Master thesis Economics and Business: Data Science and Marketing Analytics

Characteristics that influence product adoption and defection decisions:

the case of telecom service diffusion within the social network

April 2020

Abstract: Thanks to technological advancement an ease of communication between individuals is constantly increasing, which has an impact on their decisions as the consumers. As a result of this technological development it is possible to investigate the motives, which guide customer's behavior. The goal of this thesis is to investigate such characteristics which influence customer decision, in the context of the adoption and defection of a mobile add-on (call-tune) service. For this purpose, information about individuals' demographic characteristics, the use of telecommunication services and, most importantly, the social tie with an already adopter or defector of this service, is used. For the purpose of statistical inference, the statistical model of logistic regression and the machine learning model of classification decision trees are applied. The results of this research have indicated that individual's decisions can be predicted based on their demographic profiles and the previous use of telecom service. Similarly, close social neighbors and people with similar social profiles can affect an individual's decision to conform with his social tie. Finally, it is demonstrated that some individuals are more influential regarding their peer's decision to adopt or defect the service.

Keywords: social influence, social network, social contagion, peer influence, social ties, service adoption, service defection, demographic characteristics, homophily.

Student name and ID: Efthymia Ioannidou, 502498ei

Supervisor: Prof. Dr. Vardit Landsman Schwartz

Second assessor: Prof. Dr. Dennis Fok

Discipline: MSc Economics and Business

Programme: Data Science and Marketing Analytics

Table of content

1. Introduction

In the modern era, social networks are constantly gaining public attention, especially in the marketing field. Social network applications are widely considered crucial for marketeers in order to understand and apply influence marketing and customer acquisition techniques. Bristor (1985) explains that members of a social network should not be examined as individuals, since they are free actors of this system, who have impact on other people's choices and vice versa. From the published scientific papers, it can be understood that the actors of a social network are talking about products and services and they exchange opinions and feedback about them. The 'word-of-mouth' behavior can be investigated as a format of spread of information and diffusion of social influence (Lin, 1971). One of many researches in this topic, made by Rogers (2003), suggests that people tend to be consulted by others with a similar background and opinions for their decision making. Thus, researchers and companies should understand that it is easier to acquire new clients by targeting their current client's social networks.

The research on social network is about combining the investigation of different scientific sectors. This exact combination of the sciences is what makes the social network at the same time interesting and complicated. Sciences like sociology, epidemiology, marketing and data science, to name just a few, come together to help marketeers and scientists in understanding how the consumers interact. In order, to comprehend behavior of the individuals in their social exchanges it is necessary to know the fundamental concepts of social sciences. Epidemiology helps to understand how information is spread within the social network. Marketing uses and applies that knowledge to explain consumer's behavior and understand the motivation behind adoption of new products or services. Finally, data science helps to process the communication and adoption data and to find relevant insights. When all these aspects are combined, then predictions on people's social interactions and adoptions of product or services can be made. Social networks can help researchers and companies in making decisions and finding critical solutions to problems related to consumer behaviors. Granovetter (1985) demonstrates that the importance of understanding the social network helps companies to forecast future consumer actions and consumers' choices.

In the modern era of massive information flow, consumers are more suspicious to advertising than before. Based on Assael (1998) "word-of-mouth" has a stronger influential power than mass advertisement. For this reason, modern marketeers should consider various alternative ways to acquire new consumers and to retain and satisfy already existing ones. Katz et al. (2017) mentioned that promotion of a product or service has to be accompanied by a social reference; otherwise the consumers will not consider buying it. Marketeers need to know whom the consumers trust and whom do they refer to. The outcome of such a social interaction and social link is the flow of information about the product and the personal experience with it. Ansari, Koenigsberg and Stahl (2011) believe that the exchange of information provides an additional value to the customer. This information exchange is affecting consumer's opinions and behaviors, not only in the stage of awareness of the product or service, but also in the following stage of evaluation of alternatives. Wilcox and Stephen (2013) mention that the social network has a strong impact on people's feelings, behavior and decisions they make. Thus, the actors of the social network do not only transfer information about products, but also influence the choices of the members of their social network. It is intriguing to investigate who are those critical members of the social network and how close to the other members are they in the network.

