Alternatives for Telco Data Network

Colot, Christian; Baecke, Philippe; Linden, Isabelle

Published in:
Information Systems Management

DOI:
10.1080/10580530.2021.1887981

Publication date:
2021

Alternatives for telco data network: the value of spatial and referral networks for churn detection

Christian Colot[a,*], Philippe Baecke[b], Isabelle Linden[c]

[a]***Dr Christian Colot*** *is a Postdoctoral Researcher at the University of Namur in Belgium, Department of Business Administration. He has recently finished his PhD in Management sciences from the University of Namur. He also holds a Master degree in Information System from the University of Lille and Masters degrees in Statistics and Management sciences from UCLouvain. Christian is member of the Namur Digital Institute. His research interests include Big Data Analytics, Customer Relationship Management and Mobile data. His work has appeared in several international peer reviewed conferences.*

[b] ***Prof. Dr Philippe Baecke*** *is an Associate Professor at Vlerick Business School (Belgium) and Adjunct Lecturer at Trinity Business School (Ireland) with a strong expertise in big data analytics. From a research perspective, Philippe focuses on improving business insights by creatively incorporating new data types, such as geographical and social network data. In addition, he has research focus on ad targeting and measurement. His research has been published in several peer reviewed journals. Over the past years he has assisted several companies both on a strategic and operational level, with their data & analytics strategy and projects.*

[c]***Prof. Dr Isabelle Linden*** *is a Professor of Information Management at the University of Namur in Belgium. She obtained her PhD in Computer Sciences from the University of Namur. She also holds Masters degrees in Philosophy and in Mathematics from the University of Liège, Belgium. She is member of the FoCuS Research Group of the Namur Digital Institute. Combining theoretical computer science and business administration, her main research domain regards information, knowledge and artificial intelligence. She explores their integration within systems within systems that support decision making. Her works can be found in several international edited books, journals and conferences.*

[*]Corresponding author. Email: christian.colot@unamur.be

Abstract

The value of communication network has received significant attention in the literature on churn prediction, while little is known about the potential business value of alternative networks. This knowledge would help telephone companies to make timely strategic decisions in our evolving economic environment where traditional communication technologies are declining. This study assesses to which extent two alternative networks might (1) structurally substitute this network and (2) complement this network for churn prediction within telephone companies.
Keywords: Churn, Social Network Analysis, Business value, Data acquisition, Data Analytics

1. Introduction

While social network analysis (SNA) exists for almost 80 years (Granovetter, 1977), it has received in the last decade a growing interest from companies and the scientific community (Church et al., 2015; Goel and Goldstein, 2014; Hill et al., 2006). This trend is caused by multiple reasons. On the one hand, along with the development of large capacity and fast computing hardware, SNA can now be applied to much larger communities (Onnela et al., 2007). On the other hand, the emergence of device mediated social networks has given the opportunity to collect network data on a larger scale. This new network data has received special attention from companies due to its potential of improving various facets of its activities such as fraud detection (Van Vlasselaer et al., 2015), risk mitigation (De Cnudde et al.,

2019) and marketing (Bilgicer et al., 2015; De Matos et al., 2014a; Ma et al., 2014; Verbraken et al., 2014).

Specifically for telephone companies, calls and text messages, stored as call detail records (CDRs), can be converted into connections between customers leading to a communication network. This social network has proven to be particularly valuable in churn detection support systems (Dasgupta et al., 2008; Kim et al., 2014; Owczarczuk, 2010). While, for telephone companies, communication networks created based on call detail records have received significant interest in the churn literature, little is known about the value of alternative networks. On the short term, exploring alternative networks would help telephone companies to assess if a communication network is the only appropriate network to reflect the interdependence of behaviors among customers. Moreover, more fundamentally, investigating these networks is also likely to disclose long term alternatives to a communication network in an evolving business environment where traditional communication technologies such as text messages and mobile calls are declining due to the emergence of over-the-top services such as Whats'App, Messenger and WeChat (Farooq and Raju, 2019; Stork et al., 2017; Sujata et al., 2015). For instance, the number of messages generated by these applications exceeds the number of text messages operated by telephone companies since 2012 and reached a volume 4.45 times higher in 2016 (Legos, 2019). This gives rise to a real interest in investigating to investigate the value of alternative networks compared to the communication network.

In particular, past literature have proven that spatial (Yang and Allenby, 2003) and referral network (Van den Bulte et al., 2018) can deliver valuable customer insights for other marketing issues (e.g. customer acquisition). These 2 networks are particularly interesting to compare with a frequently used communication network as they might be developed inside a telephone company. Furthermore, they possess inherent characteristics that might be relevant for churn detection. First, a spatial network connects customers who live close to each other. This network comes at a low cost since it only requires customers'addresses to be collected. This network may capture some similarity traits across neighbors and potential local shocks which might affect neighbors in the same way such as a local advertisement of a competitor. Second, a referral network can also be developed by a telephone company by means of a referral program that tracks when the customer recommends the product or service. This second alternative network requires a higher cost, but it is particularly interesting in a churn setting as, by design, connected customers already communicated about the product/service in the past. This study contributes to the existing literature by assessing to which extent these 2 networks might (1) substitute CDRs-based networks on the long run and (2) complement CDRs-based networks on the short run for telephone companies in the churn identification processes. What is more, this study also investigates the analytical maturity that would be required by the telephone company to leverage these networks.

Based on our results, we find that the communication network currently remains the network the most appropriate to disclose insights on customer churn. However, if this network would become obsolete due to the competition with over-the-top services, a referral network could substitute this network as long as the telephone company is able to build advanced machine learning models. For telephone companies, these results suggest that they should start initiating referral programs and also developing their analytical skills in order to develop their long term resilience. Compared to related studies on referral networks, this paper not only confirms the value of this network for detecting churn within telephone companies but also highlights that it is able to timely replace, in an analytical mature organization, the organic communication network with a comparable value. For academia, this contribution highlights the opportunity to analyze multiple networks simultaneously either in combination or in substitution. In particular, the dilemma of a company outside the telephone industry having to choose between developing an internal network or buying an external network to leverage network effects seems a promising research avenue.

The remaining part of the paper proceeds as follows. Section 2 reports related works on customer churn prediction, spatial networks and referral networks. Section 3 discusses the empirical data from a European telephone company used in this study. Notably, three networks are retrieved and compared from this data unlike other studies on churn mentioned in the related work section where only one network is investigated. Section 4 explains the modeling techniques and evaluation metrics. Section 5 presents the findings of this study. Finally, section 6 draws main conclusions and addresses future works.

2. Related work

2.1. Social network analysis for churn detection

There is a large amount of literature highlighting the value of social network analysis to improve churn customer detection. Table 1 presents an overview of recent journal papers on this topic. Most of these papers (22 out of 24) focus on the churn behavior of telecom customers by leveraging CDRs to connect customers in a social network graph. As a result, this literature overview table shows a strong connection with the work of Óskarsdóttir et al. (2017) who investigated the value of network analysis from a telco perspective.

While these studies put forward the way they extract useful information from this network, all of them acknowledge the value of the CDRs-based network for the telephone company which organically collects it. The two remaining papers of this literature overview concern churn in the bank industry. The work of Benoit and Van den Poel (2012) studied sibling relationships between bank customers up to two hops. They found a significant improvement of churn detection while including network metrics into the model. The study of Van den Bulte et al. (2018) highlighted the life time value of a referral based on the observation of her/his referrer in a referral program of a bank. They found that referrals are less likely to churn as long as their referrer remains in the bank.

Together these studies suggest that customer churn is not an isolated outcome but it is permeable to peer effects. However, stated as such, this generalization might suffer from a lack of support given that most studies focus on a network specifically linked to the daily activity of the company i.e. CDRs of a telephone company. This last remark is particularly of importance for three reasons.

First, the source of the peer effect might in this context be due to a link between the network and the activity of the company. Theoretically, two main explanations are presented in the literature regarding the source of peer effect: homophily (McPherson et al., 2001), i.e. people belonging to the same community may share some similarities such as the same attitude with respect to a specific offer, and social influence (Okazaki, 2008), i.e. people might also directly influence each other through their interactions. For the specific case of the communication network, the source of peer effect is subject to discussion in the literature and it is probably a mixture of both explanations (Aral et al., 2009; Ma et al., 2014; Shalizi and Thomas, 2011). Besides this, these peer effects could also be explained by the fact that joint consumption of the service is ended when the peer leaves the telephone company (Van den Bulte et al., 2018). This is particularly salient if intra-provider communications get a preferential rate. This explanation is consistent with Benedek et al. (2014) which observed that customers closely embedded in the communication network are less likely to churn. This effect is called social enrichment by Schmitt et al. (2011). Consequently, conclusions from previous studies based on communication networks in the telecom industry may not apply to alternative networks as being connected in these networks do not imply joint consumption of the service.

Second, even if the communication network is relevant for telephone companies, alternative networks could deliver incremental value on top of this network as well. This is particularly the case if these networks portray a different source of peer influence. In other words, an incremental value would support the idea that alternative networks might represent different sources of peer influence. As we discuss hereafter, the proposed alternative networks have different origins, they might consequently vehicle different sources of peer influence.

Third, the emergence of over-the-top services such as Whats'App, Messenger and WeChat (Farooq and Raju, 2019; Stork et al., 2017; Sujata et al., 2015) might lead to the obsolescence of classical communication technologies operated by telephone companies, alternative networks might consequently also substitute the potential decline of a communication network in the future.

