

The Impact Of Social Network Analysis On Telecommunication Customers

Ramiz Assaf

Abstract: Telecommunication industry has proven itself as a chain of socio-economic impact. Market leaders should take advantage of data analysis consecutively as the need to sustain customers the longest possible time. This research is based on community detection analysis through identifying group members and influencers in the telecommunication industry. It is motivated by two major research questions: (1) Who are the influencers in subscribers for churn and new products adoptions? (2) Who are subscribers whose loyalty is threatened by churn around them? The use of data analysis and data mining algorithms and standard process in today's challenges was set to increase the efficiency of solving not only technical problems but business as to move toward insights and analysis based exercises. This research proposes a model based on the Similarity, Authority and Hub algorithms used in Social Network Analysis (SNA) to group members who share same patterns together, to identify network leaders in terms of authority and dissemination leaders and to extract group level insights that can be used in understanding subscribers' patterns and ways of communications (in-degrees and out-degrees) between them using IBM SPSS Modeler as the base of data mining process and Oracle as the base of corporate data warehouse. This study estimated how much reliable is to use SNA in order to shift from the traditional ways of targeting subscribers, understanding their pattern, enhancing offering efficiency and improving the process of launching new product by selecting the right target. It has contributed to shifting marketing leaders' way of thinking from individual view to network view. It has also contributed to providing decision makers with a systematic and scientific approach as to focus on network influencers, to save as much as they can and to build their offers based on it.

Index Terms: Social Network Analysis; SNA; Authority; Dissemination; SPSS Modeler; data mining; Data Analysis; SQL; SPSS Modeler.

1 INTRODUCTION

Telecommunication industry globally is one of the important industries in the modern time, it has both economic and social impact on societies. Due to the excessive amounts of data in such sector, Business Intelligence (BI) and Data Mining (DM) techniques are aggressively taking their place as the top of most telecom operators and they become one of the biggest competitive advantage in those firms. These techniques help analyzing users' activity patterns and understanding their behaviors can help reacting quickly to changes. On the other hand, social life is defined as the relations between people that set up networks. Those people are organized into a community that includes social structures. Within every social structure inside community, every individual exhibits different behaviors. A small number of individuals in each community shape the overall social structure, they are called influencers. [1] Social Network Analysis (SNA) was emerged as a key research area in modern sociology, economics, information technology, anthropology, biology, geography, communication studies, organizational studies, linguistics, and social psychology. [2] SNA was also used to study how companies interact with each other, characterizing the many of the informal connections that link employees together. [3] In telecommunications, each successful transaction consists of sender, connection link and receiver. In the mobile operations there are more than one type of connection links, such as calls, text messages, e-mail, multimedia messages, and/or Internet access. SNA calculations can be applied on both individual and group characteristics. Individual characteristics are based on self/nodal factors related to attributes and behaviors such as in/out-degree, dissemination and authority scores where group characteristics are more related to group/community level insights such as group density and group size. [4]

This research proposes a model based on the Similarity, Authority and Hub algorithms used in SNA to group members who share same patterns together, to extrapolate network

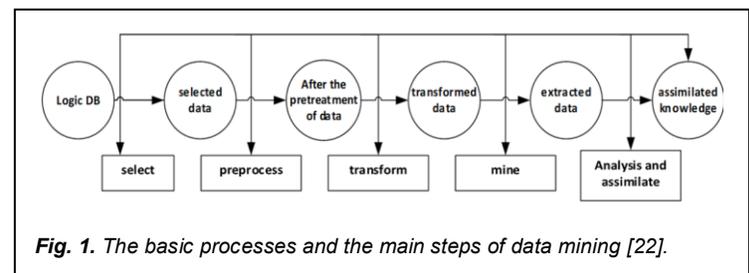


Fig. 1. The basic processes and the main steps of data mining [22].

leaders in terms of authority and dissemination leaders and to extract group level insights that can be used in understanding subscribers' patterns and ways of communications (in-degrees and out-degrees) between them. In the past, SNA had been also utilized as a part of the study of disease transmission to help see how examples of human contact help or hinder the spread of infections, for example, HIV in a populace. [5] SNA was also used in mass surveillance to determine whether or not US citizen were political threats where Total Information Awareness program was doing in-depth research on strategies to analyze social networks. One study found that happiness tends to be correlated in social networks. When someone is happy, near friends have a 25% higher chance of being also happy themselves? In addition, people at the middle of a social network tend to become happier in the future than those at the borders. Groups of happy and unhappy people were recognized within the studied networks, with a reach of three degrees of separation: a person's happiness was associated with the level of happiness of their friends' friends' friends. [6] A limited number of related prior studies have proposed approaches to use social network information in order to predict customer churn. Dasgupta et al [7] research was pioneering research in predicting customer churn using social ties between the subscribers in a telecom operator. Richter et al. [8] presented the group first churn prediction approach to predict customer churn based on the analysis of social groups