As it is mentioned previously, the members of an individual's social network play a very important role to the adoption decisions of the consumer, but it is crucial to understand that not all acquaintances are equally important and influential. As Iyengar, Van de Bulte and Valente (2011) explain, the fundamental idea of a social influence is that neither all people influence equally, nor all kinds of social ties have the same impact on people's behavior. Ma, Krishman and Montgomery (2015) suggest that the closer two consumers are, the more similar is their behavior. They continue by explaining that people who are better connected have the strongest information exchange and hence influence. As people do not communicate with the same intensity, they are not influenced by all the acquaintances equally. It is important to explore if consumers are following the choices and adoptions of their friends with similar characteristics or they are adopting the choices of their friends with different attributes. For acquiring new clients, it is important for marketeers to understand which of these powers of influence is the dominant one. Previous researches explained the phenomena related to that behavior with the terms of weak and strong ties, homophily and heterophily (Granovetter, 1973; De Bruyn and Lilien, 2008).

Marketeers make efforts to use social influences as a marketing tool, such as “word-of-mouth” and social network marketing or “peer to peer” campaigns. Social network as a marketing technique is an intriguing trend, but it needs careful investigation, because of both influence and resistance to influence can be triggered (Mourali and Yang, 2013). There is a reason to investigate internal communication techniques, because individuals are taking their decisions upon the adopting or defecting a service based on “word-of-mouth”, social influence and mimicry (Murray 1991; Wangenheim and Bayon 2004). Therefore, targeting in the new era should be raised to a new level, not only regarding segmentation, but also regarding the network level (Iacobucci, 2009). A more efficient approach for marketeers is primarily to focus at understanding the structure and the way in which social networks work and then at targeting it. A company, which utilizes the social network as a marketing technique will enhance its likelihood for successful costumer acquisition, since targeting the costumer’s friend has several advantages. Friends tend to make similar choices, therefore it is more likely that a friend likes the product, for such reasons as similar preferences, lifestyles, reference points and social influences. Moreover, a friend is already aware of the product or service, and this fact may decrease company’s expenses for product awareness. Other expenses that can be lowered comprise the customer acquisition and retention costs. These costs usually are high and that is why the social network marketing can be an efficient way to solve this problem. The most important is to attract the correct customers with the highest future earnings, who can be identified by social network analyses. Although, investigation of the social networks is time consuming, it can offer an accurate perception of the consumer’s referral decision making.

The social networks created by call data are a close simulation of the actual social network (Eagle, Pentland, and Lazer, 2009). After the creation of that kind of social network and the investigation of communication between social network’s actors, one of the desirable outcomes is the prediction of whether a consumer will adopt or not a product or service. The social network data score higher in accuracy than any demographic data (Katona, Zubcsek and Sarvary, 2011). Kleinberg (2013) highlights that the advantage of phone call communication data is associated with identifying of a high number of interactions. Still it is difficult to comprehend the impact of these interactions on individual’s decisions, which seems to be a disadvantage of this method.

This study, by investigating the characteristics of social actors in the social network based on phone call data, will try to shed some light at this problem.

The aim of this research is to investigate which are the attributes of social actors or characteristics of a social tie between those actors, that are associated with an influence on the decision-making behavior, especially in adoption and defection decisions of another social actor. Such influential characteristics that have an impact on first adoption and then on the defection of the service by an actor, will help to answer the question whether the most profound motivation for someone to adopt or defect that service is the influence caused by a friend's profile, or it is associated with a strong relationship between those two individuals.

This research contributes to the current literature by providing a deep investigation, based on statistical and machine learning models, on customers' motivations on decision making and by analyzing their social networks on a dataset of communication interactions and adoption and defection of a mobile add-on service.

In the part of study that evaluated the behavior of social actors, in order to provide information about social links of the individuals, call data is used. Additionally, data concerned the adoption and defection of the caller tune (call-tune) service provided by a telecom company is used. This service is a low-involvement technological product with low risk. More specifically, the service allows the adopter's connections listening to a specific ring tune when awaiting the reply of the receiver. The nature of this particular service spreads the information not only in such a way that the adopter is discussing and hence influencing his social link, but the call itself is creating an influence, because the friend interacts with the service by hearing this fun-tune. Peres, Muller and Mahajan (2010) suggest that even if consumers do not directly share information about this service, by discussing about it, the fact of a friend's adoption is the information by itself.

