined as $\boldsymbol{P_{ij}}$. The value of $\boldsymbol{P_{ij}}$ is 1 if the user is present on day i else it is assigned 0. The Average Active Period (AAP) is defined as

$$AAP = \frac{\sum_{j}^{users} \sum_{i}^{days} P_{ij}}{total\ number\ \ of\ users} \qquad (1)$$

Rule 2) Indegree of the user should be greater than or equal to 10:

Fig. 2 plots the net indegree or number of calls received by every user over a week and average indegree of a user over a week turned out to be 10. The horizontal axis is indicated by the users of the network by their distinct ID’s and the vertical axis has the Indegree of every user.

Rule 3) Outdegree of the user should be greater than or equal to 10:

Fig. 3 plots the net outdegree or number of calls made by

 ISBN:978-1-61208-115-1

every user over a week and average outdegree of a user over a week turned out to be 10. The horizontal axis is indicated by the users of the network by their distinct ID's and the vertical axis has the Outdegree of every user.

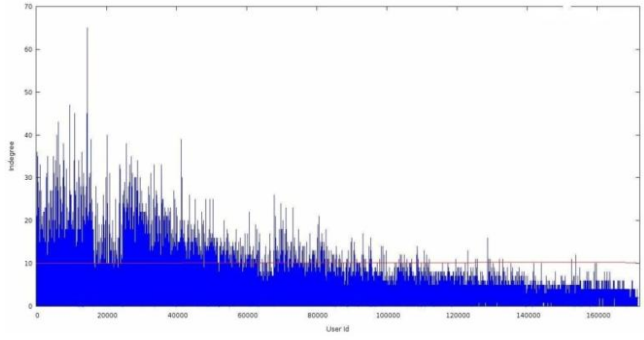

Figure 2 Indegree Vs User Id

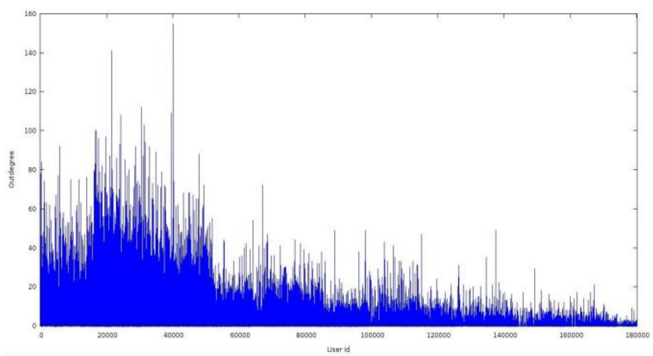

Figure 3 Outdegree Vs User Id

Rule 4) Sum of indegree and outdegree should be greater than or equal to 20:

This rule has the roots from rule 2 and 3. The sum of average indegree and average outdegree leads to 20. This rule has been made to ensure that we do not have users who just receive calls or make calls.

User who satisfies all the four rules is classified as *potential users* and the rest are called *weak users*. The latter are not considered as input set for further steps in the study. These users might increase the revenue of the network but the network or community does not shatter once they churn out and these set of users cannot be exactly put into a particular community. Addition of these users in the collection might lead into biased results. Following are the examples for weak users

1) *Call-Center*: The indegree of every call center is high when compared to its outdegree. It is impossible to categorize them into a particular community

2) *Business scheme promoters*: The indegree in this case is negligibly small when compare to its outdegree. These set of users are also considered as weak users.

At the end of this stage our data set consists of only active users. Calling number, called number and duration of the call are the only attributes which are being used. At the beginning of this step we had 1,82,000 distinct users, and after classification only 68,000 users are classified as active.

C. Community Detection

Community detection has been the toughest challenge in research area in present days. CNM performs the same greedy optimization as the algorithm [10] . What makes this perform better than the existing one is the way it handles the sparse matrices. Sparse matrices are faced in the early stages of formation of adjacency matrices and this operation is efficiently carried out in CNM using data structures for sparse matrices.