• Ramiz Assaf, Industrial Engineering Department, University of Business and Technology (UBT), Jeddah, Saudi Arabia, E-mail: r.assaf@ubt.edu.sa

or communities derived from Call Detailed Records (CDR) data. The presented approach assigns a churn score to each subscriber based on the churn score of the social group as well as personal characteristics. The results of the study reconfirm the potential of improving the current generation of customer churn prediction models by adding information that captures the social interactions of a subscriber, and indicates that group structure and membership are determinants of churn behavior. This study opens interesting alternative modeling approaches to exploit the information contained within the network structure of the customers of a telco operator. Sonamine was able to predict 25% of real churners when incorporating the SNA with Churn model. The exercise was done in a European mobile operator; their prepaid churn model was not accurate due to lack of customer demographic information. SNA Accurately predicted 25% of churners using only 5% of population lift of more than 5X of accuracy compared to previous results. [9]. In contrast to those studies, this paper will focus on presenting the vital role influencers play in each community in order to increase the conversion rate of any new product or to decrease the churn effect on the whole community through targeting its influencers and sustaining them by telecom operators.

2. METHODOLOGY

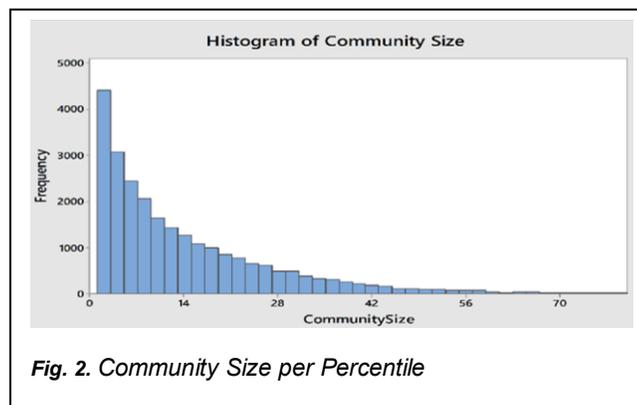


Fig. 2. Community Size per Percentile

The research methodology will be applied to one of the biggest telecommunication companies working in Palestine region. The methodology consists of six steps of data mining modelling. They are grouped into 'CRISP-DM' (CRoss Industry Standard Process for Data Mining) which means the standard process of any data mining model. [10]. The step of data understanding and extraction comes first; it deals with the data on a call detailed record (CDR) level that will be extracted from the corporate data warehouse. Second step is the data preparation which focuses more on aggregating the data on customer level and validating it to ensure a clean and correct data before transferring it to SNA tool. Third, the modelling phase will focus on applying SNA algorithms to derive decision making variables that will be used by business users. All data will be extracted from the corporate data warehouse (DW/DWH) and for the most recent three cycles. Customer influence factor (CIF) is calculated on monthly basis, that is all subscribers base will have an influence factor update on each month in order to establish a good sense of customer behavior. All measures related to CIF are established using the mean of the last three months. Mean is used to discard outliers, to reduce the impact of special peaks and to control

the average curve of customer behavior. To successfully define a SNA in any community, there will be two main algorithms inside according to IBM Knowledge Center related to the selected tool of IBM SPSS Modeler. The first step is splitting the groups and assign its members (partitioning into groups) which is a group level exercise that is based on similarity measures [11]. Second algorithm is more responsible for studying the pattern inside each divided group of members in terms of members' relations using Hyperlink-Induced Topic Search (HITS) algorithm [12].

2.1 Data Understanding and Extracting Phase

The first step in SNA modelling is to understand and extract the required data. A part of data understanding is to define customers' profiles data and behavioral data. Profile data are customers' gender, location, age and profession. While Behavioral are variables that are being updated frequently such as customer's activity on the network, activity duration, time and destination which are called Customer Detail Records (CDRs). Data was extracted on CDRs level after applying the filters eliminating inactive subscribers, machine generated subscribers or even call centers. To simplify the extraction process, due to excess amounts of data in the CDRs, monthly, regional means were the basis of the extracted from the corporate data warehouse. The final number of subscribers for the specific region are around 25,000. The 25,000 representative subscribers resulted in around 353 thousand records, this is a real case of how huge are the data in telecommunication industry (although this dataset is an aggregated CDRs). Data types are divided into two types CDRs and Profile Data. CDRs Data include: A number (Originator), B number (Terminator), Call type, Total and Mean call duration and Count of Calls. While profile data include information about: Mobile No, subscriber tenure, subscriber plan, and Region. SNA is a concept that is closely related to customer influence; thus, the analytic data model was developed according to residential customers only. These customers have the ability to influence each other. [4]

2.2 Data Preparation

A data preparation process is required to prepare all data extracted from the corporate data warehouse, and to ensure that it is used properly in the SNA study. At this point, information about customers, accounts, fixed lines, CDRs, products, and billing history is collected from the data warehouse. It is necessary to organize the data in a single dataset. Also, data processing takes a place in this phase where data checkups and validation occur. These steps include eliminating null values and outliers. The following three steps are processed during the phase of data preparation: Removing non eligible subscribers, eliminating the outliers and merging the two datasets. Removing Non Eligible Subscribers included deleting record of subscribers registered as call center or free numbers. This step also deletes any subscriber with less than three months of active subscription. In order to guarantee having normalized dataset and as a part of data preparation phase to exclude the outliers that might affect the model output, the data preparation process includes some steps in respect to the related content. The three attributes (Total Durations, Calls count and average call duration) are analyzed in terms of distribution, and the outliers are removed from the dataset of CDRs looking at the percentiles. The process itself is about doing the averages and percentiles

related to each attribute. For instance, the average community size in this case study's network is 14 persons, this indicates that each subscriber has a unique numbers of connections, to themselves, equals to 14 and in which there are some subscribers below or above this number. Percentiles provide a more distribution of customers' pattern divided into the percentages of 1-99% as shown in Figure 2.