The structure of that call data is created in form of dyads (micro-level). There is track of interactions within each dyad, volume of communication, demographic characteristics and – most importantly – the adoption and defection dates of the add-on service. The large amount of data gives the ability to generalize the findings, in a sense that the results can provide information for different data or different markets than the investigated one.

In order to materialize the analysis of the data, two statistical methods are used. The first method is the binomial logistic regression, which identifies the most relevant and significant variables for the prediction of the response variable and the direction of the relationship of each predictor with the dependent variable. Later, the method of classification decision trees is applied, to see not only the most important variables that are affecting the response variable outcome, but also the effect of each variable after taking into consideration the effect of the rest of the variables. The results of two methods applied to two different scenarios of adoption and defection are being explained and compared in order to reveal important insights for companies and marketeers.

This thesis is organized as follows. Section 2 describes the literature overview of the characteristics of social actors and their influence on the consumer behavior within the social network. Section 0 deals with the data, its structure and the variables which help in investigating the characteristics that can make a customer adopt, but also to defect the service based on a social tie with an adopter and defector of this service. Section 4 presents the research methodology. Section 5 describes main results of the empirical analysis. Finally, Section 6 summarizes the most important findings and provides a framework for further discussion and research on this topic.

2. Literature overview

The aim of the following section is to provide an overview of the current research on the subject of social network, ties of individuals within the network and the influence on consumer behavior. Later on, factors such as product involvement and other variables, which impose resistance to the social influence are investigated. Finally, different types of social links between individuals and characteristics which facilitate information sharing and consumer decision making are elaborated in the context of the current research.

Social network and social ties

Primarily, to facilitate the understanding of the forthcoming discussion on the literature overview it is necessary to familiarize with the topic of the social network and how is it structured. Iacobucci (2009) defines the social network as a system of nodes and links, which carries social interactions. The nodes are referred to as the individuals and the links to as ties or – better said – social ties. The ties include characteristics such as frequency and trust. More precisely, frequency describes the degree of communication between the individuals, trust the level of willingness to share the information. Ansari, Koenigsberg and Stahl (2011) suggest that ties can be of two formats, binary directed and weighted directed, with the distinction between them based on the existence or absence of the two-directional communication. Finally, Hartmann et al. (2008) emphasize the importance of tie activeness; especially, in the context of customer targeting by companies. Some ties are passive, which means that only one of the individuals gets influenced during the social interaction.

Literature indicates that many actors are interacting with a small number of people, while only a few interact with numerous (Onnela et al., 2007). On the contrary, only a small percentage of people is enjoying long conversations, when a majority is communicating for a limited time. The spread of information and the social influences among social ties with friends, family members and acquaintances and their impact on consumer's behavior is referred to as the "word-of-mouth" (Arndt, 1967). Flow of information on the product or service is creating the product awareness and is decreasing the risk brought by its acquisition. Different levels of risk trigger different levels of information seeking. Given that the product or service has a low risk, seeking information in the social network is less frequent, in comparison with a high-risked product or service.

Technological elevation has positively influenced the way in which people communicate and how information flows (Wuyts, 2010). Thanks to the social media and technological development, social ties become easier to be created, maintained and exhibit wider frontiers (Muller and Peres, 2019). It is important to mention here that, despite of the technological development, interactions within the social network are out of control of the company and the marketeers, since the actors are free to interchange positive or negative opinions about the product or service.

Diffusion of innovation within the social network

The social network has proven to be crucial for the spread of innovation. There appear to be two types of mechanisms in the social network: i) primarily they serve the purpose of information spread of the innovation; but also, ii) the purpose of persuasion to adopt the innovation (Wuyts, 2010). Based on several researches[1] it has been proven that the social network not only facilitates but also accelerates the spread of innovative ideas and applications. Muller and Peres (2019) explain which characteristics constitute a social network adequate for the diffusion of innovation. These comprise the cohesion, connectedness and conciseness. The first one requires that all actors are influencing and are influenced by others in the social network. The second characteristic is a large number of ties per actor. Finally, the last one demands different kinds of ties, where each tie has to facilitate the spread of information on innovation.