The calling number and the called number are taken as inputs for the community detection algorithm. CNM algorithm [4] is used for community detection. It is a hierarchical agglomeration algorithm which detects the com-munity structure and is faster than many competing algorithms. Initially every node is considered as a community and adjacency matrix A_{vw} is generated for that network. $\delta(c_v, c_w)$ denotes the connectivity. δ equals 1 if there exist an edge, else it is 0. The number of edges m in the graph and vertex x belonging to community c_v is given by

$$m = \frac{1}{2}\sum_{vw} A_{vw}\,\delta(c_v, c_w) \qquad (2)$$

The fraction of edges making one community is given by

$$\frac{\sum_{v,w} A_{vw}\,\delta(c_v,c_w)}{\sum_{v,w} A_{vw}} = \frac{1}{2m}\sum_{vw} A_{vw}\,\delta(c_v, c_w) \qquad (3)$$

Modularity is the main parameter which is taken into consideration when the communities are made. The fraction is directly proportional to the probability of the two edges merging into one community. Degree k_v of vertex v is defined as the number of edges leading to it, then the modularity (Q) is defined as

$$Q = \frac{1}{2m}\sum_{vw}\left[A_{vw} - \frac{k_v k_w}{2m}\right]\delta(c_w, i) \qquad (4)$$

The input handling has been changed to avoid the algorithm from running into an infinite loop. Eqn. 5 shows that CNM algorithm [4] works only for undirected edges. Every edge in telecom network is directed. In other terms every edge in telecom network has a source (calling number) and destination (called number). CNM does not identify such edges and considers them as undirected edge which leads the algorithm into an infinite loop. The modification which is made into it was at the input stage for algorithm. The input file format is a list of called and calling numbers.

$$\delta(c_v, c_w) = \sum_i \delta(c_v, i)\delta(c_w, i) \qquad (5)$$

The condition checkpoint which was introduced here is: If there exists an edge as A to B and B to A then swap B and A in the second case or else there would be an infinite loop between A to B and algorithm stops working. Once the nodes are swapped in the second case there will only be 2 copies of edges between A and B which resulted in dissection of loop which solves the problem. Upon every merge, the rows and columns of the corresponding communities are blended. Alternate way is to store the difference in the values of the modularity which result from the conglomeration of communities. Choose the largest among them and perform amalgamation. This would result in lesser memory storage and computational cost as these are the main hurdles when we are dealing with huge data.

The strongest user of a community is the alpha user of the community [1]. Definition of alpha user varies from research and its usage. In our research it is defined as the user of a community on whose removal there is a high probability of the network getting shattered off. The user is highly connected to most of the users in that community. The user has the capability to spread the information in less time.

D. Isolated and Non-Isolated Communities

Every community formed is given a unique ID. Every community is now considered as a node and the degree of the community is calculated. No of edges from a particular community to other community is defined as the degree. If the degree is equal to 0 then it is called an isolated community else a non-isolated community. Higher the number of isolated communities, higher is the modularity of the community as modularity is defined as the ratio of number of isolated communities to the total number of communities formed. Isolated communities are those whose members make calls only to their community members and make no calls to any other community members.

E. Generation of New edges

All the communities and the community members along with their respective CDRs which contain the calls made by them or call received by them along with the duration are taken as input. We have presented the algorithm related to the generation of new edges in mobile telecom networks.

Algorithm A graph G = (V, E) with V>0 and E>0

1) Look for two nodes x, y that are non-adjacent and share a common neighbor
2) If no such pair of nodes x, y exists, go to step 8, else goto step 3
3) The edge weight between x and the common neighbor is labeled as w1
4) The edge weight between y and the common neighbor is labeled as w2
5) Check if w1 and w2 are more than the threshold weight set by the user
6) If not go to step 8, else goto step 7
7) Add edge(x, y) and assign a weight equivalent to average of w1 and w2
8) Add element x, y to set E
9) Go to step 1.

The implementation of the algorithm is as follows:

Formation of new edges between nodes which are not connected and have a strong common node in between is the motto of the algorithm. To enhance our results, we will concentrate on every community individually and try to form new edges in between users of the community. These nodes can be easily joined because they belong to the same community. Certain rules have been laid down in computing such edges. Let’s say there are three nodes A, B, and C as shown in Fig. 1, where AB and BC are connected and B is the common node. Weight of the edge AB is w1 and weight of the edge BC is w2. A new edge AC will be formed if and only if w1 and w2 are more than the threshold weight. This threshold weight is user specific. This is used to define whether the edges w1 and w2 (already existing edges) are strong or weak. Once this new edge is formed its weight is defined as the average of w1 and w2. Formation of new edges not only increase the weight of the structure but also increase the closeness, clustering co-efficient, average talk time and average revenue of the community. The betweenness values of the community decreases which actually indicates that the probability of the network getting disturbed when a node or user moves to other network is less.

IV. RESULTS

The section describes the results achieved. The result consists of communities deduced with high modularity over duration of seven days. Another experiment which was carried on was to find out the behavior of communities over two time periods, namely weekends and weekdays. The alpha members of every community have been highlighted using Pajek software [11] which is a tool used for network analysis and visualization. In the last section we have presented the results on triads using a community of size 339 nodes as network.

