Paper 111-2010

# Creating a customer influence factor to decrease the impact of churn and to enhance the bundle diffusion in telecommunications based on social network analysis

Carlos Andre Reis Pinheiro, Markus Helfert, School of Computing, Dublin City University, Dublin, Ireland

## ABSTRACT

*Telecommunications' industry evolves into a high competitive market which demands companies to put in place an effective customer relation approach. Social network analysis can be used to increase the knowledge related to the customers' influence, arising relevant information to turns the customer experience better. SNA can be used to evaluate the customers' links and therefore clarify distinguishes aspects about the virtual communities inside the networks, allowing companies to deploy a more effective action plan to better diffuse their products/services and avoid churn.*

## THE INTERNAL SOCIAL NETWORKS BEHIND THE TELECOMMUNICATION'S ENVIRONMENTS

The main feature of any community is the relationship events between their members. Any kind of community is established and maintained based on this kind of relation events. According these brand new technologies and communication devices available currently, this type of characteristic have been increasing in relevance and becoming more evident in telecommunications scenery. From distinguish possibilities of communication and relationship, with flexibility and mobility, those communities gained new boundaries, creating new means of relationships and social networks. Based on different ways to communicate, people are ready to create new types of virtual networks. Due the dynamism and cohesion of these communities, included in a context highly technological, the people's influence can be significantly more strong and relevant to the companies.

Additionally, some kind of events can be faced as a chain process, which means that some specific points of the network can trigger a sequence of similar events due its influence and weight. The linkages between the customers can describe the way that the events will run and the weight of the linkages can represent the impact of the chain process in the network.

Understanding the way of these relationships occurs and recognizing the influences of some specific points inside the network could be a great competitive advantage, especially with respect to the event which could be performed in a chain process like churn or purchasing.

Due the almost unique relationship between the telephones and the individuals, and the influence which some person could exert over the others members of the community, the recognition of the characteristics of the social network existing inside the customer base is completely relevant to the telecom companies [1] and [2]. Discovering the central and strong nodes of those social networks and understanding how they behavior could be one of the most effective way to establish a new customer value, and to predict the impact of the churn and product adoption, but in chain process overview. Realizing the customers connections and identifying the central nodes of each chain inside the social network might be the best way to prevent a massive event of churn and hence a huge leakage of revenue. Analogously, understanding the strength of the central nodes and the way they influence the other nodes mighty be a good way to improve the production adoption process and hence increase substantially the revenue and the effectiveness of marketing campaigns.

## USING SOCIAL NETWORK ANALYSIS TO CREATE A CUSTOMER INFLUENCE FACTOR

Social network analysis can disclose the possible correlations among some particular business event, such as churn or purchasing, inside the community, proving the stronger impact when the event is triggered by a core node of the social network and the weaker influence when it is triggered by a boundary node. The influence means the number of customers which should follow the initial business event in a chain process.

Monitoring and analyzing the social network, particularly those which are interconnected, can allow companies to assess the revenue impact assigned to the business events in a chain process [3]. Based on the likelihood assigned to the churn's event and the level of influence related to the customer, companies can perform a more straight actions to avoid the chain's process initiation. Analogous can be performed to the purchasing event. Knowing the likelihood of bundle purchasing and the influence factor assigned to the customers is possible to establish a more accurate promotion to diffuse products and services.

Social network analysis in telecommunications can help companies to recognize the customer's behavior and then predict how strong are the links between the customers and the possible impact of the events among them. In this particular scenery is more important to retain an influential customer than a regular one. In fact, valuable here is the length of influence which means how big will be the chain process triggered. In the same way is more important to sell a bundle to an influential customer than to a regular one. The difference here is that the influential customer can lead other customers to churn or purchase more than a regular one.

Considering the hypothesis that there are a strong correlation between the churn and the purchasing events when they occurred inside a social network, one of the main yields of this approach is to evaluate the impact of each node in the social network, which means assess each customer in the community considering some particular business events.