Table 1: The distribution table for community size, duration per month, count of calls and duration per call attributes.

Percentile	1 st	10 th	25 th	50 th	75 th	90 th	95 th	99 th
Community Size (Number of nodes)	1	2	4	9	20	32	42	67
Duration Per Month (Minutes)	2.3	17.6	42.3	107.0	230.6	383.6	497.0	809.33
Number of calls	0.6	2.6	6.66	16.6	36.0	59.3	76.6	125.66
Duration Per Call (Minutes)	1.1	4.2	5.42	6.4	7.6	9.5	12.0	27.50

The last step is merging the two datasets, as now the targeted base went down to 24,300 subscribers. A process of merging the two datasets together (Aggregated CDRs and Profile) will take a place. This is needed to extract more profile data rather than only statistical measures. This adds another layer of validating the results with what was expected and comparing them with the current perspective (Region X generates more revenues than Y, does it include more influencers also?). The previous procedures related to understanding, cleaning and profiling of data is important to ensure a derived and reliable output.

3 Data Modeling

A social network is typically illustrated using a Sociogram [7]. In this type of visual display, individuals correspond to points, or nodes, in a space. Lines, or edges, connecting the points represent relationships between the individuals. If the relationships are directional, the edges include an arrow to indicate the direction. If the relationships have weights, the labels for the edges denote the values. Fig. 3 displays a network of four individuals (1,2,3 and 4) along with their relationship strengths (represented as total duration between them in 7,1,12,7 and 4).

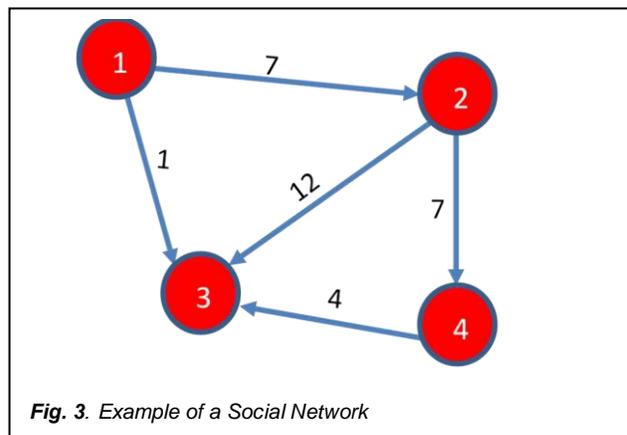


Fig. 3. Example of a Social Network

Suppose the network represents the phone calls made by individuals with the relationship weights indicating the length of the calls (duration). In this case, Node #1 called two nodes, spending the majority of time talking to node #2. A community is a group of people who have a shared pattern that distinguishes them from other communities. In this work, the initial sample of 25,000 subscribers is, with no doubt, a combination of different communities indicating influencers in each group. The group will be set to include members from 2 to 42 (10th and 95th percentiles). To summarize the process, the next steps are about dividing the sample into groups (Partitioning into groups), describing networks (network properties) and calculating influence factor (describing groups and group members).

3.1 Partitioning into Groups

Individuals who have high relationship weights (frequency of calls or total duration) between each other are grouped together, they are similar to each other and this similarity of individuals in a group is measured by weights. As a result, group identification begins by removing the weaker relationships in a network. The process of removing those weak relationships is based on the coverage threshold (θ). Coverage Threshold is defined as the fraction of the strongest relationships to retain. [13] For example, a coverage threshold value of 0.3 results in the strongest 30% of relationships being used for group identification with removing the remaining 70% of the relationships.

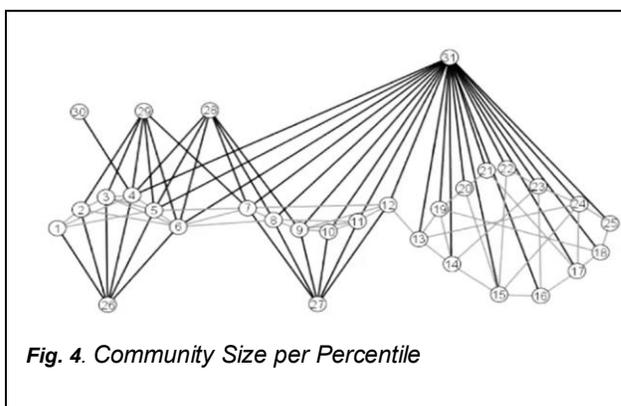


Fig. 4. Community Size per Percentile

Table 2 shows that group size limits will help in portioning the groups to either split those who have large number of members into separate groups or to completely omit those