TABLE I. ATTRIBUTES OF COMMUNITY

Period	Modularity	No. of Communities	Min Size	Max Size	Mean size
Week (7 days)	0.9970	8533	2	281	3.52
Weekdays	0.9506	9613	2	624	5.48
Weekends	0.9704	9703	2	1086	6.87

 ISBN:978-1-61208-115-1

Table I shows the behavior of communities that are formed over three time periods namely weekdays, weekends and over a week. It is shown that the communities that are formed on weekends have better modularity, mean size and max than when compared to the communities formed on weekdays. This is because the users in the network are active over the weekends. The above table shows attributes of communities that are formed over two time periods namely weekdays and weekends. Communities that are formed on weekends have a better modularity, mean size and max size than the communities formed on weekdays. This is because the number of calls and the duration of calls are high over the weekends.

Table II shows the number of communities and their respective degrees. As mentioned earlier a community with degree 0 is called an isolated community where the users belonging to that community only make intra-community calls.

Interesting results which were deduced by the formation of community over a week

1. 8510 out of 8533 communities are isolated
2. Every community has an alpha member. Hence there are 8533 alpha members in the network
3. Number of communities which have only one edge to other communities: 8

TABLE II. COMMUNITIES AND THEIR DEGREE FORMED OVER A WEEK

Degree	No. of Communities
0	8510
1	8
2	4
3	4
4	5
5	2

TABLE III. DISTRIBUTION OF ISOLATED COMMUNITIES FORMED OVER A WEEK

Community Size(size)	No. of Communities
1 < size <10	8374
10 < size < 20	86
20 < size < 30	22
30 < size < 40	14
40 < size < 50	9
50 < size < 60	5
60 < size < 100	2

Table III describes the distribution of isolated communities over a week. Size of a community indicates the number of users in that particular community. 8374 communities are of size less than 10, which indicates that there are more number of smaller communities when compared to communities of size 100.

Fig. 4 represents the hierarchical representation of a community of users based on their net degree in ascending order from the top. This is a sample representation of one of the community whose size is 339. The node at the bottom of the diagram is the alpha member. It can be clearly seen that the number of connections between the alpha user to the other users in the community. The users at a particular level have equal degree.

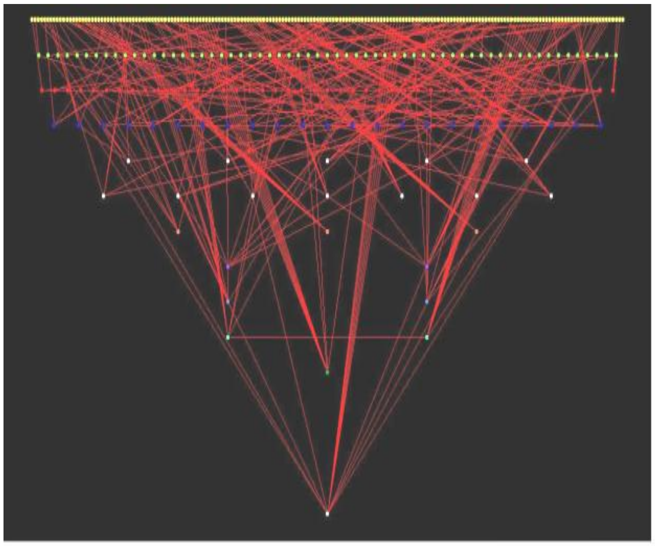

Figure 4 Hierarchical representation of nodes based on degree

TABLE IV. BEFORE AND AFTER GENERATING NEW EDGES

Measure	Before	After	% change
Closeness	0.0741	0.11	48.4
Average Talk time	152.3 units	198.6 units	30.4
Average Revenue	228.45	300.7	31.6
Betweenness	0.443	0.441	(0.45)
Modularity	0.992	0.9910	0.08

Fig. 5 shows the triads formed on a community of size 339 nodes with weight threshold 100. Red colored edges (353) are the existing set of edges and yellow colored edges (311) are new edges on addition of which triads are formed.