## CUSTOMER INFLUENCE FACTOR MODELING

In order to assess the correlation among the churn and bundle acquisition events is quite important to analyze the events in a chain's perspective. Are the events related to each other? Are the correlation assigned to the events related to any customers' attribute? Are some correlations about the strength of the links and the business events?

Those questions can be addressed using social network analysis rather than traditional analytical models.

Due the huge complexity related to the human being relationships, the main challenge involving social network analysis is the capacity to recognize the patterns of behaviors among the individuals and the impact of each event can hold in terms of individual's influences.

The knowledge about the length of the customers' influence can be used to define a new corporate value allowing companies to establish distinguish loyalty and selling campaigns. This new perspective can change substantially the way the companies manage their customers, evaluating them based on the influence in business events' correlation instead their isolated revenue information. Particularly in the telecommunications' market this new approach can be tremendous relevant due the natural social networks hidden inside the data.

In order to highlight the influential customers against to the regular ones a social network approach will be performed over the data, basically the call detail records. A distinguish differentiation among the customers can be raised by this technique, identifying the customers' influence and hence the kind of business action which should be performed in terms of customer relation or marketing campaigns. The influence factor can reveals the customers who are able to trigger some particular events in a chain perspective. This kind of knowledge is more relevant in a telecommunication's scenery than traditional attributes and makes the companies able to execute more effective actions for business processes.

The most traditional measures to depict a social network are the first order centrality and the second order centrality [4]. The first measure describes the number of straight connections from some particular node. The second one describes the number of connections the nodes related to the original one has [5].

Those two measures of the network topology will be use to compose the customer influence factor once they represent the how many customers a particular one can straightly reach and how many customers can be reached indirectly.

Considering a particular node, once the first order centrality presents the amount of nodes directly connected with it, this measure is quite relevant for those events where a straightly influence is involved. It is well known as a number of friends a particular customer has and can straightly influence. The event of churn could be considered as this kind of event. The second order centrality shows the number of nodes indirectly connected with it and is more related to events where a straightly influence is not required. Events which require spread diffusion such as bundle acquisition or product adoption could be considered as this kind of event. Analogously, this one is well known as how many friends the friends of a particular customer have. Both types of events will be focus of examples onward.

Even though those two measures about social network analysis are the most regular and commonly used, the majority of applications using social network analysis technique do not consider two directions of relationship [6]. When it is applied for co-authorship or friendship networks, the edges between the nodes really do not require a bidirectional vector. If a node A is friend of a node B, node B is also friend of node A. The link between them doesn't need a direction to establish who is friend of whom. They are simply friends. The same thing happens with the co-authorship. In spite of the order of appearance, authors publish together and in that way, the direction of the edge is not required indeed. They simply publish in conjunction.

However in the telecommunications' scenario is quite relevant to consider the direction of the relationship [7]. Some node A, or in that case, a customer A, is not just connected to node B, or to the customer B. In telecommunications the customers A hold some connections with the customer B and the customer B can holds another sort of connections with the customer A. The bidirectional vector should be taken into consideration. In this case, the customer A can make 10 times more calls to the customer B than the customer B call to the customer A. This differentiation in the edge should represent that customer A is more active in terms of calling maker than the customer B. Customer B in this particular case is more a calling receiver than the customer A. In terms of network activity, and also in terms of revenue generation, customer A could be considered more important than the customer B.

Additionally the two different directions in relation with the connections have distinct values in the telecommunications scenario. The incoming calls hold a distinct value than the outgoing calls. In that way, when the customer make calls a different weighting should be consider in comparison when the customers receive calls. Once one particular customer can make and receive calls inside the network, both directions of calls should be taken into consideration but using distinct weights in the final influence factor calculation.

Therefore the formula created to calculate the customer influence factor should considers the first and the second order centralities for the incoming and outgoing calls distinctly. The number of calls and the total duration between the pairs of customers connected are being considered as well as a way to establish the strength of the relationship between the customers. A relation of the total duration over the number of calls between a pair of customers can be used alternatively as a way to establish the strength of the relationship between them.