Social Contagion & Peer Influence

The social network and people's communication within it result in social influence, which depends on the kind of relationships between the individuals. Katona, Zubcsek and Sarvary (2011) define social influence as the reliance of social communication on decision making. Based on Assael (1998), individuals' behavior and decision making are influenced by external factors like culture, nationality and the social class they belong to. Moreover, social influences can result from different types of social relationships. Kleinberg (2013) suggests that these relationships can either be friendly or antagonistic. Because friendship is related to trust and easy flow of information, it

[1] See, for example, Valente and Davis (1999), Goldenberg et al. (2009) and Muller and Peres (2019).

exerts an influence on consumer's choices and tendencies. As Narayan et al. (2011) have said, peers' decisions seem to influence the actor's decisions. The influence by peers can be sufficiently strong to change the preferences and the willingness to pay for a product's attribute.

Katona, Zubcsek and Sarvary (2011) found that members in the same social network exert three times more powerful influence on consumer's decision making. People who communicate more frequently, such as family members or colleagues, have a stronger impact on consumer's choices and opinions, than people who communicate occasionally. As it has been proven, family members have a stronger impact on consumer choice. Friends' opinions have a slightly lower impact. Finally, Ma, Krishman and Montgomery (2015) argue that social influence is of a dynamic nature, which means that it can change its strength over time. Importantly, a previous influence of an individual can change in the future.

The influence becomes applicable when the consumer discusses and acquires a product or service that has already been adopted and accepted by his social network. As pointed by Narayan, Rao, Saunders (2011), peer influence and "word-of-mouth" are particularly important factors that influence consumer's behavior when the product purchase involves a higher uncertainty (i.e., in the case of the first use of the product or service, and when information about the product's characteristics is limited). As a matter of fact, if customers have already used the product or service, or if they can find information about it elsewhere, they will not seek for the friend's advice. But in the opposite scenario, they will likely contact their social network for advice.

Resistance to social influence

In contrast to the researches who demonstrated the power of social influence on individuals' decisions, Mourali and Yang (2013) investigated the effect of opposition (resistance) to this influence. Some people tend not to take into consideration other people's choices. This resistance is a way to express distinction and a pretended independence. A characteristic that is considered to be strongly associated with such a resistance, is the power, in a sense that, the more powerful an individual is, the more independent are his decisions. Another reason for such a resistance is the feeling of certainty, which makes the individuals to stop seeking for information inside their social network, because they already have it.

Involvement

Product involvement describes the degree of customer commitment and reflection which is exercised before product acquisition and has an impact on consumer's behavior in terms of information seeking and decision making. Different consumers demonstrate unique levels of involvement in various products; hence, they get influenced differently. Involvement can help in understanding their motives and behaviors. For example, a person with high involvement in the product may not seek for information and may not get influenced by his social network. Consumers, who are characterized by a high level of involvement, tend to communicate more often about their choices about the products and services; irrespective of the fact whether the feedback about the product is negative or positive (Assael, 1998).

Another finding is that the degree of involvement increases as the variety of options is higher, due to the higher risk associated with the number of options (Assael, 1998). One of the explanations of this phenomenon is that the customers need more information to decrease their risk of adopting the service or product. Moreover, involvement can be affected by the situation in which the product is purchased (Assael, 1998). Especially, the degree of involvement is higher in the cases of a high social pressure, which increases the level of self-consciousness. Finally, Reingen and Kernan (1986) demonstrate that the structure of the social network varies based on the service or product. Especially, in the context of a low involvement service, its low risk would be associated with insignificant information searching. In this research the analyzed adoption and defection decisions are related to such a service.

Adoption and Defection

Recently, due to the technological development and the accessibility of network data, researchers of many domains have investigated the impact of the diffusion process and service adoption. Among them, Hill, Provost and Volinsky (2006), based on the telecommunication data, revealed that customer communicating with an adopter of a particular service has an increased likelihood of adopting as well. This phenomenon is also shown in the work of Katona, Zubcsek and Sarvary (2011), based on data related to a web-based networking service. However, Boorman (1974), Friedkin (2006) and Centola and Macy (2007) mention that in many cases social influence of a single member of the social network may not be sufficient to convince the actor to adopt a product or service, as the likelihood adoption increases when more actors in the social network adopt the service. Another interesting topic related to the service adopting is its timing. As explained by Iyengar, Van de Bulte and Valente (2011), a consumer who has adopted in the past and then defected, has a lower probability to be satisfied in comparison with someone, who still is using the product or service.