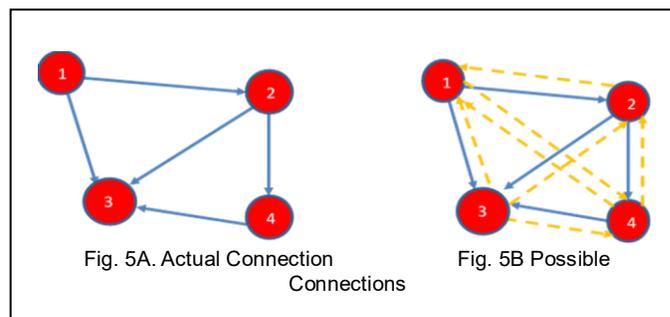
Table 2: Coverage threshold factor distribution

Coverage Threshold (Θ)	10%	15%	30%	50%	60%	80%	90%	95%
Total Nodes in Network	16,222	17,184	19,085	20,581	20,720	22,509	22,509	22,509
Total Links in Network	34,382	38,416	43,856	44,532	44,978	41,392	41,392	41,392
Total Number of Groups	1,903	1,862	1,633	1,487	1,471	1,197	1,197	1,197
Mean Group Size	8.52	9.23	11.69	13.84	14.09	18.8	18.8	18.8
Mean Group Density	0.3	0.31	0.3	0.29	0.29	0.25	0.25	0.25
Mean In-Degree	2.12	2.24	2.3	2.16	2.17	1.84	1.84	1.84
Mean Out-Degree	2.12	2.24	2.3	2.16	2.17	1.84	1.84	1.84

who have smaller members. Normally, the group limit is a minimum of two members and will be, in this case study, a maximum of 42 members. The groups remaining are called 'core groups' [13]. Table 2 illustrates the relation between coverage threshold and average group size by carrying out experiment with different θ from 10% to 95%. Table 2 show terms like "Total Nodes in Network" which refers to total subscribers resulted after applying each coverage threshold and omitting weaker relationships, "Total Links in Network" is number of connections that a network member originated to another network member, "Total Number of Groups" stands for count of output communities inside the dataset, "Mean Group Size" is important to Table 2 and shows that the more coverage threshold is, the less tight relationships weight (which might include weak relationships), the more average group members (including non-core members). The last three coverage thresholds (80%, 90% and 95%) resulted in same distribution of group members which concludes, for our sample dataset, that applying any of these thresholds will result in the same. In another meaning, it's is not significantly different to apply 80%,90% or 95% as coverage threshold. Splitting members into groups is based on the similarity between those members within the group and the dissimilarity between the groups and where nodes inside the group connected with many more edges than between groups [11]. Similarity is measured through examining the neighbors of each member and the strength of relation between members and their neighbors, two nodes are similar if when you start random walk from those two nodes, those random walks will meet soon. Random Walk is "an algorithm provides a good relevance score between two nodes in a weighted graph, that is how closely two nodes are related and it has been successfully used in numerous settings, like automatic captioning of images, generalizations to the connection subgraphs, personalized PageRank, and many more" [14]. There are two types of similarity measures considered: Structural Equivalence: Members within the group have a structural similarity compared to each other, they share the same neighbors and approximately the same relation strength. In order to calculate the structural similarity, Jaccard Similarity, Cosign Similarity and/or Pearson Correlation Coefficient can be used. [4] Regular Equivalence: Two nodes are regularly equivalent if they are equally related to equivalent others, in another words, they have the same pattern of connecting to others. Regular equivalence provides nothing for the group identification in terms of community, but it does when it comes to extract the similar pattern within all subscribers' base. [11].

3.2 Describing Networks

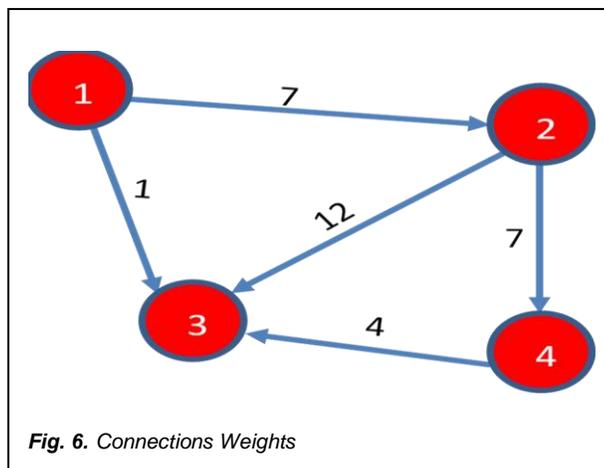
Information about networks, groups, and individuals needs to be extracted into descriptive characteristics that will allow cross-comparisons and inclusion in predictive models based on a set of key performance measures. One of the most used examples is comparing a group of nodes to another within the network as to describe the characteristics of the former compared to the latest, other example is compare individuals in the network to others to characterize them or to identify the most important ones. In order to describe social networks, density and degree are the most common used measures. They both reflect connectivity, density focuses on the entire network or network subs groups where degree describes the individuals in the network and characterizes them. [15]Network



density is defined as the proportion of actual connections (relationships) between nodes in a network (group) divided by total possible connections in that group [16]. The more actual connections, the denser is the network and the more cohesive are the nodes in the network. Density ranges from 0 to 1 taking into considerations that connections are either one way or two ways between the nodes (directed or undirected). Some nodes may not have direct connections to other nodes, some nodes may have only outgoing connection to other nodes (the relation is not reciprocated). Figure 5 stimulates a group with actual connections (lines in blue) compared to all the possible connections (dashed lines in yellow). [17] The possible connections calculation is the permutation where the size in the set is always two (a connection is made between two nodes).The formula for Network Density (ND):

$$ND = AC / P_2^k \quad (1)$$

where AC is the number of actual connections, is count of nodes in the network, P_2^k is the 2-permutations of k. The network density for the above sample of Figure 4 $ND = 5/(4!(4-2)!) = 5/12 = 0.42$



