SOCIAL NETWORK ANALYSIS EVALUATING THE CUSTOMER'S INFLUENCE FACTOR OVER BUSINESS EVENTS

Carlos Andre Reis Pinheiro[1] and Markus Helfert[2]

[1]School of Computing, Dublin City University, Dublin, Ireland
cpinheiro@computing.dcu.ie
[2]School of Computing, Dublin City University, Dublin, Ireland
markus.helfert@computing.dcu.ie

ABSTRACT

The telecommunication industry has evolved into a highly competitive market, which requires companies to put an effective customer relation approach in place. In order to increase customer relationship management, social network analysis (SNA) can be used to increase the knowledge related to the customers' influence. SNA can improve relevant information that helps to increase the customer experience. It can be used to evaluate the customers' relations and therefore clarify distinguishing aspects about the internal communities inside the entire network, allowing companies to deploy a more efficient action plan to better diffuse their products/services and avoid the customer's churn. In this paper we present an approach to evaluate the customer influence factor related to business events.

KEYWORDS

Social Network Analysis, Pattern Recognition, Customer Behaviour, Data Mining, Knowledge Discovery

1. THE INTERNAL SOCIAL NETWORKS BEHIND TELECOMMUNICATIONS

One of the most important characteristic of any type of community is described by the kind of the relationship among its members. Communities are usually established and maintained due to the kind of relationships. However, as the availability of new technologies and communication devices increased, the types of characteristics of communities have been increasingly relevant. This has becoming even more evident in the telecommunications sector, with many new possibilities of establishing relationships among customers. New communities have been established and have exceeded traditional boundaries. This has increased the flexibility and mobility and has often created new meanings of relations among members, and thus establishing different forms of social networks. Due to the dynamics of communities, framed in a context of highly evolving technologies, estimating the customers' influence is extremely relevant to telecommunications' companies.

In addition to the importance of business communities, some types of business events are occurring as a chain process, where some specific points of the network can trigger a sequence of similar events according to its influence and weight. Also, relations among customers can describe the way that events will occur within the network, as well as the weights of the links can foresee the impact of the chain process. By understanding the way these relations take place and recognizing the influences of some specific nodes inside network, we can provide significant business knowledge. This is especially interesting in relation to events which occur in a chain perspective such as churn or purchasing [9].

DOI : 10.5121/ijaia.2010.1410

Due to the particular correlation between phones and individuals, and the influence which some person could exert over others, the pattern recognition for social networks inside customers' bases is particular relevant to telecom companies [1] [2]. Analysing these virtual communication networks, and discovering central and strong nodes, can contribute to significantly increase the understanding about customers' behaviour. Understanding this behaviour is one of the most effective approaches in order to establish a new customer's value, allowing companies to predict the impact of churn and product adoption. In our research we have considered this behaviour in a chain perspective. By observing the customers' relations and identifying the central nodes of each social network appears to be a suitable way to prevent a massive event of churn or substantially increase the product diffusion.

2. USING SOCIAL NETWORK ANALYSIS TO CREATE A CUSTOMER INFLUENCE FACTOR

Social network analysis can identify possible correlations among particular business events within the communities, such as churn or purchasing. The correlation indicates that the impact is higher, when the event is triggered by a customer with a high influence value [10]. Influence represents the number of customers, which are expected to follow the initial event in a chain perspective; or in other words the number of customers who will follow the initiator's behaviour when a churn or buy event occurs. By monitoring and analyzing the social network, we can assess the revenue impact in relation to the business events [3]. Thereby are those relations, which are interconnected of particular interest. Based on the likelihood established to the churn's event and the level of influence related to the customers, companies are able to determine suitable actions to avoid the chain's process initiation, either of churn of bundle diffusion.

Social network analysis in telecommunications can help companies to recognize the customer's behaviour and therefore predict the strength of relations among customers. Furthermore, it is possible to identify the impact of events within networks. In this particular scenario it is more important to retain an influential customer than a regular one, regardless how much they contribute in terms of revenue. In fact, in a network perspective, value represents here the level of influence, which describes the extent of events in a chain process. Analogously, it 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.