Defection of an actor is found to have a strong influence on his social ties. Researchers (Haenlein and Libai, 2013; Nitzan and Libai, 2011) have shown that the fact of social's actor defection will exert an influence in the direction of defection. The actors in a social network, who are talking negatively about the product or service, are either considering to defect, or they have already done it. In addition, they are influencing others to do the same. Rationally, as the number of social ties increase, the number of defected neighbors in the social network is increasing as well. However, heavy users are less likely to be influenced by defected neighbors, because they rely on their own expertise regarding the product or service. Finally, the lower the cost of defection of the service the higher the likelihood that an actor will be influenced by his defected neighbor.

Landsman and Nitzan (2019) demonstrate that the two powers of adoption and defection are acting together within the social network and affect decisions of the customer. The exposure to adopters or defectors of a service is influencing the decision of an individual. An interaction with an adopter of the service will decrease the probability to defect, however as time passes the effect fades, and the probability of defection will increase again. Similarly, an exposure to a defector of the service will decrease the probability of the consumer to adopt. Finally, the researchers mention that the most recent social exposures are more powerful.

This research contributes to the literature by combing communication data with the adoption and defection data. In comparison with the previous research (e.g., Katona et al., 2011) this study assesses unbiased data, since it contains both information on individuals who decided to adopt the service, and on those who refrained from purchasing it. Additionally, in comparison with the work of Katona et al. (2011) this research analyzes the information on actual communication between the individuals, which was not available for these authors. Finally, previous researches (e.g., Katona et al., 2011 and Nitzan & Libai, 2011) have investigated the field of social influence on the adoption and defection, respectively; however, this research investigates both scenarios. Also, the methodology that was used to evaluate the data seems to be innovative, since in previous researches the machine learning models were not used.

Weak and strong ties

Granovetter (1973) indicates that social ties can be categorized into weak and strong ones. The first ones are the relationships with acquaintances and the second ones with friends and family. Wilcox and Stephen (2013) among others[2] emphasize the importance of a high involvement in social relationships in the definition of tie strength. Additionally, the more similarities two people share, the more probable is that they share a strong tie (Precker, 1952). Moreover, a strong tie should involve frequent interactions and communication (Granovetter, 1973). Rationally, a sporadic communication does not have the same influence as a frequent and long-term relationship. As a matter of fact, nowadays, the communication and its maintenance among all kind of ties become easier, the flow of information seems to depend on the strength of ties and not as much on the number of them.

The literature demonstrates that not all social relationships are related to the same information sharing and have the same degree of influence on the individual's decision (Granovetter, 1973 and Ansari, Koenigsberg and Stahl, 2011). The strength of the tie is responsible for the flow of information of high complexity, since a complex information requires strong interactions to be transferred (Hansen, 1999). Moreover, as pointed by Granovetter (1973) and Reingen and Kernan (1983), people are addressing firstly their strong ties for information seeking. People tend to adopt products and services, which are accepted by their social ties. Muller and Peres (2019) describe this phenomenon as the normative pressure. Tie strength is also important in the context of the moral hazard. Frenzen and Nakamoto (1993) found that individuals are not willing to spread information of moral hazard to weak ties. This fact does not occur in the case of strong ties. Finally, it follows that interactions within the social network enhance individual's self-esteem. Wilcox and Stephen (2013) suggest that this happens especially during an interaction of an actor with his strong ties.

[2] Refer also to Onnela et al. (2007) and Stadtfeld and Pentland (2015).

Granovetter (1973) was the first, who suggested that weak ties have the crucial role in social networks and subsequent research[3] has confirmed these findings. Most importantly, weak ties are useful as the communication channel enabling a quicker spread of information between not similar actors of the social network. They also enable spreading the information that otherwise would not be transferred. Onnela et al. (2007) explain that if the weak ties were absent in the social network, it would disintegrate. This is unlike the removal of strong ties, which would only cause a local failure. Therefore, weak ties are important for the maintenance of social network and strong ties are important to support the communication within the group locally, at the micro-level.

Katona, Zubcsek and Sarvary (2011) demonstrate that it is possible to find where the actor is located in his social network. According to the methodology of Muller and Peres, 2019, this location is primarily characterized by an average number of social ties of this actor. His position can also be predicted from the composition of these ties. Van de Bulte and Wuyts (2007) mentioned that tie strength has two components: the intensity, which is measured with the duration of calls and their interaction frequency, and the valence, which depends on how the actors are willing to exchange information. In comparison to the valence, which is not easily measurable, the frequency of communication can be observed in real call data. Relevant scientific articles[4] define the frequent communication as five calls per month during three-month period. On the other hand, Haenlein and Libai (2013) introduce a method of filtering the duration of call data and include only members who communicate more than 1% of the total duration of phone calls of an individual.