Nodal Degree individual's connections (relationships). The more connections that an individual has (both in and out), the more this individual is important in the network as s/he is involved in the most relationships. This individual acquires information from a variety of sources and spreads it to a large number of other group members. So, degree is defined as total number of connections involving each node in a network and is classified to either 'in-degree' or 'out-degree'. In the first, a node is considered to be the target involving how many connections were made to this node. Conversely, out-degree is the number of connections made where the node is the source. In-degree is often treated as a measure of prestige. Higher in-degree values correspond to more relationships ending at that node. In other words, those individuals are contacted by a high number of other individuals. Many other nodes are initiating relationships with the node. On the other hand, out-degree is treated as a measure of centrality. Higher values correspond to more relationships originating from that node. Those individuals contact a high number of other individuals. [18] For nodes in the network of Fig. 5A, the out-degree values reveal that node #3 is the most central. Based on the in-degree values, nodes #2 and #4 have more prestige than #1 and #3. Although, Node #2 is more active compared to node #4 (it has a total degree of three compared to two).

3.3 Describing Groups and Groups Members

In addition to the density, in-degree, and out-degree, other statistics describe group dynamics. In particular, authority and dissemination scores offer measures of the social status of the individuals within the groups. The role of each individual in a group is vitally important when trying to predict the behavior of both the group and its members. Fig. 6 indicates the weights of relationships between group nodes and in terms of duration. For example, node #2 called node #3 by 12 minutes, and at the same time node #3 received 12 minutes from node #2. Who are authority and dissemination influencers? Authority and dissemination algorithms are based on the weights of relationships between group members. The more an individual receives from others, the more they ask him or her for information or opinions, which is called Authority. On contrary, the more an individual connects to other members, the more he has an ability to spread the word. This is called dissemination. Table 3 summarizes the previously mentioned group on the level of each node. There is two widely used algorithms to calculate the authority and dissemination, either to use the PageRank or HITS algorithms and taking into

considerations that both algorithms are based on the right understanding of adjacency matrix resulting in the score of importance of each node inside the network. [19]

Table 3: illustrating Figure 6 details

Node #	In-degree	Out-degree	Incoming Duration	Outgoing Duration
1	0	2	0	8
2	1	2	7	19
3	3	0	17	0
4	1	1	7	4

3.4 PageRank and HITS

Devi, et al. defined both algorithms, summarized the advantages and disadvantages of each algorithm and exercised them on a network of webpages (represented as nodes A and B). Link analysis algorithms are used to calculate webpage rank. An example of these algorithms are Pagerank and HITS algorithm, they are different link analysis that employ different models to calculate the rank. Both algorithms give importance to links rather than the content of pages. In the Pagerank, rank score of a page is divided evenly over the pages to which it links. But, HITS algorithm rank pages according to their Authority and Hub of a page. Google search engine uses Pagerank algorithm due to the features of link feasibility, less query time cost, efficiency and less susceptibility to localized links. HITS is used by IBM in all of its applications and search engine. Results demonstrate that HITS calculates authority nodes and Hub correctly. HITS may also be combined with other information retrieval based rankings. [19] One of the main advantages of HITS over PageRank is that the first uses the input of three important parameters (In-degree, Out-degree and relationship weights) where the latest is based on 'In-degree' only. Regarding the disadvantages, the first requires more processing time (due to the variability of inputs and relations) where the latest is considered to be more efficient in time manner. What to use is based on the context. IBM Modeler uses HITS algorithm as in the same for IBM Search Engine.

3.4.1 Authority Score

Authority scores measure the importance of an individual corresponding the number of relationships ending at him or her indicating how much an individual receives from other members of a group. The more incoming connections to this node, the more s/he has authority compared to the group members who are asking for opinions or information. Authority scores ranges from 0 to 1, the closer to 1 indicates a closer authority and the higher potential of being the authority leader of that group. [20]. The confidence of authority for the leader (overall strength of the authority leader) is computed through dividing highest authority score in the group by lowest authority score, following the below equation:

$$\text{Authority Leader Confidence} = \frac{\text{Max (Authority Score)}}{\text{Min (Authority Score)}} \quad (2)$$

3.4.2 Dissemination Score

In contrast, dissemination score measures the importance of an individual with the number of relationships originating from him or her indicating how much an individual connects to other members of a group. If a particular person contacts many other people in the group, that person can significantly affect the opinions of the entire group. The closer the dissemination

score is to 1, the more the node connects to the other group members. The **dissemination leader** for the group is the one who has the maximum dissemination score among other members [20]. The confidence of dissemination for the leader (overall strength of the dissemination leader) is computed through dividing highest dissemination score in the group by lowest dissemination score, following the below equation:

$$\text{Dissemination Leader Confidence} = \frac{\text{Max (dissemination Score)}}{\text{Min (dissemination Score)}} \quad (3)$$