Therefore the customer influence factor is taking into consideration the way the customers are originating their calls, providing a good sense of knowledge how they play inside the social network as a call maker.

In this way, three distinct attributes are being considered to calculate the customer influence factor, the first order centrality, the second order centrality and a relation of the amount of call over the total duration. All those attributes are related to the originating calls only, depicting the way the customers make calls inside the social network.

Analogously, the same attributes are taking into consideration when the customers receive their calls, depicting the way they play as a call receiver.

Hence, the first order centrality, the second order centrality and also a relation between the amount of calls and the total duration are included in the customer influence factor formula in order to represent the customers' behavior as a receiver calls inside the social network.

However, the outgoing calls hold a distinct value than the incoming calls when some corporate dimensions are being considered. In terms of revenue for instance, the outgoing calls have a higher value for the company than the incoming calls. In this way, the first order centrality, the second order centrality and the relation between the number of calls and the total duration are being considered with different weights, due to their value for the company.

Once the important thing here is to establish a differentiation between the incoming calls and the outgoing ones, it is possible to use distinct type of weights such as the retail price of calls, the cost, the profit, and so on. Due the high complexity in order to establish some particular measures like the cost of the call or even worst the profit, it is easier to use simply the cost of the calls to define relation between those two types.

Besides the relation between the values of the calls, could be interesting to consider also the amount of calls in relation to those particular two types mentioned previously. The relation between the values of the calls in this way should take into consideration the value and the amount of calls for each one those two types.

It is important to notice that the values of the calls vary according to the time of the day and also to the workdays and weekends.

In this particular case study, considering all those types of differentiation, the relation between the incoming calls and the outgoing ones is about 12.215. This means that all components in relation to the incoming calls, such as the first order centrality, the second order centrality and the relation between the number of calls and the total duration should be divided by 12.215 in order to differentiate the two types of calls in terms of corporate value.

Finally, one additional factor is included in the customer influence formula. Once the content assigned to the attributes can be quite disperse, due the highly distinction among the customers' behavior, is quite common find a huge difference of values between two particular customers in terms of the number of originating calls, receiving calls, or total duration. In order to smooth those differences, making the whole measures more normalized it was used the coefficient of variation for some specific factors inside the formula.

Therefore, the coefficient of variation was applied to the first order centrality, the second order centrality and the relation between the number of calls and the total duration in order to normalize the magnitude of the measures. This process allows distinct measures such as first and second order centralities as well as the relation between the number of call and the total duration to be compared and used n a single formula.

The social network related to the telecommunications companies holds internal and external connections. For instance, considering just the residential customers from a particular telecom operator, they communicate with other residential customers but with telephones from other distinct operators as well. Although the residential customers can exert influence over another residential customers all network should be considered in order to estimate the real length of the customer influence. In that way, the relations between residential customers have some type of weight and the other connections, between residential and non residential customers, including the ones from outside the network, have another weight in the influence factor formula.

Summarizing the calculation process, the first and the second order centralities are taking into account as well as the sum of calls and the total of duration. These measures are calculated for both internal residential network and entire network. In addition, those measures are calculated in separately considering the incoming and outgoing calls. For all the measures assigned to the incoming network a multiplicative factor is applied in order to differentiate the value of incoming and outgoing calls.

Finally, aiming to normalize the magnitude of the measures so it should be possible to use them in a massive process, considering the entire customer base, a coefficient of variation is applied to some particular components in the formula. This is performed to decrease the dispersion of the metrics in more applicable scores for real world business problems, allowing them to be used into a single formula.

The customer influence factor is calculated in a monthly basis and all measures are established considering the mean of the last four months. This is done to discard the outlier numbers and also to decrease the impact of some particular peaks and troughs in the average curve of the customers' behaviors.