In this research the influence of the strong ties is investigated regarding the customers' decision upon service adoption and defection. Since information on the actual kind of relationships between the two individuals was not available, this research defines the strength of relationships basing on the frequency of their communication in comparison with such a frequency regarding other people.

[3] Refer to, e.g., Brown and Reingen (1987).
[4] Refer to e.g., Ma, Krishman, Montgomery (2015).

Homophily and heterophily

Homophily and heterophily are scientific terms which have an important role in the social network investigation. As described by Brown and Reingen (1987), homophily describes the social tie between actors, who share similar demographic characteristics such as age, gender and social class[5]. Friendship is the social tie with the highest level of these similarities (De Bruyn and Lilien, 2008). Homophily can also be found in similar professions (e.g. police officer and fireman) and there is a higher probability that those professionals will exhibit social interactions (Centola, 2015). A social tie with no demographic similarities is characterized by heterophily.

Literature has revealed that people prefer to be surrounded by others with similar characteristics (McPherson et al., 2001; Centola and van de Rijt, 2015). For example, Stadtfeld and Pentland (2015) mention that people need gender homophily to create a social tie with their partner's friend. Besides, when people are reducing the number of members in their social networks, the ones who remain share the homophily with an actor (Iacobucci, 2009).

De Bruyn and Lilien (2008) in their article, mention that when two individuals share similar demographic and/or social attributes they have also a higher probability of exchanging information. This is because the social interaction, information flow and hence the "word-of-mouth" is usually more frequent among this kind of people. The same research has found that these similarities are inspiring trust, which facilitates the flow of information, especially when the transmitted information requires reliability. Furthermore, Nitzan and Libai (2011) with their work, gave evidence that homophily influences customer retention. However, at the same time, homophily has a power of retarding innovations, as homophilic social networks are not exposed to innovations (Goldenberg et al., 2009). Both of these factors should be taken into consideration by companies in their customer targeting strategies.

As it has been mentioned before, social ties have an impact on adoption of innovative products and services. According to the finding of Berger and Heath (2007) people react positively to adoption of people with homophilic characteristics, contrary to the

[5] Social class is the hierarchical segmentation of people in a society. Inside the same segment people have a similar social status, while outside of it, the status is different. Social class is a mix of referral attributes such as income, education, profession and prestige (Assael, 1998).

adoptions of dissimilar individuals, where negative reactions are more common. Moreover, Brown and Reingen's (1987) research gives an evidence that actors are affected more by information provided by social ties characterized by homophily.

Despite the fact that many researches take into consideration homophily, while in terms of influence and product diffusion, heterophily is also very important. De Bruyn and Reingen (1987) explain that if there are no similar acquaintances, who have domain knowledge, the actor will search for information outside of his homophilic social circle. It is very possible that, in such a case the actor will be strongly influenced by these heterophilic social ties. Blau and Schwartz (2018) make further comment that homophily cannot be the only factor responsible for the maintenance of social network and that heterogeneity is crucial regarding its format. Some of the reasons for such a strong influence comprise difference of age and higher social class, which are also associated with power and expertise (De Bruyn and Lilien, 2008). For example, older people influence the younger ones (Katona, Zubcsek and Sarvary, 2011) and the individuals with a higher social status or class have an impact on the others (Strodtbeck et al., 1957).

Influence versus association

In the research of social influence within the social network, homophily plays a very important role, because it can hide relevant information. More specifically, after investigating the adoption decisions of individuals who share homophilic social ties, there will be always a question on the real motivation behind this adoption. The main question regards with the identification of whether the adoption results from the influence of social tie or it is simply the result of similar tastes and high level of homophily of two individuals. This phenomenon is related to the identification of the actor within his social network.

According to the research of Aral et al. (2009) the decision on service adoption can be attributed to homophily and not to the social influence in about 50% of the cases. Similarly, Ma, Krishman and Montgomery (2015) mention that if homophily were not taken into consideration during prediction of service adoption, then the effect of social tie influence can be overestimated. To solve this problem, the same researchers, took into consideration both the communication of dyads of actors and the level of homophily between them as the variables in their models.[6] Similarly, in this research, homophily is considered in the models, together with other attributes (such as age, social class and higher number of social ties) that can forecast influence.