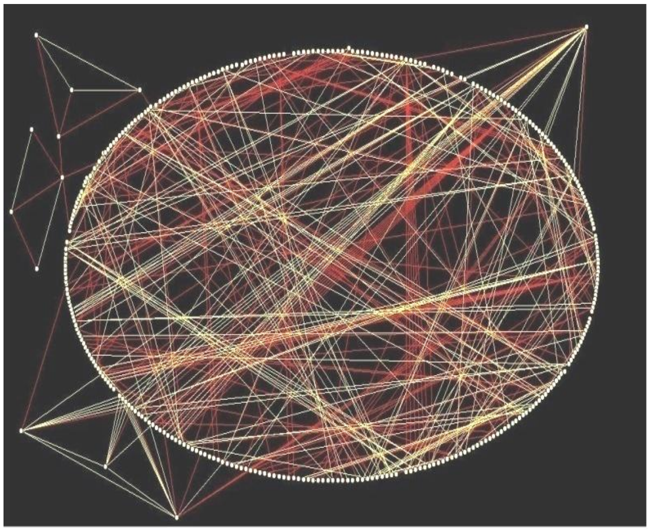

Figure 5 Triads formed on a community

V. DISCUSSION

In this section, we discussed about the advantages of the research based on our experiment of triad formation with a community of size 339 nodes as an example. Various graph theory attributes [2] which depict the strength of a network have been calculated and compared under two scenarios namely before generation of new edges and after addition of new edges.

The attributes which were used to show the success of our method are *Closeness, Average talk time of the community, Average Revenue of the community, Betweenness* and *Modularity*. Closeness is the inverse of sum of the shortest distances between each and every node in the network. Betweenness is a measure which depicts how much the network can be shattered if a person is inactive or churned.

From Table IV, we infer the following:

1) Closeness between nodes is increased by 48.4% from the earlier case which implies of how close the nodes are knitted. The shortest path between any two nodes has decreased and hence the there is an increase in the measure
2) Average Talk time has increased by 30.4% as new edges lead to rise in number of calls.
3) Average Revenue has increased by 31.6%. Number of calls has increased which will automatically trigger the revenue to higher levels.
4) Betweenness has decreased by 0.45% which shows that the amount of shatter or disturbance caused by a particular user has lessened.
5) There is not much increase in modularity as our model was of size 339 nodes and modularity would only increase if there are new users added to the community. In our cases we are only working on new edges. The increase of 0.08% can be explained as formation of one or two isolated communities as new edges have been formed.

VI. CONCLUSION

Based on the results we can conclude that all the steps followed in our experiment have enriched the results starting from the initial step of filtering out users who's either indegree or outdegree is 0. Then our assumption of concentrating on active users would lead to better results has also been proven by achieving communities of higher modularity. Generation of new edges have been proved to be an advantage for the operator as it bonds its users much higher than the earlier stage which would in turn increase his revenue. Considering other parameters of CDR like SMS, GPRS, would lead us to find much more potential nodes which later can be used as data for our experiments. Moreover the Customer ranking can be considered at community level. The usage of location-based information can strengthen our claim of triad formation in mobile networks.

REFERENCES

[1] S. Doyle, "Social network analysis in the Telcom sector Marketing applications" Journal of Database Marketing & Customer Strategy Management (2008) 15, 130 – 134. doi: 10.1057/dbm.2008.8

[2] M. Saravanan, G. Prasad, S. Karishma and D. Suganthi, "Labeling communities using structural properties," International Conference on Advances in Social Networks Analysis and Mining, 2010. doi: 10.1109/ASONAM.2010.49

[3] K. Wakita and T. Tsurumi . "Finding community structure in a mega-scale social networking service". in Proceedings of IADIS international conference on WWW/Internet 2007, pp.153-162, 2007.

[4] A. Clauset, M. E .J. Newman, and C. Moore, "Finding community structure in very large networks". Phys. Rev. E 70, 066111, 2004. doi: 10.1103/PhysRevE.70.066111

[5] M. Granovetter, "The strength of weak ties," The American Journal of Sociology, vol. 78(6), pp. 1360–1380, 1973.

[6] M. E. J. Newman and M. Girvan, "Finding and evaluating community structure in networks," Phys. Rev. E 69,026113, 2004. doi: 10.1103/PhysRevE.69.026113

[7] S. Milgram, "The small world problem," Psychology Today, 1967.

[8] A. Bookstein, "Informetric distributions, part i: Unified overview," American Society for Information Science, Vol. 41, pp. 368–375, 1999. doi: 10.1002/(SICI)1097-4571 (199007)

[9] R. L. Vincent D. Blondel, J. L. Guillaume and E. Lefebvre, "Fast unfolding of communities in large networks," Vol. 2008, No.10. doi: 10.1088/1742-5468/2008/10/P10008

[10] M. E. J. Newman "Modularity and community structure in networks", physics/0602124 Proceedings of the National Academy of Sciences (USA) 103, pp. 8577—8582, 2006.

[11] http://vlado.fmf.uni-lj.si/pub/networks/pajek/.