The customer influence factor is used in a corporate perspective in order to improve the customer loyalty process and also the bundle diffusion initiative. Considering a time window of 6 months in the past in relation of events of churn and bundle purchasing, a set of correlation measures have been established to proof the relationship between the customer influence factor and some business events in particular.

## CORRELATION BETWEEN THE CUSTOMER INFLUENCE AND THE PAST EVENTS OF CHURN

Using the customer influence factor measure based on the last four months of network data, and considering the onwards six months to establish the correlations, the following figures is about the performance of the customer influence in terms of the churn events.

The average customer influence factor was calculated to 769,104 residential customers. In a particular month, 10,624 have left the company, performing the churn event in that particular case.

Taking 1,000 residential customers randomly, they relate with another 5,076 distinct phones considering the entire network, which 1,262 are residential customers of the telecommunications company in that study. Considering the subsequently forward three months, 18 residential customers from those 1,262 have left the company as well. In other words, those 1,000 random residential customers usually relates with another 1,262 residential customers, all the ones they can exert some sort of influence. From these relationships, just 18 residential customers have followed them in the same event, the churn, in the subsequently months. We are considering in this study that those 18 residential customers who made churn have been led by those random thousand original residential customers in the chain's process in some specific way.

Taking the same amount of 1,000 residential customers, but ranked now by the average revenue in the company, they relate themselves with another 6,955 distinct phones, which 1,156 are residential customers of the same telecom operator. From these relationships, 19 residential customers left the company in the subsequently three months forward.

Now, considering the same amount of one thousand residential customers, but ranked by the influence factor, they relate with another 16,991 distinct phones which 5,018 are residential customers who belongs to the same telecom company. From that amount, 130 residential customers have made churn in the subsequently three months forward.

Considering the almost eleven thousand residential customers who left the company in a particular month, a random thousand residential customers from that amount have led additional 18 residential customers to make churn. The top thousand residential customers according the average revenue from that same amount have led additional 19 residential customers to make churn. And finally, the top thousand influential residential customers from the same eleven thousand residential customers who left the company have led additional 130 residential customers to make churn.

In terms of capacity to span their actions, considering the subset of thousand residential customers who have left the company in particular month, from the random process, each 56 residential customers who left the company have led another 1 to follow them in the churn event. From the revenue ranking, each 53 residential customers who left the company have led another 1 residential customers to follow them in the churn event. And finally, from the influence factor ranking, each 8 residential customers who left the company have led another 1 to follow them in the event of churn.

In terms of business actions, if the company intends to deploy a retention process, it should be aware that each 56 residential customers in average will affect or lead another one residential customer to make churn event. And each 8 influential residential customer will affect or lead the same another one residential customer to make churn. That is with no doubt a big difference in terms of span when the subject it retain the customer or increase their loyalty.

Although the performance in terms of span is quite bigger when considered the influential residential customers, 622% higher than the random process, the hit rate of the influence factor is also more effective.

Certainly, as the influential customers relate with a higher number of other customers, should be expected that they influence more individuals in absolute terms. In another perspective, it is expected that a higher number of related customers make churn when they are related with the influential customers than they are related with a random or average customers. Comparing both subsets of customers, there are more than five thousand customers related to the influential ones and just a little bit more than one thousand customers related to the random or average ones. It should be expected that from that five thousand related customers a higher number of individuals make churn in comparison with the one thousand subset.

In spite of the higher absolute number of subsequently events associated with the influential customers, they also have a better performance in a relative analysis. Even though taking into consideration those five thousand related customers, one hundred and thirty made churn in the subsequently months, which represents 2.6% of the possible customers to be affected. Considering the random subset of customers, from the one thousand related customers, just eighteen customers made churn in the subsequently months, representing just 1.3% of the possible affected customers.

Taking into consideration the relative performance, from each seventy related customers assigned to the random subset, just one is affected by the initial event of churn. However, considering the related customers assigned to the influential ones, from each thirty nine possible customers to be affected, one is influenced to follow the initial event of churn. This represents a performance 81% better than the random or average customers.