Authors and year	Application domain	Network	Edges	Weight	Number of features	Method	Main insight
Ahmad, A.K., Jafar, A., Aljoumaa, K. (2019)	Telco	communication	directed	weighted	+2000	Decision Tree, Random Forest, Gradient Boosted Machine Tree "GBM" and Extreme Gradient Boosting	The way to use big data plateform to do the analysis. Best results obtained with the inclusion of social network metrics and XGBoost
Ascarza, E., Ebbes, P., Netzer, O., & Danielson, M. (2017)	Telco	communication	undirected	unweighted	17	random assignment between focus and control group, t-tests and treatments effects trough diff in diff models	spill overs effects of maketing initiatives on contacts of targeted customer
Backiel, A., Baesens, B., & Claeskens, G. (2014)	Telco	communication	undirected	number of seconds of call	117	network featurization, classifiers (LR, Cox)	SNA valuable for prepaid telco customers
Baras, D., Ronen, A., & Yom-Tov, E. (2014)	Telco	communication	outgoing	number of calls	4	diffusion algorithm of Dasgupta, classifiers (DT, LR)	Usefulness of diffusion processes to make predictions
Benedek, G., Lublóy, Á., & Vastag, G. (2014)	Telco	communication	undirected	weighted/ unweighted	7	snowball sampling and simulation	Influence of ego embedness within the network
Benoit, D. F., & Van den Poel, D. (2012)	Bank	Kinship	undirected	unweighted	31	RF	SNA valuable to detect bank churners
Dierkes, T., Bichler, M., & Krishnan, R. (2011)	Telco	communication	undirected	number and duration of calls	78	Logistic regression ,Markov logic with individual and network variables	Usefulness of Markov logic networks to make predictions but featurizatior on first hop give better results
Ferreira, P., Telang, R., Godinho De Matos, M. (2019)	Telco	communication	undirected	unweighted	10	ordered probit regression	Impact of the churn of peers on the behaviour of the customer in term of activity, choice of a tariff plan and churn likelihood
Godinho de Matos, M., Ferreira, P., Belo, R. (2018)	Telco	communication	undirected	unweighted	19	randomized experiment, linear probability model	On top of targeting likely churners, churn is additionally reduced by targeting also their contacts
Haenlein, M. (2013)	Telco	communication	directed	duration of calls	14	stratified Cox Proportional Hazards model	Social effects limited to outgoing communications and in time
Keramati, A., Jafari-Marandi, R., Aliannejadi, M., Ahmadian, I., Mozaffari, M., & Abbasi, U. (2014)	Telco	communication	undirected	unweighted	11	DecisionTree, Artificial Neural Networks, K-Nearest Neighbors, and Support Vector Machine, mixture of these classifiers	Methodology combining these 4 classifiers deliver superior results
Kim, K., Jun, C. H., & Lee, J. (2014)	Telco	communication	undirected	call duration	37	SPA diffusion process, Louvain method for community detection, classifiers (logistic regression, MLP)	Combining diffusion process and community detection deliver superior results
Mitrovi, S., Baesens, B. , Lemahieu, W. , De Weerdt, J. (2018)	Telco	communication	directed	weighted	83	multiple-criteria feature selection (AUC and process time) based on pareto optimum, classifiers (LR,RF)	Best models obtained with local features (for operational efficiency) augmented direct network features (for predictive performance)
Mitrovi, S., Baesens, B., Lemahieu, W., De Weerdt, J. (2019)	Telco	communication	undirected	unweighted	27	tcc2vec, classifier (LR)	Tcc2vec approach better leverages RFM information by enriching the call graph and using representation learning
Mitrovi, S., De Weerdt, J. (2019)	Telco	communication	directed	unweighted	9	meta-path, classifier (LR)	Meta-path better leverages augmented call graph with RFM variables by using a representation learning adjusted for heterogeneous network
Nitzan, I., & Libai, B. (2011)	Telco	communication	undirected	call duration and homophily	14	proportional hazard models	Social effect of churn, moderated by tenure, connection degree and homophily
Óskarsdóttir, M., Bravo, C., Verbeke, W., Sarraute, C., Baesens, B., & Vanthienen, J. (2017)	Telco	communication	directed/ undirected	weighted/un weighted	14	relational learners (relational classifiers and collective inference methods) and non-relational learners (LR, ANN and RF)	best results from benchmarking obtained with: non relational learner, inclusion of network variables, no collective inference , binary weight anc undirected network
Óskarsdóttir, M., Van Calster T., Baesens B., Lemahieu, W., Vanthienen J. (2018)	Telco	communication	directed/ undirected	weighted	8	classifiers (similarity forest, nearest neighbour, LR and RF)	Dynamic nature of communication network taken into account with Similarity Forest
Phadke, C., Uzunalioglu, H., Mendiratta, V. B., Kushnir, D., & Doran, D. (2013)	Telco	communication	undirected	number of calls, duration and neighborhoo d overlap	10	propagation method leading to a supplementary feature, stochastic gradient boosting	SNA useful for churn prediction, proposition of a propagation method
Postigo-Boix, M., Melús-Moreno, J. L. (2018)	Telco	communication	undirected	weighted	7	ABM modelling	Usefulness of using ABM modelling to capture peer influence. This model includes customer profile and activity to connect peers.
Van den Bulte, C., Bayer, E., Skiera, B., & Schmitt, P. (2018)	Bank	referral	directed	unweighted	10	panel model with person and dyad random effect, Cox proportional hazard model	Referred people less likely to churn as long as their referrer do not churn
Verbeke, W., Martens, D., & Baesens, B. (2014)	Telco	communication	undirected	duration of calls	58	relational learners (relational classifiers and collective inference methods) and non-relational learners (LR, ADT, Bag, RF, BN)	Effect of out of first hop, proposition of combining relational and non relational learner in one final model
Zhang, X., Zhu, J., Xu, S., & Wan, Y. (2012)	Telco	communication	undirected	weighted/un weighted	39	propagation with classifiers (DT, LR and NN)	SNA useful for churn prediction, proposition of a propagation method
Zhu, T., Wang, B., Wu, B., & Zhu, C. (2011)	Telco	communication	directed/ undirected	weighted/un weighted	8	K-Means	Clustering approach on individual network feature leads to a categorisation of customer roles in the network
This study	Telco	Communication/ referral/spatial	undirected	weighted	9	LR, NN, RF	Communication network does not need to be complemented by other networks for telco if advanced techniques are used, but might be substituted by an internal referral networks for other companies

Table 1: Literature overview on the use of SNA for customer churn prediction

For these reasons, this study investigates the value of alternative networks that might be developed inside a telephone company. In particular, 2 alternative networks are analyzed, namely spatial and referral networks. Academic research has already shown that these networks are valuable for other marketing settings such as customer acquisition. Their relevance for this study is discussed hereafter.

2.2. Spatial networks

One of the oldest alternative networks investigated are spatial networks that are developed based on customers'addresses. In this case, customers are considered to be connected when they are living close to each other. The reason of connecting spatial neighbours is that they might exhibit socio-demographical similarities such as the same level of income or the same age. They might also be exposed to the same local events. The value of this network has already been proven in product acquisition (Baecke and Van den Poel, 2012, 2013; Bell and Song, 2007; Bilgicer et al., 2015; De Matos et al., 2014b; Iyengar et al., 2011; Manchanda et al., 2008; Yang and Allenby, 2003). In a study conducted by Manchanda et al. (2008), spatial proximity was used as a proxy for social contagion across physicians. This spatial proximity was found to be a significant predictor for the adoption of a new drug. To the best of our knowledge, the value of spatial similarity in a churn context received few attention in the literature. Haenlein (2013) tested the interaction effect of sharing the same zip code and the directionality of the relationship, but they found no significant results. By contrast, Moeyersoms and Martens (2015) also studied the inclusion of zip codes in churn detection models and found positive results. Compared to these two last studies, the present study investigates the value of spatial network at a more granular level i.e. GPS coordinates instead of zip codes.

2.3. Referral networks

While address information is one of the most frequently passively collected customer characteristics (Verhoef et al., 2003), organisations can also collect information regarding the influence of referrals related to its activity. This influence has been studied for a long time ago (Whyte Jr, 1954). For instance, Villanueva et al. (2008) estimated the influence of words of mouth (WOM) on firm value to be almost twice as much as the value of the new acquired customer by a marketing campaign. In term of method, the opportunity to derive more insights by studying this interpersonal influence based on network data instead of a classic survey has been highlighted by Reingen and Kernan (1986).

In particular, the fact of using referral network data enables to better estimate the value of peer influence. This value has been typically studied as a one-time event at the moment of the referral (Aral and Walker, 2014; Beaman and Magruder, 2012; Godes and Mayzlin, 2009; Hinz et al., 2011; Wirtz and Chew, 2002). However, this peer influence may also lead to additional value on the long run (Castilla, 2005; Fernandez et al., 2000; Garnefeld et al., 2013; Labianca et al., 2013; Schmitt et al., 2011). For instance, to determine the effect of a referral network on employee performance, Castilla (2005) compared the relative evolution of new workers in a call center depending on whether they were referred by an employee or not. He found a positive impact of referral on long term productivity and loyalty of these referred new workers. Referral networks have also been studied in a customer life time value perspective. In a study conducted by Garnefeld et al. (2013), it was shown that the churn rate of existing customers participating to a referral program drop by 63,16% compared to look alike customers in the next 12 months.