3. CUSTOMER INFLUENCE FACTOR MODELLING

In order to evaluate the correlation between churn and bundle acquisition events it is important to analyze events in a chain's perspective. Due to the huge complexity related to the relationships, the main challenge involving social network analysis is the capacity to recognize the patterns of behaviours among individuals and therefore the impact of each event in terms of individual's influences. The knowledge about the length of the customers' influence can be used to define a different value perspective allowing companies to establish distinguished loyalty and selling campaigns. This new perspective can change substantially the way companies manage their customers, evaluating them based on the influence in terms of business events' correlations in contrast to the traditional focus on isolated revenue information. Particularly in a telecommunication market this new approach can be tremendous relevant due the natural social networks hidden in the data.

In order to compare the influential customers with regular ones we have employed a social network analysis approach over call detail records. We also included some additional demographic and financial information. A distinguished differentiation among customers can be highlighted by this technique, identifying the customers' influence and hence the sort of business' actions that would be performed in terms of a customer relationship campaign. The influence factor can reveal the customers who are able to impact on others when a particular business event is triggered in a chain perspective.

Due to the importance of relations in communities, we used first and second order centralities, two basic network metrics, to describe the network [4]. The first measure describes the number of direct connections from a particular node. The second measure describes the number of connections a node has that is related to the original one [5]. The two measures are used as main element to calculate the customer influence factor. It represents the number of customers a particular one can directly reach as well as how many customers can be reached indirectly. For each node, the first order centrality presents the amount of nodes directly connected with it. This measure is relevant for those events where a direct influence among customers is involved. The second order centrality describes the number of nodes indirectly connected with it and is more related to events where a direct influence is not obviously involved. Events which require spread diffusion such as bundle acquisition or product adoption would be suitable to this type of event.

Even though those two measures are well established in social network analysis and are 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 links among the nodes usually do not require a bidirectional vector. If a node A is friend of a node B, hence node B is also friend of node A. The link between them does not need a direction to establish the "friend relationship". They are simply friends. The same issue occurs with representing co-authorship. Regardless of the order of appearance, authors publish together and therefore, the direction of the edge is not required. They simply publish in conjunction.

In contrast to these typical application areas, the direction of the relationships is relevant in telecommunication applications [7]. A particular node A, or in that case, a customer A, is not just connected to a node B, or to a customer B. In telecommunications, customer A is connected to customer B and customer B may be connected to customer A. These two distinct connections between customers A and B may be different in terms of recency, frequency and strength. Therefore, a bidirectional vector should be considered. For instance, customer A can make 10 times more calls to customer B than customer B to customer A. This differentiation in the connection should represent that customer A is more active in terms of calls than customer B. Customer B is more a calling receiver than customer A in this particular case. In terms of network activity, and also in terms of revenue generation, customer A could be considered more important than customer B. In telecommunications, representation of the social networks can be quite dynamic [8], where leaders and followers might swap roles over the time, leading in some periods and following in another.

Additionally, the two different directions of those connections can hold distinct values in the terms of revenue. The incoming calls have 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 on 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 relationship among 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 among them.

The customer influence factor takes into consideration the way the customers originate 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 were considered in order to calculate the customer influence factor, which are, the first order centrality, the second order centrality and a relation of the amount of calls over the total duration. All these attributes are assigned to the originating calls only, depicting the way the customers make calls inside the social network. Analogously, the same attributes were considered when the customers receive their calls, describing 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' behaviour 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.

One important issue here is to establish a differentiation between the incoming calls and the outgoing ones. With this 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 it is valuable 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 weekdays and weekends.

This approach was applied to a big European telecommunications company, reaching very reasonable results in terms of business. 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. The content assigned to the attributes can be quite disperse. For instance, due the highly distinction among the customers' behaviour, it is quite common to 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 and making the whole measures more normalized we 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 calls and the total duration. The measures can be compared and used in 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, we assign a particular weight for the relations between residential customers as well as for the other connections, between residential and non residential customers, including the ones from outside the network.

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 that it will 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 on 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' behaviours.

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.

4. 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, in the following we illustrated some figures about the performance of the customer influence in terms of the churn events.

The average customer influence factor was calculated based on 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 other residential customer to 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 significant bigger when considered the influential residential customers, 622% higher compared with the random process, the hit rate of the influence factor is also more effective.