## CORRELATION BETWEEN THE CUSTOMER INFLUENCE AND THE BUNDLE DIFFUSION PROCESS

According an analogous procedure, but considering the bundle purchasing event instead the churn occurrence, it was performed some correlation analysis in respect to the customer influence factor performance.

As mentioned previously, the customer influence factor was calculated to 769,104 residential customers. In a particular month, 20,480 have purchased a bundle of particular services of the company.

Similarly, taking randomly 1,000 residential customers from those who have purchased some bundle, they relate with another 29,216 distinct phones considering the entire network, which 6,847 are residential customers of the telecommunications company in that particular study. Considering the subsequently forward three months, another 885 residential customers from those 6,847 residential customers related have purchased some bundle as well. In other words, those 1,000 random residential customers usually relates with another 29,216 residential customers, all the ones they can exert some sort of influence, and from those relationships, 885 another residential customers have followed them in the same event, the bundle purchasing, in the subsequently months. Analogously, we are considering in this study that those 885 residential customers who have purchased some bundle were led in some way by the original random thousand residential customers in this chain’s process.

Taking the same amount of 1,000 residential customers, but ranked now by the average revenue, they relate with another 47,041 distinct phones considering the entire network, which 9,235 are residential customers of the same telecom. From these relationships, 1,165 residential customers have purchased some bundle in the subsequently three months forward.

Now, considering again the same amount of one thousand residential customers, but ranked by the influence factor, they relate with another 64,366 distinct phones considering the entire network, which 21,558 are residential customers who belongs to the same company. From that amount, 6,454 residential customers have purchased some bundle in the subsequently three months forward.

Considering the almost twenty seven thousand residential customers who have purchased some bundle in a particular month, a random thousand residential customers from that amount have led additional 885 residential customers to purchase some bundle. The top thousand residential customers according the average revenue from that same amount have led additional 1,165 residential customers to purchase some bundle. And finally, the top thousand influential residential customers from the same twenty seven thousand residential customers who purchased some bundle have led additional 6,454 residential customers to purchase some bundle onwards.

Analogous to the churn event process, the bundle purchasing event have presented a good performance in terms of capacity to span, considering the same subset of thousand residential customers who have purchased some bundle in particular month. From the random process, each 10 residential customers who purchased some bundle have led another 9 residential customers to follow them in the bundle acquisition. From the revenue ranking, each 10 residential customers who purchased some bundle have led another 12 residential customers to follow them in the bundle purchasing event. Finally, from the influence factor ranking, the same each 10 residential customers who purchased some bundle have led another 65 residential customers to follow them in the same event of acquisition.

Again, in terms of business actions, if the company intends to launch a bundle diffusion campaign, it should be aware that each 10 residential customers in average will affect or lead another 9 residential customers to acquire some bundle onwards. However, the same each 10 influential residential customer will affect or lead 65 additional residential customers to acquire some bundle on the forward months. That is with undoubtedly a huge distinguishing in terms of span when the bundle diffusion process matters.

Although the performance in terms of span is significantly bigger when considered the influential residential customers, 629% higher than the random process, the hit rate of the influence factor is also more effective when the event is analyzed in a perspective of chain.

Considering the 885 residential customers from the 6,847 related customers who have followed the original thousand random purchaser customers, the hit rate are 13%. In other words, from the possible 6,847 residential customers who possibly could be affected in a chain process such as bundle acquisition, 885 residential customers were influenced indeed. In the same way, considering the 1,165 residential customers from the 9,235 related customers who have followed the original top thousand purchaser customers in terms of revenue, the hit rate is also 13%. However, considering the 6,454 residential customers from the 21,558 related customers who have followed the original thousand influential purchaser customers, the hit rate is about 30%.