Hence, as supported by Van den Bulte et al. (2018) in our literature review, another alternative network which might add value to customer churn detection models is a referral network that connects referrers and referrals. This network is more costly than the spatial network but has the advantage that, by construction, social influence has already been observed during the acquisition phase of the product/service.

The current study compares the value of a frequently used communication network for customer churn detection with two alternative networks, namely a spatial and referral network which might represent other sources of peer influence. First, the value of each network is assessed separately. Next, it is analyzed on top of the communication network. By this, our study contributes to tactic and strategic decisions of telephone companies. First, on the strategic level, we investigate if alternative networks might substitute communication networks on the long run if the current expansion of over-the-top services might lead to the obsolescence of the latter network. Second, on the tactic level, we study the added value of two supplementary networks. To the best of our knowledge, the present study is one of the first studies that compares the value of multiple networks simultaneously (Ansari et al., 2011; Johnson et al., 2012).

3. Data

The data for this study was collected in 2013 from a European telephone company that targets a prepaid customer segment. It covers a period of 9 months from begin November 2012 to end July 2013. The database includes customer data, airtime transactions, CDRs and referral program data. To the best of our knowledge, this database is very unique in the sense that, beyond Call Data Records, this database also contains addresses of customers and data related to a referral program. This allows us, as we shall discuss later, to derive not only a communication network (as other related studies mentioned in section 2.1.) but also a spatial and a referral network. From a legal point of view, this data collection and treatment complies with GDPR regulation (The European Parliament, 2016) as scientific research is considered as a compatible goal to the initial goal of the treatment of the data (article 9.2.j). In this line, data collected was also anonymised as much as possible.

The raw customer data includes 256 581 customers. Data filtering was performed to ensure to get retail customers with regular air time activity and communication behavior. Details of the selection process are given in table 2. The final customer selection contains 91 085 customers.

Step	Selection	Reason	Remaining observations
1	All customers	Data available	256 581
2	Country residents	Non residents might possess specific reasons to churn due to their reduced link with the country	255 763
3	Valid address	To be able to derive a spatial network	253 117
4	Linked phone number	To filter customers already left	239 885
5	At least one outgoing and one incoming call or text on period 11/2012-01/2013	To filter inactive customers or without bidirectional activity	168 511
6	Filter flop 5% (less than 7 outgoing call/text*) on period 11/2012-01/2013	To filter customers without a regular phone use	
7	Filter top 5% (more than 1 863 outgoing call&text*) on period 11/2012-01/2013	To filter customer with extreme phone use which might be the sign of professional activity highly linked to	150 074
8	At least one top-up within the 3 months preceding 2013/05/01 – Customer selection	Minimum top-up activity considered to remain customer	91 085

Table 2: Selection process (* call counts for 1 and text for 0,5)

This study makes use of a predictive model to estimate if customers are likely to churn in the next 3 months which is operationalised as not buying airtime within this time frame. More specifically, 9 months of data are split in 3 periods of equal length. The first trimester is used to draw the communication network. The second trimester is intended to observe the churn behavior of the customers who are linked with the focal customer. The last trimester constitutes the dependent period where churn behavior of the focal customers are observed. Our selection includes 1 668 churners which churned during this last period.

The model is based on 15 independent variables that are designed based on data available before the start of the third semester. These variables are classified into 4 categories: base variables, communication variables, spatial variables and referral variables as presented in Table 3. Base variables include 6 socio-demographical and transactional variables.

Sub-category	Feature
Base	Total net income divided by number of tax reports in the neighborhood
	Age of the customer
	Gender of the customer
	Time from enrolment
	Time from last transaction
	Evolution of transactions
Communication network	Number of connections (i.e. degree)
	Average churn behavior among peers weighted by communication during trimester 2
	Total communication time spent during trimester 1 with defecting peers (of trimester 2)
Spatial network	Number of neighbors who live at most at 3.28 kilometers
	Average churn behavior among peers weighted by proximity index during trimester 2
	Number of defecting peers weighted by proximity index during trimester 2
Referral network	Number of connections (i.e. degree)
	Average churn behavior among peers during trimester 2
	Number of defecting peers during trimester 2

Table 3: List of independent variables

The three other categories include variables derived from 3 underlying networks. For each of them, given the network characteristics, 3 network-based independent variables have been created. This feature creation process has been standardised as much as possible over the 3 networks so that observed effects can be fully allocated to the network and not the feature creation process. The first feature contains the number of connected customers in this network. It measures the level of social enrichment as defined by Schmitt et al. (2011) which correspond in this study to the well-being of being connected to other mobile users of this telephone company. Customers having a higher level of social enrichment are expected to be less prone to churn. The two other features measure the average exposition and total exposition to churners. Customers exposed to churners are more likely to churn as well. As mentioned in the related work section, this similar outcome might stem from homophily (McPherson et al., 2001) or social influence (Okazaki, 2008). These two features are built as follows:

$$AverageChurnBehavior(i)_j = \frac{\sum_{p=1}^{d} W_p * Churn_p}{\sum_{p=1}^{d} W_p} \tag{1}$$

$$NumberOfDefectingPeers(i)_j = \sum_{p=1}^{d} W_p * Churn_p \tag{2}$$

where i represents the ego customer, j the network (i.e. communication, spatial or referral), p a direct connection (peer) of i, d the degree, W_p the weight, $Churn_p$ a dummy variable (1=churner, 0= not churner). In particular, the weight expresses the level of proximity between connected customers in the network. It is based on the exchanged communication time for the communication network, on the proximity index between residences for the spatial network and on a fix value of 1 for the referral network as this network denotes the presence or absence of a referral link without specific information on the strength of this link. The network specific definition of each feature is given in table 3.

The reminder of this section discusses in more details the specificities of each network construction in turn. With respect to the communication network, as discussed in the related work section, previous literature has already observed similar behaviors between people who communicate regularly by call and/or text messages (Óskarsdóttir et al., 2017). In this study, people are linked in the communication network if (1) they have communicated at least 5 times and (2) if they have both initiated at least one communication during the first month of the first trimester meaning that the communication is regular and bidirectional. This network definition is similar to Ma et al. (2014). The overall communication strength between 2 customers is estimated on the whole first trimester. It is expressed as the total number of seconds of communication where, following Kusuma et al. (2013), a text message accounts for 30 seconds of communication.

Concerning the spatial network, the underlying idea is that customers who are living close to each other have a higher chance to influence each other and/or to exhibit the same behavior. This network is created based on the addresses of the customers. More specifically, these addresses were converted into GPS coordinates using Pygeocoder, a Python package designed for Google Geocoding (Python Community, 2018). Next, Euclidean distances between the customers'coordinates were computed. In order to select only close neighbours into the ego network of each customer and to make it more computationally feasible, only customers living maximum 3.28 kilometres away from the ego network are considered. This distance equals the average distance of first percentile neighbours of each customer. Further, this distance metric was converted into a proximity index using the following formula proposed by Yang and Allenby (2003):

$$pi_{i,j} = \frac{1}{e^{d_{i,j}}} \tag{3}$$

where $pi_{i,j}$ is the proximity index between addresses of customer i and customer j, and $d_{i,j}$ is the corresponding euclidean distance expressed in kilometers between these addresses.

Finally, this proximity index is used to weight the average churn percentage of neighbors and the number of defecting peers during trimester 2. By this, the behavior of a direct neighbors receive a high weight, while a neighbor located far away only slightly impact the independent variable. In this process, preliminary tests showed that the choice of the cut-off point (i.e. 3.28 kilometers) does not significantly affect the final results as the proximity index exhibits a power law shape. For example, while a direct neighbor gets an index of 1, a neighbor living 3 kilometers away gets an index of 0.0498.

For the referral network, some influence might be expected between referral and referrer when one of them churns as they already discussed the product or service when the referral joined the telephone company. Referral marketing has been intensively used by the European telephone company of this study to grow its customer base: 58.19% of the customers were referred by existing customers who receive a reward for each successful referral. On top of this operational marketing strategy, investing in this program is also valuable from a data point of view since it enables to develop an alternative network for network analysis. For this network, no specific restriction was set. This network has also no weight since this is typically not measured in a referral program.

4. Methodology

The goal of this study is to investigate if alternative networks i.e. spatial and referral networks might complement or substitute a communication network for churn detection in a telephone company. In order to address these research questions from a practical perspective, the level of analytical maturity of the telephone company that would be required to leverage these networks is taken into account. In particular, Lismont et al. (2017) defines a correspondence between this level of maturity and the complexity level of models accessible. Their study makes the distinction between 4 levels of maturity. While the first level called "no analytics" includes companies which do not use analytics, the most differentiating factor between the 3 others in term of models, is the choice of simple and understandable models for levels 2 and 3 (called respectively "analytics bootstrappers" and "sustainable analytics adopters") versus complex models for companies with the highest level of analytical maturity (called "disruptive analytics innovators"). This is why, in this study, we compare the results of understandable models with the results of complex models. For the specific choice of models within these two categories, Lessmann et al. (2015) defines logistic regression (LR) as the industry standard for understandable models and points to neural network (NN) as being the best performing individual classifier and to random forest (RF) as the best performing ensemble classifier. Consequently, in this study, we compare the results of LR, NN and RF models. Note that this choice matches well with research practices as these three classifiers are also the most frequently occurring modeling techniques used in previous literature as illustrated in the methods column of table 1. The rest of this section further discusses these models and the choice of evaluation metrics.