4. Results

4.1 Applying SNA

The concept of SNA in the previous sections was discussed starting from Group analysis measures such as number, density, size and confidence of groups and Individual analysis like nodal degrees, authority and hub scores. In order to apply the SNA on the projected sample, IBM SPSS Modeler is used applying HITS algorithm through IBM SPSS Modeler SNA Group Analysis Node. Below is a snapshot for one group community (group numbers of #59) result generated by this Node as an example. [21]

4.2 Group #59 Results

In this sub section, a visual representation of group #59 relationships weights and directions in addition to the group analysis and Individual analysis tables output will be discussed.

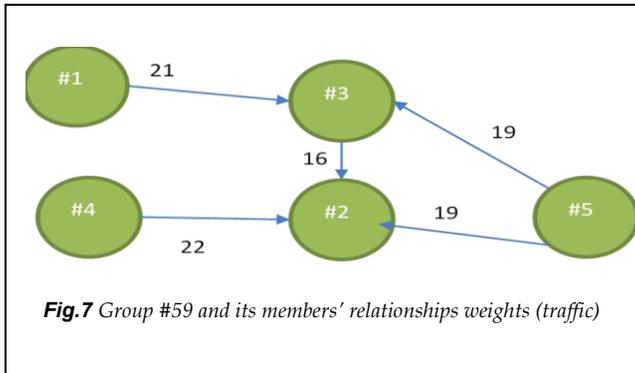


Table 4. SNA Group Analysis Results for Group 59

Group Number	59
Size	5
Density	0.25
Max Authority Score	0.455877
Min Authority Score	0.101223
Authority Confidence Score	4.50368
Max Dissemination Score	0.26307
Min Dissemination Score	0.145487
Dissemination Confidence Score	1.80821

Table 5. Group Analysis Individual Results for Group 59

Node Number	Member #1	Member #2	Member #3	Member #4	Member #5
Authority Score	0.101	0.455	0.240	0.101	0.101
Authority Rank	5	1	2	3	4
Dissemination Score	0.225	0.145	0.182	0.182	0.263
Dissemination Rank	2	5	3	4	1
In-Degree	0	3	2	0	0
Out-Degree	1	0	1	1	2
Group Authority Leader	0	1	0	0	0
Group Authority Leader Confidence	0	4.503	0	0	0
Group Dissemination Leader	0	0	0	0	1
Group Dissemination Leader Confidence	0	0	0	0	1.808

Examining Table 5, Member #2 has the highest Authority Score, which means that all other network members, as mentioned earlier, refer to him before taking any decision. Having member #2 outside the network means that there will be no reference for other group members which will lead to a higher probability for others to leave the network as we can notice that member #2 had the highest in-degree as well (although highest in-degree will not guarantee having a member as influencer since the weight is traffic based), this is the churn influence of leader on other members. For those who will adopt, Member #5 has the highest rank type 2 which is more responsible to disseminate (spread the word) to other network members. Having member #5 convinced in a service and buying it will lead to a higher probability for others to do so as we also can notice that member #5 has the highest out-degree, this is effect of dissemination of other network members. Group members 1,3 and 4 will be more probable to have their loyalty threatened as result of having both members #2 and #5 not satisfied.

5. Discussion of Results

The previous four sections discussed theoretical parts of the work along with methodology section which focused more on the drivers related to SNA model in terms of their definitions, calculations and algorithms applied. This section will answer the research main questions. A model of subscribers' network was built to determine the connections and the strength of those connections by which one can conclude who is influenced and who is influencing. With this map of the social community within the grasp, SNA can help telecom operators to make better marketing decisions. It can also help predict

churn and retain customers by understanding how influencers will cause a ripple effect through communities, acquire new customers through member get member campaigns, and help cross sell and up-sell with targeted viral marketing. Most importantly, this SNA model will allow telecom operators to change the way they target customers to include those with secured ROI which means less offering cost. Since the cost of acquiring a new subscriber is high, it makes sense to start by retaining our existing customers. SNA uses community circles to pinpoint these "contagious churners". These are subscribers who have not churned yet, but are being influenced by their friends and social group to churn. Once we have found the influencers we can target them with disproportionate benefits,

after all it is better to target one and allow their influence to naturally do its job rather targeting everyone but the key here is who is that 'one'. The final result of SNA came up with very rich insights that need to be studied and analyzed by the company under study. Those insights will help them understand their subscribers' patterns in terms of number of communities, average community size, number of connections and how many influencers are there. The use of SNA results will also assist in defining the type of subscriber that the company needs to target based on their KPIs. So if it was meant to introduce a new product in the market, then the strategy here is 'to spread the word' by using the dissemination factor. But if it was meant to acquire new members or to increase sales on a new product then the strategy is 'to target the influencers' by using the authority factor. SNA will provide those two key players inside every community! As an example, in a traditional approach which is used by most of companies through targeting its current customers, they will start looking at how much those customers generate revenues? How much do they utilize the traffic? When do they talk? And start to build an offer expecting uplift equal to 5%. However, on an individual level when incorporating SNA, their target will be more focused on a community view instead of individual which will have an impact of all of community members instead of one individual who might not be the influencer. Regardless if the KPI was set to reduce the churn, stretch the revenue or increase the base, the cost will be also focused on the community key players! Those who will for sure response and let other members do, those who we can expect an ROI of them and their affected followers. Table (5-1) shows a summary of the previously selected sample in regards of total members (nodes), links (connections), groups (communities) and other group details.