Certainly, as the influential customers relate with higher number of other customers, we should also expect that they might influence more individuals in absolute terms. In another perspective, it is expected that a higher number of related customers may leave when they are related with influential customers than they are related with average ones. Comparing both subsets of customers, there are more than five thousand customers related to the influential ones and just a slightly more than one thousand customers are related to the averages. It should be expected therefore that from these five thousand related customers a higher number of individuals leave when compared to the subset of one thousand regular customers.

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, 130 churned 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 70 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 39 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.

5. Correlation between the Customer Influence and the Bundle Diffusion Process

We applied our approach to similar scenario, by considering the bundle purchasing event instead the churn occurrence. We performed 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 27000 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

Over the last year, the telecommunications market is characterised by a high degree of dynamic. When the market changes the data related to it changes as well. In this study we provided a SNA approach related to the telecommunication sector. We provided insight into the model as well as its application. However, the social network analysis model should be monitored and assessed to adapt itself to new business realities, which will be represented through new types of data or data content.

Approaches of social network analysis are quite adaptable to different changes that happen in data, pursuing the customer's behaviour 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 behaviour 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 enables telecommunications companies to establish and perform business actions in a manner to retain more customers with less effort and also to diffuse some bundles of products and services with less cost of operation.

In contrast to considering isolated characteristics, our research indeed presents a noevel approach in that regard that the value is establish by assigning valued to the customers, considering their influence's attributes and their relations' weights. Taking into account events that occur 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 behaviour, 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 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 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.

ACKNOWLEDGEMENT:

This project has been partly funded by Enterprise Ireland (http://www.enterprise-ireland.com/) under the Innovation Partnership Scheme.

REFERENCES

[1] Degenne, M. F. (1999) *Introducing Social Networks*, Sage Publications, London.

[2] Knoke, D., Yank, S. (2007) *Social Network Analysis*, Sage Publications, London.

[3] Dasgupta, K., Singh, R., Viswanathan, B., Chakraborty, D., Mukherjea, S., Navati, A. (2008) Social ties and their relevance to churn in mobile telecom networks, *11th International Conference on Extending Database Technology: Advances in Database Technology*, No. 261, pp. 668-677.

[4] Carrington, P. J., Scott, J., Wasserman, S. (2005) Models and Methods in Social Network Analysis, Cambridge University Press, New York.

[5] Wasserman, S., Faust, K. (1994) Social Network Analysis: methods and applications, Cambridge University Press, New York.

[6] Kempe, D., Kleinberg, J. Tardos, E. (2003) Maximizing the spread of influence through a social network, *Ninth ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 137-146.

[7] Seshadri, M., Machiraju, S., Sridharan, A. (2008) Mobile call graphs: beyond power-law and lognormal distributions, *14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 596-604.

[8] Kathleen M. Carley. (forthcoming) Dynamic Network Analysis, *Summary of the NRC workshop on Social Network Modeling and Analysis*, Ron Breiger and Kathleen M. Carley (Eds.), National Research Council.Dynamic Network Analysis.

[9] Papadopoulos, S., Skusa, A., Vakali, A., Kompatsiaris, Y., Wagner, N. (2009) Bridge Bounding: A Local Approach for Efficient Community Discovery in Complex NetworksData Analysis, *Statistics and Probability (physics.data-an); Physics and Society (physics.soc-ph)*.

[10] Porter, M., Onnela, J., Mucha, P. (2009) Communities in Networks, Statistics and Probability (physics.data-an); Physics and Society (physics.soc-ph).

Authors

Carlos Andre Reis Pinheiro is Affiliated Researcher at the School of Computing, Dublin City University, Ireland. He accomplished a Post Doctorate research at the Business Informatics Group, DCU. He holds a Doctor in Computer Sciences from the Federal University of Rio de Janeiro, Brazil. Before joined DCU he worked at Brasil Telecom and accomplished a Post Doctorate research at IMPA, Brazil

Markus Helfert is a Lecturer in Information Systems and Head of the Business Informatics Group at the School of Computing, Dublin City University, Ireland. He holds a Doctor in business administration from the University of St. Gallen, Switzerland, a Master-Diploma in Business Informatics from the University Mannheim, Germany, and a Bachelor of Science from Napier University, Edinburgh, UK-Scotland.