Logistic regression is defined as follows:

$$P(\mathrm{Y}_i|X_i) = \frac{1}{1 + e^{-(\beta_0 + \beta_1^t * X_i)}} \tag{4}$$

where $P(\mathrm{Y}_i|X_i)$ is the estimated probability for customer i to churn within the 3 next months, X_i is a vector containing specific values of selected variables, β_0 contains the intercept and β_1 is a vector of coefficients of the same degree as X_i. As discussed above, this technique is widely used in industry, especially for companies with a moderate to medium level of analytical maturity. This choice is merely linked to the understandable nature of this technique (Lismont et al., 2017). For the same reason, this technique is also often investigated in churn prediction litterature such as Backiel et al. (2014) which studies the lifetime prediction of telco customers based on a communication network.

The neural network chosen in this study is a feed forward neural network which contains one input layer, one hidden layer and one output layer. Every neuron of one layer is connected to all neurons of the next layer which is particularly appropriate to identify potential interaction effects. To compute the input of a node of a given layer to the next layer, a logistic link function is used. The input layer contains as many neurons as independent variables introduced in the model, while the output layer contains one neuron representing the outcome i.e. churn behavior. The number of nodes in the hidden layer is fine tuned for each model using 30% of the data. A grid search is applied that ranges between the number of nodes in the input layer and the number of nodes in the output layer. Compared to LR, this technique is more complex. Consequently, as discussed above, this technique is only accessible to organisations with a high level of analytical maturity Lismont et al. (2017). It is also the best performing individual classifier Lessmann et al. (2015). In churn prediction literature, this classifier is included, among others, in Zhang et al. (2012) to study the influence of peers in the decision to churn through a propagation process.

Random forest is an ensemble method which aggregates the individual predictions of different decision trees. This technique makes use of two techniques to get different decision trees from the same data: randomisation of observations and randomisation of independent variables. The randomisation of observation is obtained by bootstrap i.e. random selection of observations with replacement. This bootstrap includes as many observations as the original dataset. On the average, every bootstrap contains 2/3 of the original data. The randomisation of independent variables is performed at each node split. Instead of testing all variables, some variables are selected at random as candidates for split. This number of variables along with the number of trees were fine tuned using the same scheme as Lessmann et al. (2015):

- number of trees: 100, 250, 500, 750, 1000
- number of candidate variables: $\sqrt{m * (0.1, 0.25, 0.5, 1, 2, 4)}$

where m is the number of variables. Like neural network, this technique is only used by organisations with an high level of analytical maturity (Lismont et al., 2017). This technique is also the most performing ensemble method (Lessmann et al., 2015). With respect to churn prediction literature, this method is frequently used. See table 1. For instance, Mitrović and De Weerdt (2019) investigates the operational efficiency of features using random forest.

In term of modeling design, in a first step, meta parameters of NN and

RF are fine tuned using 30% of the data with an embedded cross-validation scheme. Next, based on the best meta-parameters identified the first step, the models are developed and evaluated using the remaining 70% of the data. Also this is based on an embedded cross-validation scheme. For this last process, LR also includes a feature selection scheme. More concretely, each training set resulting from the cross-validation is again split into two sets to perform a backward feature selection based on the Area Under the Roc Curve (AUC) metric that is presented below. The first one (2/3 of training set) is used to go through the backward feature selection scheme. The second set (1/3 of training set) is used to select the best visited model.

In addition, for the selected models we compare the expected maximum profit generated by a campaign (Verbraken et al., 2012). Compared to the top lift approach, this metric does not predefine a threshold for the selection of the campaign. More concretely, this approach identifies the maximum profit achievable by varying the threshold level for selection. The selected threshold is consequently the one leading to the maximum profit. Furthermore, this metric includes a stochastic approach to assess the impact of the offer on the likely churners. Other parameters are deterministic. Most of them are input parameters of the method with default values based on past studies and insights from telephone companies (Verbraken et al., 2012). It is the case for the customer life time value of remaining customers (€200), the direct cost of contacting a customer (€1) and the cost of the incentive to remain (€10). One parameter which has received less attention in the litterature is the churn rate of the telephone company. In particular, Verbraken et al. (2012) already illustrates that the profits heavily depends on the churn rate in the organization. In that study, for datasets where the churn rate was less than 2%, the profit was zero or close to zero (i.e. up to 0.051), while for datasets where the churn rate was 14.14% the profit was 5.734 . This is because the proportion of true positive remains proportionally low in the selection compared to false negative even for small fractions of the population base targeted. Hence a lot of false negatives benefit from the incentive (leading to cost) as well. Therefore the profit realized on true positive who accept the offer do not compensate for the cost. Note that given the strong loyalty of the customers of the telephone company in this study, the churn rate is extremely low i.e. 1.83%. Hence, in order to increase comparability, we have simulated churn rates of 5%, 10% and 15% by performing random under sampling. What is more, we have also simulated variations of other parameters of the metric to assess the robustness of our results: a higher customer life time value for remaining likely churners (€300 instead of €200), no cost of direct contact as text messages might be free for the telephone company (compared to €1) and a reduced incentive cost (€5 instead of €10).

To further interpret the importance of each feature within a specific model, the variable importance is computed based on the study of Breiman (2001). After fitting a model, the performance is now assessed on a modified hold-on set. This modified hold-on set contains the same data as the original data except for one feature where values are permuted. Breiman (2001) did it on the specific case of the random forest where a different permutation scheme is made on the out-of-bag set for each tree to get robust results. This method can be generalized to any classifier by shuffling values of the same hold-on set multiple times. The average performance obtained compared to the initial performance gives the mean decrease in performance.

To conclude this section, the methodological hypothesis of this research are reflected. The main perspective here considered is predictive modeling i.e. the prediction of a future outcome. This perspective is notably different from a causality perspective which aims at testing the causal influence of specific factors identified from theory. To get however a better view on the individual influence, the variable importance of features is investigated. With respect to theoretical sources of influence of factors investigated, the related works subsection 2.1. discusses the sources of peer influence for the study of social networks, but, in our predictive modeling perspective, we do not try to disentangle them. Overall, research in a predictive modeling perspective does not provide a sound causal perspective i.e. why it is important but it provides insights on the managerial value of these factors.

5. Results

In order to assess the similarity and dissimilarity of churn pattern across the 3 networks, this section reports in a first subsection a bi-variate analysis between networks characteristics and churn. Results on the value of network-based features in predictive models are then reported in the two next subsections according to the two research questions: (1) might a spatial and/or a referral network substitute the frequently used communication network and (2) would a spatial and/or a referral network complement this communication network for telephone companies ?

5.1. Descriptive analysis

Figure 1 provides, for every network, the proportion of customers and the churn percentage by the number of peers, i.e. connected customers from the same telephone company. For the communication network, there is a clear decrease of the churn rate when looking at a growing number of peers. This result is consistent with Benedek et al. (2014) which highlighted that social enrichment i.e. embeddedness within the communication network reduces churn. Initially, the same pattern can be recognized based in a referral network. The more a focal customer is connected to referred customers the lower the churn probability of that focal customer. However, at a certain threshold, the churn rate increases again. This might be triggered due to the financial incentive linked to the referral program. For the spatial network, there is no clear relationship between the number of customers living in the neighborhood and the churn rate.

Figure 2 has a similar layout as figure 1, showing for every network, the proportion of customers and churn percentage, but now by the number of defecting peers. In accordance with previous literature, this churn percentage grows along with the number of defecting peers whatever the network. For example, someone having no defecting peer in the referral network has a churn rate of 1.61%, while someone connected to 2 or more defecting peers in this network reaches a churn rate of 5.77%. What is remarkable about the data in these figures is the similarity of churn pattern across the communication network and the referral network. This indicates that a well-built alternative referral network has the potential to reflect well a communication network based on CDR data. Further, we can also observe for both networks that someone not connected is more likely to churn than someone connected with one churned customer. This illustrates again the importance of social enrichment as discussed for Figure 1.

Overall, based on these 2 figures, the referral network is a promising network to investigate. On the one hand, this network shares similarities of pattern with the communication network in term of influence of the number of defecting peers on the churn likelihood suggesting that it also inherits some social enrichment properties. This supports the hypothesis that a referral network might substitute a communication network. On the other hand, the fact that the referral network, unlike the communication network, exhibits a rising churn rate starting from a degree of 4 which tends to show that this network might also reflect opportunistic behaviors of the customer that might also drive churn later on in case of an offer from a telephone company competitor. This phenomenon of opportunistic behavior in referral programs has already been discussed in the literature (Jin and Huang, 2014; Tuk et al., 2009; Verlegh et al., 2013; Wirtz et al., 2013). Hence, the referral network might complement a communication network. With respect to the spatial network, this descriptive analysis tends to show that the spatial proximity barely influences the churn behavior.