Table 6. Summary Statistics

Name	Value
Total Nodes in Network	17,184
Total Links in Network	38,416
Total Number of Groups	1,862
Mean Group Size	9.23
Mean Group Density	0.31
Mean In-Degree	2.24
Mean Out-Degree	2.24

At a coverage threshold equal to 15%. SNA model indicated that there are 1,862 final groups for the dataset of size equal to 17,184 subscribers (the difference between 24,000 and 17,184 is due to coverage threshold, this is referred as "Total Number of Groups". "Total Links in Networks" (in-degree and out-degree) were 38,416 links with an average of 2.24 of links per group member. Finally, the "Average Group Size" was 9.23 with an "Average Group Density" equal to 0.31.

5.1 Who, when they churn, would take few friends with them?

A group of 9 members will have two influencers, the authority leader and the dissemination leader. If the authority leader has churned the rest of members will have no authority leader to ask for his/her opinion – due to his highest in-degree level and weight. So, they will simply follow him/her to a new telecom operator. Network operators are keen to sustain their

subscribers' base the longest possible period of time. SNA will assist to understand their base structure and to avoid the churn of network communities key players which when they churn will take their friend with them. Hence, the company should offer a very aggressive offers for those influencers who are likely to churn in order to prevent them, and their community members, from doing so.

5.2 Who, when they adopt, would push a few friends to do the same?

On the other hand, if the dissemination leader decided to adopt to a new product, possessing the powerful tool 'the word of mouth' is due to highest out-degree. Then the rest of network members are expected to adopt. Instead of targeting all base and paying a lot of money to 'spread the word' which will affect their campaigns or product release to the market, branding and awareness costs, operators can hire the network community dissemination leaders to do the job.

5.3 Who are Subscribers whose Loyalty is threatened by churn around them?

Finally, the rest of 7 members are the group followers that their loyalty will be threatened by the churn of the influencers. In their authority and dissemination scores, they will have lower scores compared to their leaders, they are flagged as '2-5' in their 'Authority rank' and 'Dissemination rank' of a group equal to five members as the example Table 7 of group 59 discussed earlier.

Table 6: Authority rank' and 'Dissemination rank' for group 59

Node Number	Member #1	Member #2	Member #3	Member #4	Member #5
Authority rank	5	1	2	3	4
Dissemination rank	2	5	3	4	1

6 Conclusions

Dynamicity environment of telecommunication industry, high-level of competition and increased customers' expectations have made necessity of getting awareness of increasing offering mechanism, predicting influencers power overall other community members and utilizing subscribers pattern to understand network structures. This study defines SNA as a method that helps operators to plan, monitor, analyze, and manage customers' network more effectively by providing a network view for the decision makers. With a shared purpose, a consistent data model, real-time information, easy-to-use tools, and streamlined processes, it's much simpler to align data driven insights with strategy that will support faster decisions and boost performance to achieve business goals. SNA drives organizations to focus on the key drivers of value as they relate to corporate strategy and scientific offering and targeting processes. SNA provides fact-based guidance for value-based decision making. This research has contributed to providing decision makers with a systematic approach for targeting operator's customers' base with a consideration of the involved relationships among its members. The similarity algorithm used to analyze subscribers pattern of similarity and relationship weight as to group them together and eliminate weak relationships. Inside every group, there are two types of influencers used for managing other group members, the first is the authority leader who is defined as the reference for every decision other members take and the dissemination

(hub) leader who is responsible for 'spread the word' to other group members. The two influencers should be used smartly by marketing owners or decision makers, the analysis of their strength and compare the traditional ways of targeting vs the new projected and scientific based model is also measurable. A well-organized SNA model is constructed to facilitate the solving process. It is our belief that SNA has reached the compromise and will be useful for many other cases as it has been in the past. In particular, SNA has broken through the management community to be widely used by CEOs. This widespread use is certainly due to its ease of applicability and the structure of SNA that follows the intuitive way in which decision makers solve problems. SNA is based on DW input of data and the use of SNA algorithms; the more accurate data and available we have is the more SNA model is reliable. It's the data mining models basis! Based on this research, one can see that SNA will solve existing cases in for the studied company that were proposed in this research for: (1) Reducing churn through targeting community leader and focusing the investment of offers on him instead of all community members. (2) For increasing product launch efficiency when introducing new product and testing it on a specific members- those who have the power to share the details- instead of random test. (3) For focusing on whole network ROI and network leader instead of other normal group members. At the end, it's better to target specific subscribers instead of targeting whole network. In this research, it is recommended to run SNA model every three months as a starting point, then the marketing owner will have the decision either to stretch or shorten this period. It's based on how customer base is increasing or decreasing, the acquisition of new subscribers or the churn of the existing. It differs in a company with a stable and mature state of base and with no changes in the market related to network expansion or launching new products than a company with a challenging market. This could be a reason behind the success of SNA model as well. As stated before, the purpose of this research is to develop a data mining model that will assist telecom network operators to understand their subscribers' base pattern and the way they connect. It's moving from individual view to network view which will enhance the offering methodology, reduce churn and increase new product launch efficiency. The analysis has focused on the process and ways of mapping the result of SNA model with the company's strategy. Incorporating traffic from other destinations, like competitors, and focusing on how the company's subscribers interact with competitors' base will assist in understanding the strategy of the latest and could be a future research objective along with incorporating more demographic variables like gender. Through shaping a competitor SNA model, the company will be able to measure competitors influence on its network along with their regional distribution which might help in predicting if, at someday, the competitors might acquire its base.