Opinion Leaders

There are some specific attributes that increase the probability of an individual to become the opinion leader. Mainly, the opinion leaders either become leaders because of who they are, or because of their deep specific knowledge or because of strong influence they exert on the social ties (Weimann, 1994). Moreover, in many cases people with a higher financial and social status are opinion leaders (Weimann, 1991). Opinion leaders are also characterized by their willingness to give their advice and to share their knowledge and information about innovations (Muller and Peres, 2019). Goldenberg et al. (2009) have found that opinion leaders can also become early adopters, since they came across the product or service earlier. This occurs from their extended social network connections. Finally, it is worth to mention that relevant literature explains that opinion leaders in one field are not necessarily leaders in another (Merton, 1949; Myers and Robertson, 1972; Katz et al., 2017).

Opinion leaders are important for information diffusion. Van de Bulte and Yoshi (2007) demonstrated that opinion leaders are responsible for the transfer of information in a group of people of low social activity, like individuals with small number of members in their social network. However, the information flow concerning opinion leaders seems to be unidirectional, since Iyengar, Van de Bulte and Valente (2011) mention that actors who consider themselves as opinion leaders are less influenced by behavioral adoption and information flow of other actors.

Haenlein and Libai (2013) suggests that companies should focus not only on opinion leaders, but also on revenue leaders, who are clients with a high life-time value. In many cases those actors are not only revenue, but also opinion leaders. This assumption is based on the homophily effect, by which these revenue-opinion leaders will interact with other homogenous actors in their social network (Goldenberg et al., 2009). However, companies should be aware of the power of a group of opinion leaders and

[6] Similar methodology is followed by other researchers; confer, e.g., (Landsman and Nitzan, 2019).

their ability to help, but also to put obstacles in the contagion of a product, depending on their willingness (Goldenberg et al., 2009).

One kind of opinion leaders is the social hubs. As described by Goldenberg et al. (2009), social hubs are social network actors who are characterized by a large number of social links. Typically, social hubs create social ties in different types of circles. As the degree of an individual is increasing, this individual can influence more people (Goldenberg et al., 2005).

A suggested measure to quantify social hubs is the number of ties, which exceeds three times the standard deviations of such a number in the social network (Goldenberg et al., 2009). In this research it is mentioned that social hubs are more probable to adopt earlier, because of their exposure to a higher degree of connections (higher number of social ties). The same researchers categorize social hubs into the “innovative” and “follower” ones. Their results show that “innovative hubs” increase the speed of adoptions and the “follower hubs” increase the number of adopters.

The literature is not unanimous whether the social hubs have a strong influencing power on customers’ behavior. Social actors who are both early adopters and innovators may have a high influencing power on service adoption (Goldenberg et al., 2009). However, at the same time, Nitzan and Libai (2011) show that the higher number of the social ties an individual has, the lower influence he has on actor’s behavior. The same researchers mention that such individuals are more likely to defect the service, because of their higher exposure to the defection related to the extent of their social network. In order to investigate this controversial role of people with extremely high number of social members in their networks, this research characterizes such individuals as the social hubs and investigates their impact on the service adoptions and defections. In contrast to the previous papers, in order to further understand which individuals are more influential, this categorization is applied both to the already adopters/defectors and to their social neighbors.

3. Data description

In this section, the description of data used in the empirical analysis is presented. The final dataset contains call data and adoption data of the service and was created by merging the two sets of data provided by a telecom company.[7]

The first dataset contains communication data of active customers of a telecom company, over the 12-month period of 2008. The data provides information about the existence of communication between the provider's customers, and additionally the monthly duration of the calls and numbers of the text messages exchanged. Filtering was applied to investigate the adoptions of costumers in January 2008. Similarly, the dataset includes the connections that communicated with these costumers at least once in the period of four months from January to April 2008. These communication interactions are represented in a network that contains links. Every link constitutes the existence of the communication between a dyad of people. This dyad has two components or two social actors, the *"focal"* and the *"friend"*. *"Friend"* is the person with whom the *"focal"* customer (an individual who has already adopted or defected the service) has communicated at least once during that time.

The second dataset contains information about the relevant adoption and defection dates of all customers who adopted a particular add-on service in January 2008. Thanks to this dataset it is possible to identify and track adoptions and defections, in a format of day, month and year.