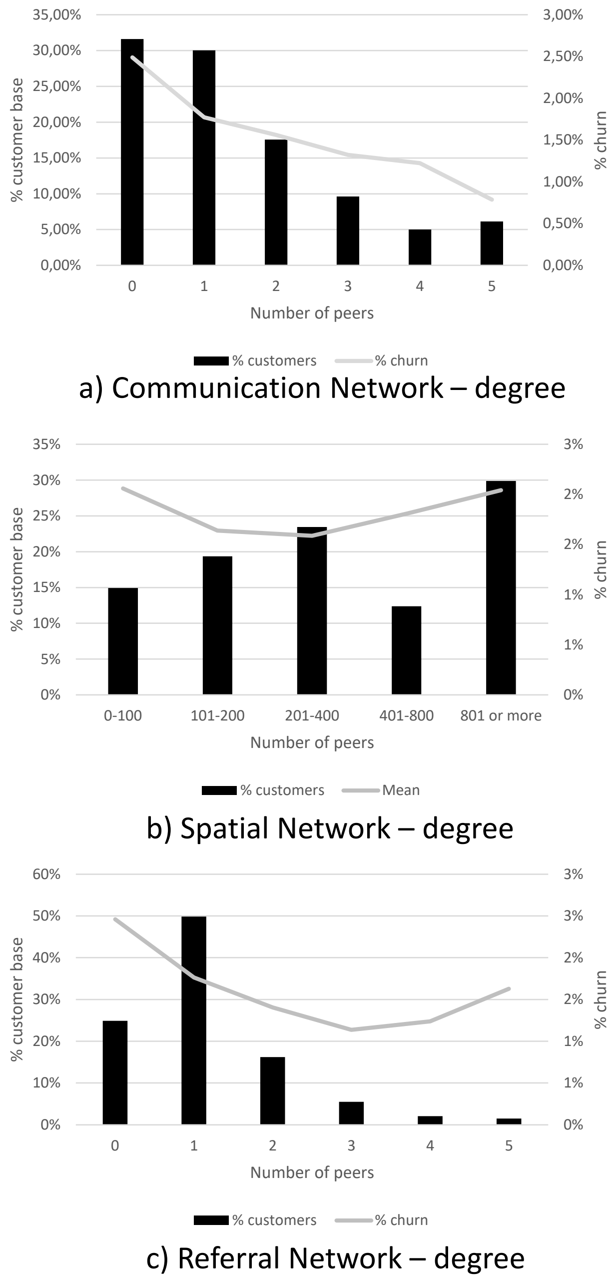
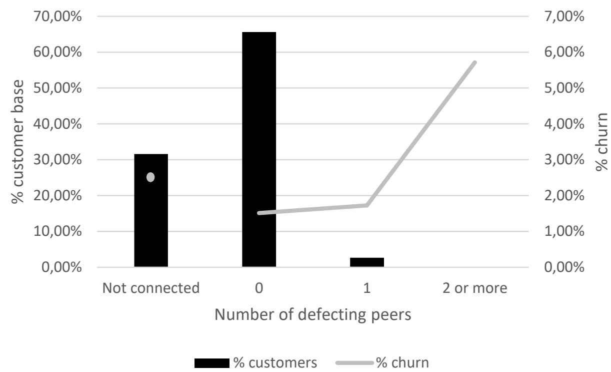
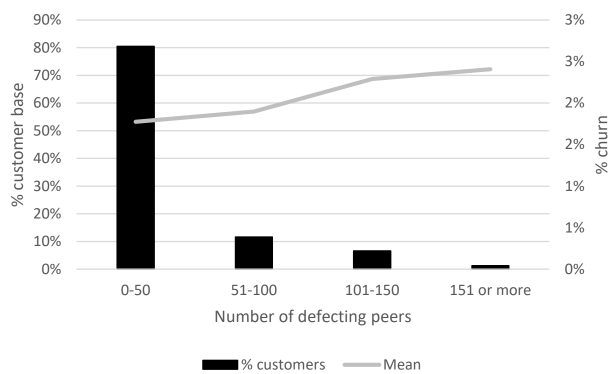


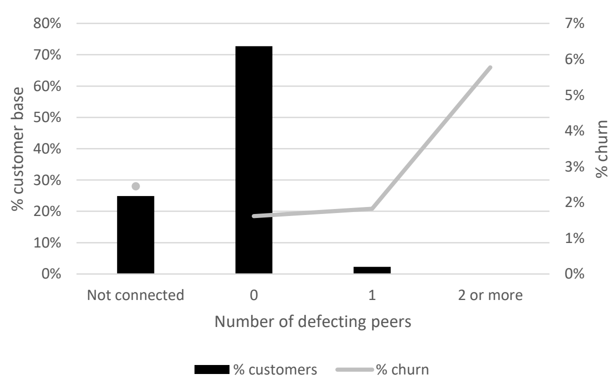
Figure 1: % customers and % churn by number of peers for communication network, referral network and spatial network



a) Communication Network – defecting peers



b) Spatial Network – defecting peers



c) Referral Network – defecting peers

Figure 2: % customers and % churn by number of defecting peers for communication network, referral network and spatial network

5.2. Substitute communication network perspective

We discuss here the perspective of a telephone company that would like to assess if 2 alternative networks i.e. a spatial or a referral network might replace a communication network in their churn prediction model if the growing use of over-the-top services would lead to the decline of traditional communications. Table 4 presents the AUC and top lift 10% results obtained with logistic regression, neural network and random forest for every intermediate model constructed starting from a base model to a full model that also includes all features of spatial and of referral network. The base+communication model is also displayed to make comparisons. Further, table 5 presents the p-values of the DeLong test for the models included in the substitution perspective in order to assess whether the observed differences between models are significant.

At the 5% significance level, spatial network does not provide a significant additional value on top the base model whatever the classifier. For the referral network tough, the underlying metrics deliver substantial incremental value to predict churn on top of the base model (from 1.32% AUC difference with NN to 1.84% with LR). Furthermore, in accordance with the bivariate analysis, when comparing the performance of the model "Base + Referral" with the performance of the model "Base + Communication", we see that the referral network is already able to capture quite a big proportion of the improvement resulting from the communication network. See figure 3. For a logistic regression model, an improvement of 1.84% is reached compared to the base model, while a model based on communication data is only able to do 1.13% better than the model based on a referral network.

Using more complex models (i.e. neural network model or random forest), this improvement is not even significant anymore using the Delong Test (see table 5). These AUC results are consistent with top lift 10% results: top lifts for the referral model are even slightly better with NN (0.17% incremental lift) and RF (0.21% incremental lift).

	Model	AUC LR	AUC NN	AUC RF	LIFT 10% LR	LIFT 10% NN	LIFT 10% RF
Complement	Full model	63.64	68.32	67.98	3.37	4.14	3.86
Complement	Base + Communication + Referral	63.64	68.32	67.98	3.37	4.14	3.86
Complement	Base + Communication + Spatial	63.26	68.32	67.48	3.33	4.14	3.65
	Base + Communication	63.26	68.32	67.48	3.33	4.14	3.65
Substitute	Base + Referral + Spatial	62.13	67.68	67.98	3.31	4.31	3.86
Substitute	Base + Referral	62.13	67.68	67.98	3.31	4.31	3.86
Substitute	Base + Spatial	60.29	66.36	67.25	2.81	3.73	3.70
Substitute	Base	60.29	66.36	66.35	2.81	3.73	3.40

Table 4: AUC and top lift 10% by model, classifier and perspective (expressed in %)

Classifier	Models compared	Base + CN	Base + SN + RN	Base + RN	Base + SN	Base
LR	Base + CN	1	0.0204	0.0204	<0.0001	<0.0001
	Base + SN + RN	0.0204	1	1	<0.0001	<0.0001
	Base + RN	0.0204	1	1	<0.0001	<0.0001
	Base + SN	<0.0001	<0.0001	<0.0001	1	1
	Base	<0.0001	<0.0001	<0.0001	1	1
NN	Base + CN	1	0.2430	0.2430	0.0008	0.0008
	Base + SN + RN	0.2430	1	1	0.0153	0.0153
	Base + RN	0.2430	1	1	0.0153	0.0153
	Base + SN	0.0008	0.0153	0.0153	1	1
	Base	0.0008	0.0153	0.0153	1	1
RF	Base + CN	1	0.3944	0.3944	0.7355	0.0499
	Base + SN + RN	0.3944	1	1	0.2386	0.0005
	Base + RN	0.3944	1	1	0.2386	0.0005
	Base + SN	0.7355	0.2386	0.2386	1	0.0894
	Base	0.0499	0.0005	0.0005	0.0894	1

Table 5: Substitution perspective - p-values of DeLong Test (CN= communication network, SN= spatial network and RN= referral network)

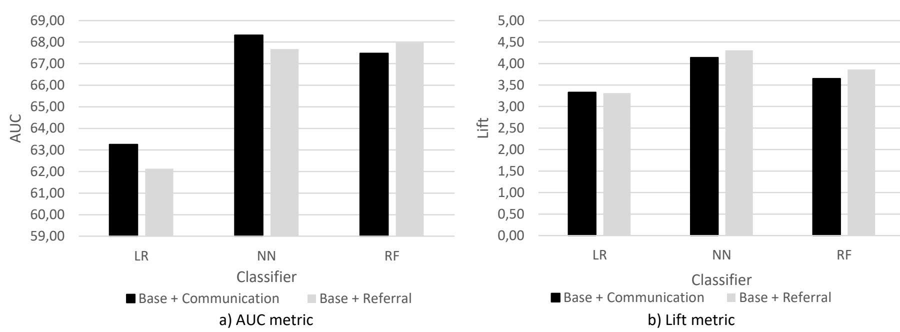


Figure 3: Substitute perspective main result: comparison of model "Base + communication" with model "Base + Referral by classifier, AUC and Lift"

The comparison of the performance of the referral network and communication network is further investigated in term of monetary performance using the Expected Maximum Profit Criterion (EMPC)(Verbraken et al., 2012) for different scenarios of churn rate, cost and profit. See figure 5. Overall, results of these simulations show that the monetary value of both models are very close.