Taking into consideration the relative performance, from each eight related customers assigned to the random subset, just one is affected by the initial event of bundle acquisition. However, considering the related customers assigned to the influential ones, from each three possible customers to be affected, one is influenced to follow the initial event of bundle acquisition. This represents a performance 132% better than the random or average customers.

## CONCLUSION

Telecommunications market is quite dynamic, and some particulars scenarios change very often. When the market changes the data related to it changes as well. The social network analysis model should be monitored and assessed to adapt itself to new business realities, which will be represented through new type of data or data content.

However social network analysis are quite adaptable to different changes which happen in data, pursuing the customer's behavior in terms of actions, usage, consuming and relationships. The adaptable feature is fundamental in a market characterized by high competition, as telecommunications. In this market the conditions in respect to the customer's behavior can change very rapidly and should track to follow the new business needs.

The dynamical characteristic in relation to the customers should be reflected on data and hence the model basis on data is able to recognize this sort of change.

Ranking the customers based on their influence factor rather than a set of isolated attributes make possible for telecommunications companies establishing and performing business actions in a straightforward way to retain more customers with less effort and also to diffuse some bundles of products and services with less cost of operation.

The innovation factor here is indeed the approach to establish the value assigned to the customers, considering their influence's attributes and their relations' weights more than their isolated characteristics. Taking into account events which happens in a chain process such as churn and bundle diffusion, this approach may represent a completely distinguish way to accomplish some important business goals.

The traditional way to evaluate the customers is usually according to their billing, demographic information or even based on their behavior, using clustering or segmentation models. However, due the virtual communities created within the telecommunications' networks, it is mandatory to establish a distinct manner to value the customers, considering now their importance and influence in those communities. Following this approach companies will be able to retain more than just the high value customers but instead, they will maintain the relations inside the networks which means more products and services usage. Analogously, targeting the customers for a marketing campaign to diffuse some new bundle based on the customer influence factor can allow companies to span their products and services in the best way, through their own customers relationships rather than based on selling procedures.

According to the past events the customer influence factor has presented additional gains in terms of business effectiveness and operational performance. The target process to select the best customers to trigger a particular campaign was 81% better in terms of hit rate using the influence factor than a random process. Also, when the target is based on the influence factor the hit rate is 132% better than the random process.

Nevertheless the gains in terms of effectiveness are quite relevant, the performance in terms of absolute numbers, which is, the capacity to span a particular business process is even better. The usage of the influence factor to execute a retention process has presented a performance 622% better than the random approach, which means, is possible to reach seven times more customers in a retention process than using the random approach. Analogously, the performance of the influence factor in a bundle diffusion process is 629% better than the random approach. Starting from the same amount of customers, it is possible to reach again seven times more customers to diffuse some particular bundle when using the customer influence factor than the random process.

## REFERENCES

1. Degenne, M. F.: Introducing Social Networks, Sage Publications, London (1999).
2. Knoke, D., Yank, S.: Social Network Analysis, Sage Publications, London (2007).
3. Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Navati, A.: Social ties and their relevance to churn in mobile telecom networks. In: 11th International Conference on Extending Database Technology: Advances in Database Technology. No. 261, pp. 668-677 (2008).
4. Carrington, P. J., Scott, J., Wasserman, S.: Models and Methods in Social Network Analysis, Cambridge University Press, New York (2005).
5. Wasserman, S., Faust, K.: Social Network Analysis: methods and applications, Cambridge University Press, New York (1994).
6. Kempe, D., Kleinberg, J. Tardos, E.: Maximizing the spread of influence through a social network. In: Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 137-146 (2003).
7. Seshadri, M., Machiraju, S., Sridharan, A.: Mobile call graphs: beyond power-law and lognormal distributions. In: 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pp. 596-604 (2008).

## CONTACT INFORMATION (HEADER 1)

Your comments and questions are valued and encouraged. Contact the author at:

Name: Carlos Andre Reis Pinheiro, Markus Helfert
Enterprise: Dublin City University, School of Computing
E-mail: [cpinheiro, markus.helfert]@computing.dcu.ie