Acknowledgment

The author would like to thank Wael Abu Rezeq for his contribution to this work.

References

- [1] W. Barry, S. Berkowitz and eds., "Social Structures: A Network Approach,," Cambridge, 1988.
- [2] L. Freeman, "the Development of Social Network Analysis," Vancouver, 2007.
- [3] S. Wasserman and K. Faust, "Social Network Analysis: Methods and Applications," Cambridge , 1994.
- [4] W. AbuRezeq, The Impact of Social Network Analysis On the Palestinian Telecom Industry, Jawwal - Palestine, Nablus: An-Najah National University Masters' Thesis, 2017.
- [5] D. L. Passmore, Social network analysis: Theory and applications, 2011.
- [6] H. Fowler and A. Christakis, "Dynamic spread of happiness in a large social network: longitudinal analysis over 20," 5 December 2008.
- [7] K. Dasgupta, R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjee and A. A. Nanavati, "Social ties and their relevance to churn in mobile telecom networks," University of Maryland Baltimore County, India, 2008.
- [8] Y. Richter, E. Yom-Tov and N. Slonim, "Predicting customer churn in mobile networks through the analysis of social groups," in Proceedings of the SIAM International Conference on Data Mining SDM , Ohio, USA, April 29 - May 1, 2010.
- [9] Sonamine, "Whitepapers," 20 Jul 2010. [Online]. Available: http://www.sonamine.com/home/index.php?option=com_phocadownload&view=category&id=2:whitepapers&download=5:social-network-analysis-for-telecommunications-marketing&Itemid=1.
- [10] SPSS Inc., CRISP-DM 1.0, SPSS, 2000.
- [11] M. E. J. Newman, "Detecting community structure in networks," University of Michigan, Michigan, 2004.
- [12] Stanford, "Hubs and Authorities," 2008. [Online]. Available: <http://nlp.stanford.edu/IR-book/html/htmledition/hubs-and-authorities-1.html>. [Accessed 24 Jun 2016].
- [13] IBM Corporation, "Partitioning into groups," IBM Knowledge Center, 2019. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.0.0/sna_user_guide_project_ddita/sna/client/sna_overview_statistics_density.html#sna_overview_statistics_density. [Accessed 04 April 2019].
- [14] H. Tong, C. Faloutsos and J.-y. Pan, "Random Walk with Restart and Its Applications," Pittsburgh, 2006.
- [15] IBM Corporation, "Describing networks," IBM Knowledge Center, 2019. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.0.0/sna_user_guide_project_ddita/sna/client/sna_overview_statistics.html#sna_overview_statistics. [Accessed 05 April 2019].
- [16] IBM Corporation, "Network density," IBM Knowledge Center, 2019. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.0.0/sna_user_guide_project_ddita/sna/client/sna_overview_statistics_density.html#sna_overview_statistics_density. [Accessed 04 April 2019].
- [17] IBM Corporation, "About social network analysis," IBM Knowledge Center, 2019. [Online]. Available: <https://www.ibm.com/support/knowledgecenter/SS3J5>

- 8_9.2.0/com.ibm.i2.anb.doc/about_social_network_analysis.html. [Accessed 05 April 2019].
- [18] IBM Corporation, "Nodal degree," IBM Knowledge Center, 2019. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.0.0/sna_user_guide_project_ddita/sna/client/sna_overview_statistics_degree.html#sna_overview_statistics_degree. [Accessed 04 April 20169].
- [19] P. Devi, A. Gupta and A. Dixit, "Comparative Study of HITS and PageRank Link," International Journal of Advanced Research in Computer and Communication Engineering, vol. III, no. 2, p. 6, 2014.
- [20] IBM Corporation, "Describing groups and group members," 2019. [Online]. Available: https://www.ibm.com/support/knowledgecenter/SS3RA7_18.0.0/sna_user_guide_project_ddita/sna/client/group_analysis_describe.html. [Accessed 05 April 2019].
- [21] I. SPSS, "IBM SPSS Modeller," 2013. [Online]. Available: <http://www-01.ibm.com/software/analytics/spss/products/modeler/>. [Accessed 2 October 2015].
- [22] Y. Feng, Application of Data Mining in CRM System, 2007.