To further investigate which metrics of the referral network are the most important, the variable importance of the best performing classifier i.e. random forest is displayed in figure 5. While transactional metrics contribute the most, the 3 underlying metrics of the referral network approximately add the same contribution to the model. Randomly permuting the values of one of them decreases the AUC from 1.63% to 2.01%.

5.3. Complement communication network perspective

We discuss now the perspective of a telephone company that would like to enhance the benefits of peer effects beyond the potential given by a communication network. From this perspective, closer inspection of the models using the communication network in table 4 shows that this network remains significantly the most appropriate to identify customer churn. Adding spatial network features to the model do not contribute to the model. Adding referral features is able to slightly improve the logistic regression and the random forest classifiers. See figure 6. However, these improvements are not statistically significant (see table 6). Top lift 10% deliver the same results where a slight improvement is only observed for random forest (0.21%).

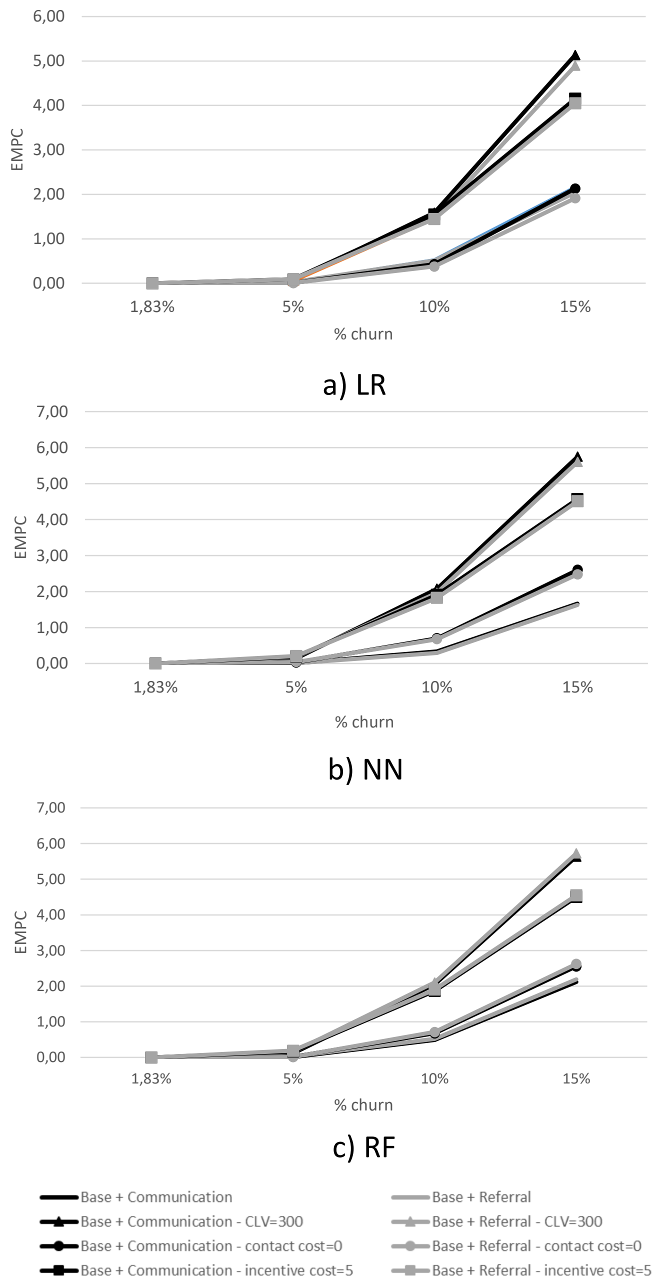


Figure 4: Comparison of Expected Maximum Profit Criterion between base+communication and base+referral models by classifier

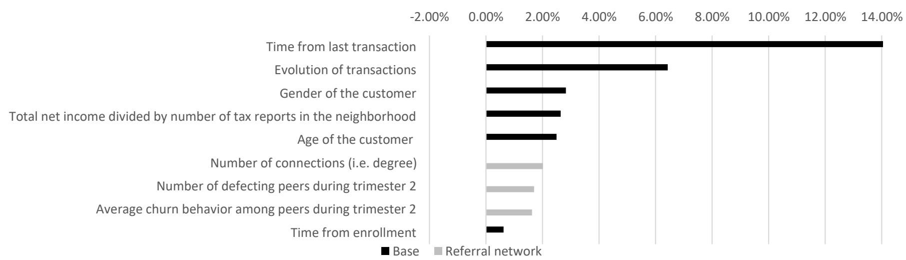


Figure 5: AUC Variable importance of metrics for base+referral model - random forest classifier

Although, the referral network individually adds significant value to a base model, it is reflecting to much of a similar network information to add value on top of the communication network. Hence, previous research on customer churn within telephone companies which has only taken into account this communication network remains relevant.

Looking further into the contribution of metrics derived from this communication network, figure 7 shows the variable importance of the base+ communication model with the best performing classifier i.e. neural network. The number of connections in this network is the most prevailing feature. The second most important feature is the total communication spent with defecting contacts, while the average churn behavior among contacts does not exhibit a contribution to the model. Note that there are 2 features with a negative variable importance. The magnitudes of these variable importances are however extremely low (up to -0.21%). This means that these features along with "Time from enrollment" (0.11%) do not have a significant impact on the AUC performance of the model.

Classifier	Models compared	Full model	Base + CN + RN	Base + CN + SN	Base + CN	Base
LR	Full model	1	1	0.0594	0.0594	<0.0001
	Base + CN + RN	1	1	0.0594	0.0594	<0.0001
	Base + CN + SN	0.0594	0.0594	1	1	<0.0001
	Base + CN	0.0594	0.0594	1	1	<0.0001
	Base	<0.0001	<0.0001	<0.0001	<0.0001	1
NN	Full model	1	1	1	1	0.0008
	Base + CN + RN	1	1	1	1	0.0008
	Base + CN + SN	1	1	1	1	0.0008
	Base + CN	1	1	1	1	0.0008
	Base	0.0008	0.0008	0.0008	0.0008	1
RF	Full model	1	1	0.3944	0.3944	0.0005
	Base + CN + RN	1	1	0.3944	0.3944	0.0005
	Base + CN + SN	0.3944	0.3944	1	1	0.0499
	Base + CN	0.3944	0.3944	1	1	0.0499
	Base	0.0005	0.0005	0.0499	0.0499	1

Table 6: Complement perspective - p-values of DeLong Test (CN= communication network, SN= spatial network and RN= referral network)

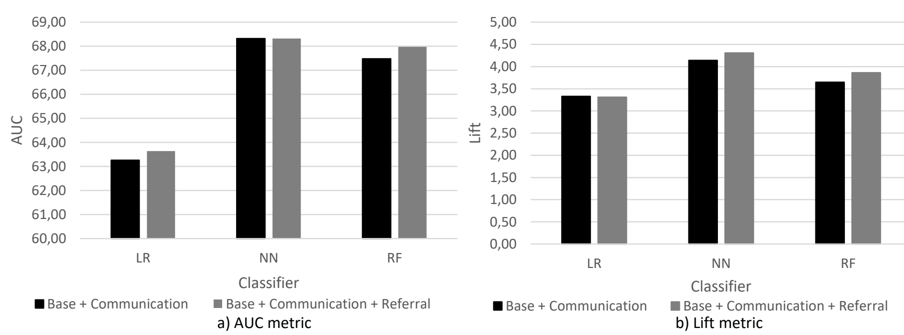


Figure 6: Complement perspective main result: comparison of model “Base + communication” with model “Base + Referral by classifier, AUC and Lift”

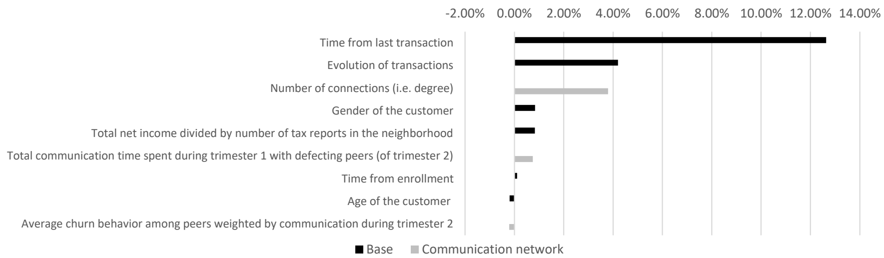


Figure 7: AUC Variable importance base+communication model - neural network classifier

6. Discussion and Conclusion

The main goal of this study is to determine the potential of alternatives to the communication network for the identification of customer churn, this last network being commonly used in the literature. Two perspectives are addressed depending on the temporal dimension considered: (1) the capacity of alternative networks to substitute a communication network on the long term if traditional communications would decline due to the competition of over-the-top services and (2) the incremental value of alternative networks for telephone companies on the short term. What is more, the required analytical maturity of the company is investigated to leverage these networks.

Suggested by literature review, the value of spatial network and referral network were tested. These networks are particularly interesting to investigate as the spatial network, unlike the referral network, is free of charge while the referral network has the advantage to draw links between customers having previously communicated about the telephone company by design. To assess these contributions, data from an European telephone company was analyzed. A communication network was extracted through their Call Data Records, a spatial network was retrieved by drawing links between customers living close to each other and a referral network was built on their referral program.

Results of this study highlight the opportunity for telephone companies to develop their long term resilience in a competitive environment where traditional communication technologies operated by telephone companies are more and more neglected in favor of mobile services operated by giant techs such as Facebook (Farooq and Raju, 2019; Stork et al., 2017; Sujata et al., 2015). Marketing managers of these companies might previously have considered to start a referral program, however this could have been to be too costly to be beneficial on itself. This study illustrates that this investment should be considered in combination with the data augmentation benefits it delivers for churn detection models. This study shows that, a referral network could be a very efficient alternative to the communication network as we found no statistical difference between the performance of both networks when an advanced classifier (i.e. neural network or a random forest) was applied. This result also encourages telephone companies to develop their analytical maturity to be able to leverage the referral network once the communication network would become too partial to reflect peer influence in the future.

At least on the short term, a communication network remain the most appropriate network to identify potential churners. Spatial and referral networks hardly add value on top of this network. Previous research on customer churn within telephone companies that has only taken into account this communication network remains consequently relevant. Only the referral network-based features can add some value when the company is at a early stage of analytical maturity and sticks to basic logistic regression analysis techniques. This study would discourage managers from a telecom organization to invest in referral programs when the main objective for this would be data augmentation.

While spatial networks prove to be valuable for customer acquisition models, this effect is not observed in a churn context whatever the perspective. This result might indicate that homophily effects (McPherson et al., 2001) are more valuable for product adoption models, while social influence (Okazaki, 2008) with direct communication adds relatively more value in churn models. This hypothesis however, assumes that spatial networks tend to capture more homophily effects because people living close to each other might share some similarities, while communication and referral networks tend to resemble social contagion as they portray interactions among individuals and active recommendations. In this research, no data was available to distinguish the source of peer effect. Further research could investigate the link between network types and the source of peer effect in order to support this hypothesis.

Another limitation linked to our data concerns the choice of a predictive modeling approach. This means that these models predict the probability of churn, while the likelihood that customers would accept the commercial offer is ignored. This is consistent with the papers listed in our literature review. A more recent approach, called uplift modeling, consists of identifying potential churners that are most likely to remain if they receive an offer to stay. This target audience is not necessarily the same as the one identified in our approach due to the fact that customers' sensitivity to commercial offers differ, especially if they have a tendency to churn. An uplift modeling approach can significantly add value (Devriendt et al., 2018), however it also requires additional data. More specifically, often through A/B testing, data were collected of a sample of customers that received a commercial offer and another sample that did not in combination with the churn behavior of both samples. Unfortunately, In this study, these data were not available. Future research might assess the usefulness of alternative networks in an uplift modeling approach.

A last limitation concerns the fact that we use the data of one telephone company. This leads to potential generalization issues: do the findings apply to other datasets ? An idealistic approach would be to use all datasets. It would nevertheless practically be impossible to access mobile data of all telephone companies as this access is subject to significant discussions with the telephone company. What is more, the time needed for operationally getting the data and hardware limitations for such amounts of data also limit the possibility to access more mobile data. However, we believe that this generalization issue is mitigated by the fact that data of mobile companies have typically the same structure. This is why most studies using mobile data focus on one set of data. Follow-up studies might include data from multiple telephone companies to strengthen the external validity of the results, especially from other continents.

To conclude, the present study highlights the interest of investigating multiple networks among the same set of individuals at the same time. Indeed, this line of research has received few attention compared to the analysis of a single network (Ansari et al., 2011; Johnson et al., 2012). Analyzing multiple networks at once allows to study research questions linked to the substitution capability or complementary of these networks. In particular, in the light of the present research, it would be interesting to further investigate for organizations without organic network if peer effects would be better leveraged through the acquisition of an external network such as a communication network or by an internal initiative such as a referral network.

References

Ahmad, A. K., A. Jafar, and K. Aljoumaa
2019. Customer churn prediction in telecom using machine learning in big data platform. *Journal of Big Data*, 6(1):28–51.

Ansari, A., O. Koenigsberg, and F. Stahl
2011. Modeling multiple relationships in social networks. *Journal of Marketing Research*, 48(4):713–728.

Aral, S., L. Muchnik, and A. Sundararajan
2009. Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks. *Proceedings of the National Academy of Sciences*, 106(51):21544–21549.

Aral, S. and D. Walker
2014. Tie strength, embeddedness, and social influence: A large-scale networked experiment. *Management Science*, 60(6):1352–1370.

Ascarza, E., P. Ebbes, O. Netzer, and M. Danielson

2017. Beyond the target customer: Social effects of customer relationship management campaigns. *Journal of Marketing Research*, 54(3):347–363.

Backiel, A., B. Baesens, and G. Claeskens

2014. Mining telecommunication networks to enhance customer lifetime predictions. In *Artificial Intelligence and Soft Computing*, Pp. 15–26. Springer.

Baecke, P. and D. Van den Poel

2012. Including spatial interdependence in customer acquisition models: a cross-category comparison. *Expert Systems with Applications*, 39(15):12105–12113.

Baecke, P. and D. Van den Poel

2013. Improving customer acquisition models by incorporating spatial autocorrelation at different levels of granularity. *Journal of Intelligent Information Systems*, 41(1):73–90.

Baras, D., A. Ronen, and E. Yom-Tov

2014. The effect of social affinity and predictive horizon on churn prediction using diffusion modeling. *Social Network Analysis and Mining*, 4(1):232–245.

Beaman, L. and J. Magruder

2012. Who gets the job referral? evidence from a social networks experiment. *American Economic Review*, 102(7):3574–93.

Bell, D. R. and S. Song

2007. Neighborhood effects and trial on the internet: Evidence from online grocery retailing. *Quantitative Marketing and Economics*, 5(4):361–400.

Benedek, G., Á. Lublóy, and G. Vastag

2014. The importance of social embeddedness: Churn models at mobile providers. *Decision Sciences*, 45(1):175–201.

Benoit, D. F. and D. Van den Poel

2012. Improving customer retention in financial services using kinship network information. *Expert Systems with Applications*, 39(13):11435–11442.

Bilgicer, T., K. Jedidi, D. R. Lehmann, and S. A. Neslin

2015. Social contagion and customer adoption of new sales channels. *Journal of Retailing*, 91(2):254–271.

Breiman, L.

2001. Random forests. *Machine learning*, 45(1):5–32.

Castilla, E. J.

2005. Social networks and employee performance in a call center. *American journal of sociology*, 110(5):1243–1283.

Church, E. M., L. S. Iyer, and X. Zhao

2015. Using interest graphs to predict rich-media diffusion in content-based online social networks. *Information Systems Management*, 32(3):210–219.

Dasgupta, K., R. Singh, B. Viswanathan, D. Chakraborty, S. Mukherjea, A. A. Nanavati, and A. Joshi
2008. Social ties and their relevance to churn in mobile telecom networks. In *Proceedings of the 11th international conference on Extending database technology: Advances in database technology*, Pp. 668–677. ACM.

De Cnudde, S., J. Moeyersoms, M. Stankova, E. Tobback, V. Javaly, and D. Martens
2019. What does your facebook profile reveal about your creditworthiness? using alternative data for microfinance. *Journal of the Operational Research Society*, 70(3):353–363.

De Matos, M. G., P. Ferreira, and D. Krackhardt
2014a. Peer influence in the diffusion of iphone 3g over a large social network. *MIS Q.*, 38(4):1103–1134.

De Matos, M. G., P. Ferreira, and D. Krackhardt
2014b. Peer influence in the diffusion of iphone 3g over a large social network. *Mis Quarterly*, 38(4):1103–1134.

Devriendt, F., D. Moldovan, and W. Verbeke
2018. A literature survey and experimental evaluation of the state-of-the-art in uplift modeling: A stepping stone toward the development of prescriptive analytics. *Big data*, 6(1):13–41.

Dierkes, T., M. Bichler, and R. Krishnan
2011. Estimating the effect of word of mouth on churn and cross-buying

in the mobile phone market with markov logic networks. *Decision Support Systems*, 51(3):361–371.

Farooq, M. and V. Raju

2019. Impact of over-the-top (ott) services on the telecom companies in the era of transformative marketing. *Global Journal of Flexible Systems Management*, 20(2):177–188.

Fernandez, R. M., E. J. Castilla, and P. Moore

2000. Social capital at work: Networks and employment at a phone center. *American journal of sociology*, 105(5):1288–1356.

Ferreira, P., R. Telang, and M. G. De Matos

2019. Effect of friends' churn on consumer behavior in mobile networks. *Journal of Management Information Systems*, 36(2):355–390.

Garnefeld, I., A. Eggert, S. V. Helm, and S. S. Tax

2013. Growing existing customers' revenue streams through customer referral programs. *Journal of Marketing*, 77(4):17–32.

Godes, D. and D. Mayzlin

2009. Firm-created word-of-mouth communication: Evidence from a field test. *Marketing science*, 28(4):721–739.

Godinho de Matos, M., F. Pedro, and B. Rodrigo

2018. Target the ego or target the group: Evidence from a randomized experiment in proactive churn management. *Marketing Science*, 37(5):793–811.

Goel, S. and D. G. Goldstein

2014. Predicting individual behavior with social networks. *Marketing Science*, 33(1):82–93.

Granovetter, M. S.

1977. The strength of weak ties. In *Social networks*, Pp. 347–367. Elsevier.

Haenlein, M.

2013. Social interactions in customer churn decisions: The impact of relationship directionality. *International Journal of Research in Marketing*, 30(3):236–248.

Hill, S., F. Provost, and C. Volinsky

2006. Network-based marketing: Identifying likely adopters via consumer networks. *Statistical Science*, 21(2):256–276.

Hinz, O., B. Skiera, C. Barrot, and J. U. Becker

2011. Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing*, 75(6):55–71.

Iyengar, R., C. Van den Bulte, and J. Choi

2011. Distinguishing between drivers of social contagion: Insights from combining social network and co-location data. *Dartmouth College, Hanover, NH, Scheduled.*

Jin, L. and Y. Huang

2014. When giving money does not work: The differential effects of monetary versus in-kind rewards in referral reward programs. *International Journal of Research in Marketing*, 31(1):107–116.

Johnson, R., B. Kovács, and A. Vicsek

2012. A comparison of email networks and off-line social networks: A study of a medium-sized bank. *Social Networks*, 34(4):462–469.

Keramati, A., R. Jafari-Marandi, M. Aliannejadi, I. Ahmadian, M. Mozaffari, and U. Abbasi

2014. Improved churn prediction in telecommunication industry using data mining techniques. *Applied Soft Computing*, 24:994–1012.

Kim, K., C.-H. Jun, and J. Lee

2014. Improved churn prediction in telecommunication industry by analyzing a large network. *Expert Systems with Applications*, 41(15):6575–6584.

Kusuma, P. D., D. Radosavljevik, F. W. Takes, and P. van der Putten

2013. Combining customer attribute and social network mining for prepaid mobile churn prediction. In *Proc. the 23rd Annual Belgian Dutch Conference on Machine Learning (BENELEARN)*, Pp. 50–58.

Labianca, G., G. Kane, Alavi, and S. Borgatti

2013. What's different about social media networks? a framework and research agenda. *MIS Quarterly*, 38(1):274–304.

Legos

2019. Whatsapp tuera-t-il le sms? https://www.legos.fr/will-whatsapp-kill-the-sms/.

Lessmann, S., B. Baesens, H. Seow, and L. Thomas

2015. Benchmarking state-of-the-art classification algorithms for credit scoring: A ten-year update. *European Journal of Operational Research*, 247(1):124–136.

Lismont, J., J. Vanthienen, B. Baesens, and W. Lemahieu
2017. Defining analytics maturity indicators: A survey approach. *International Journal of Information Management*, 37(3):114–124.

Ma, L., R. Krishnan, and A. L. Montgomery
2014. Latent homophily or social influence? an empirical analysis of purchase within a social network. *Management Science*, 61(2):454–473.

Manchanda, P., Y. Xie, and N. Youn
2008. The role of targeted communication and contagion in product adoption. *Marketing Science*, 27(6):961–976.

McPherson, M., L. Smith-Lovin, and J. M. Cook
2001. Birds of a feather: Homophily in social networks. *Annual review of sociology*, 27(1):415–444.

Mitrović, S., B. Baesens, W. Lemahieu, and J. De Weerdt
2018. On the operational efficiency of different feature types for telco churn prediction. *European Journal of Operational Research*, 267(3):1141–1155.

Mitrović, S., B. Baesens, W. Lemahieu, and J. De Weerdt
2019. tcc2vec: Rfm-informed representation learning on call graphs for churn prediction. *Information Sciences*. https://doi.org/10.1016/j.ins.2019.02.044.

Mitrović, S. and J. De Weerdt

2019. Churn modeling with probabilistic meta paths-based representation learning. *Information Processing & Management*, P. 102052.

Moeyersoms, J. and D. Martens

2015. Including high-cardinality attributes in predictive models: A case study in churn prediction in the energy sector. *Decision support systems*, 72:72–81.

Nitzan, I. and B. Libai

2011. Social effects on customer retention. *Journal of Marketing*, 75(6):24–38.

Okazaki, S.

2008. Determinant factors of mobile-based word-of-mouth campaign referral among japanese adolescents. *Psychology & Marketing*, 25(8):714–731.

Onnela, J.-P., J. Saramki, J. Hyvnen, G. Szab, D. Lazer, K. Kaski, J. Kertsz, and A.-L. Barabsi

2007. Structure and tie strengths in mobile communication networks. *Proceedings of the National Academy of Sciences*, 104(18):7332–7336.

Óskarsdóttir, M., C. Bravo, W. Verbeke, C. Sarraute, B. Baesens, and J. Vanthienen

2017. Social network analytics for churn prediction in telco: Model building, evaluation and network architecture. *Expert Systems with Applications*, 85(C):204–220.

Owczarczuk, M.

2010. Churn models for prepaid customers in the cellular telecommunication industry using large data marts. *Expert Systems with Applications*, 37(6):4710–4712.

Phadke, C., H. Uzunalioglu, V. B. Mendiratta, D. Kushnir, and D. Doran
2013. Prediction of subscriber churn using social network analysis. *Bell Labs Technical Journal*, 17(4):63–76.

Postigo-Boix, M. and J. L. Melús-Moreno
2018. A social model based on customers' profiles for analyzing the churning process in the mobile market of data plans. *Physica A: Statistical Mechanics and its Applications*, 496:571–592.

Python Community
2018. Python interface for google geocoding. https://pypi.org/project/pygeocoder/.

Reingen, P. H. and J. B. Kernan
1986. Analysis of referral networks in marketing: Methods and illustration. *Journal of marketing research*, 23(4):370–378.

Schmitt, P., B. Skiera, and C. Van den Bulte
2011. Referral programs and customer value. *Journal of Marketing*, 75(1):46–59.

Shalizi, C. R. and A. C. Thomas
2011. Homophily and contagion are generically confounded in observational social network studies. *Sociological methods & research*, 40(2):211–239.

Stork, C., S. Esselaar, and C. Chair

2017. OTT-Threat or opportunity for African Telcos? *Telecommunications Policy*, 41(7-8):600–616.

Sujata, J., S. Sohag, D. Tanu, D. Chintan, P. Shubham, and G. Sumit

2015. Impact of over the top (ott) services on telecom service providers. *Indian Journal of Science and Technology*, 8(S4):145–160.

The European Parliament, C. o. E. U.

2016. On the protection of natural persons about the processing of personal data and on the free movement of such data, and repealing directive 95/46/ec (general data protection regulation). *Official Journal of the European Union*, 679:1–88.

Tuk, M. A., P. W. Verlegh, A. Smidts, and D. H. Wigboldus

2009. Sales and sincerity: The role of relational framing in word-of-mouth marketing. *Journal of Consumer Psychology*, 19(1):38–47.

Van den Bulte, C., E. Bayer, B. Skiera, and P. Schmitt

2018. How customer referral programs turn social capital into economic capital. *Journal of Marketing Research*, 55(1):132–146.

Van Vlasselaer, V., C. Bravo, O. Caelen, T. Eliassi-Rad, L. Akoglu, M. Snoeck, and B. Baesens

2015. Apate: A novel approach for automated credit card transaction fraud detection using network-based extensions. *Decision Support Systems*, 75:38–48.

Verbeke, W., D. Martens, and B. Baesens

2014. Social network analysis for customer churn prediction. *Applied Soft Computing*, 14:431–446.

Verbraken, T., F. Goethals, W. Verbeke, and B. Baesens

2014. Predicting online channel acceptance with social network data. *Decision Support Systems*, 63:104–114.

Verbraken, T., W. Verbeke, and B. Baesens

2012. A novel profit maximizing metric for measuring classification performance of customer churn prediction models. *IEEE transactions on knowledge and data engineering*, 25(5):961–973.

Verhoef, P. C., P. N. Spring, J. C. Hoekstra, and P. S. Leeflang

2003. The commercial use of segmentation and predictive modeling techniques for database marketing in the netherlands. *Decision Support Systems*, 34(4):471–481.

Verlegh, P. W., G. Ryu, M. A. Tuk, and L. Feick

2013. Receiver responses to rewarded referrals: the motive inferences framework. *Journal of the Academy of Marketing Science*, 41(6):669–682.

Villanueva, J., S. Yoo, and D. M. Hanssens

2008. The impact of marketing-induced versus word-of-mouth customer acquisition on customer equity growth. *Journal of marketing Research*, 45(1):48–59.

Whyte Jr, W. H.

1954. The web of word of mouth. *Fortune*, 50(5):140–143.

Wirtz, J. and P. Chew

2002. The effects of incentives, deal proneness, satisfaction and tie strength on word-of-mouth behaviour. *International journal of service industry management*.

Wirtz, J., C. Orsingher, P. Chew, and S. K. Tambyah

2013. The role of metaperception on the effectiveness of referral reward programs. *Journal of Service Research*, 16(1):82–98.

Yang, S. and G. M. Allenby

2003. Modeling interdependent consumer preferences. *Journal of Marketing Research*, 40(3):282–294.

Zhang, X., J. Zhu, S. Xu, and Y. Wan

2012. Predicting customer churn through interpersonal influence. *Knowledge-Based Systems*, 28:97–104.

Zhu, T., B. Wang, B. Wu, and C. Zhu

2011. Role defining using behavior-based clustering in telecommunication network. *Expert Systems with Applications*, 38(4):3902–3908.