The add-on service, which is helping to investigate consumers' behavior on adoption and defection decision making, is called Caller Ring-Back Tone (CRBT), also known as *Fun-Tone*. It enables the customer to download a ring-back tone, which will be the tone, instead of the regular ring-back tone, which people who call this individual hear while awaiting the connection. At this point, it is useful to provide some descriptive information about this service. Firstly, the acquisition cost was 1.6 dollars per month, and there was an additional cost of 1 dollar for downloading a new song. Secondly, to defect the service, clients needed to communicate with a provider's representative.

[7] Those datasets have been also used in other papers, in order to describe customer's reaction to influence and in general consumer behavior: Nitzan and Libai (2011) (where the authors use the communication dataset, but not the dataset which tracks service adoption and defection) and Landsman and Nitzan (2019) (where both datasets are used).

Furthermore, the service is a low-involvement technological product with a low risk (see Section 2, sub-section on product involvement). Finally, there was no promotion of this service to attract new customers; hence the scenario in which two customers come across the same external stimuli that might influence their adoption or defection decision in the same period of time was very unlikely.

3.1.Variables definition

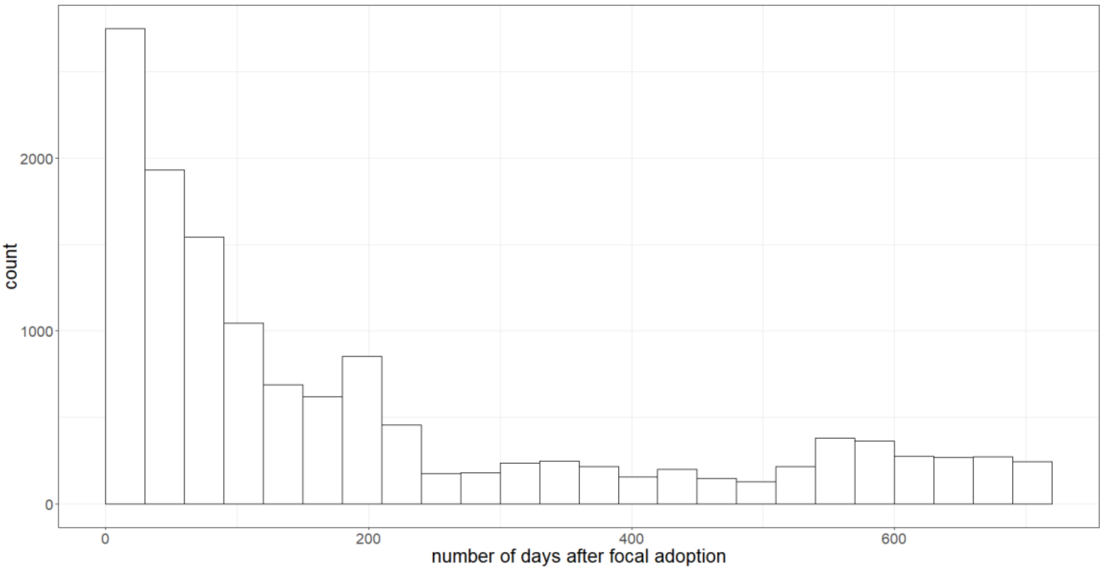
<u>Response variables</u>

This research investigates the characteristics that motivate and influence the add-on service's adoption and defection decision making. The respective response variable will be either the indicator of service adoption or defection. Firstly, the response variable that facilitates the adoption investigation is a categorical variable with the value of *"1"* when the *"friend"* adopted the service after the *"focal"* (i.e., after January 2008) and the value of "0" if he did not. Similarly, for the problem of defection decision making, the response variable is a categorical attribute, with the value of *"1"* when the friend defected after the *"focal"* and the value of *"0"* if he did not. Finally, in accordance to previous researches[8], different thresholds have been applied in order to investigate the *"dynamic effect"* of influence that a *"focal's"* adoption or defection decisions cause in her/his *"friend"*.

The average time after adoption of the add-on service by the *"friend"* is c. 197 days. Below, it is presented the count of friends who adopted the service within a specific number of days. Each bar refers to 30 days after the *"focal"* adoption.

[8] E.g., Landsman and Nitzan (2019).

Figure 1: Histogram of the number of days of *"friend"* adoption after the *"focal"*



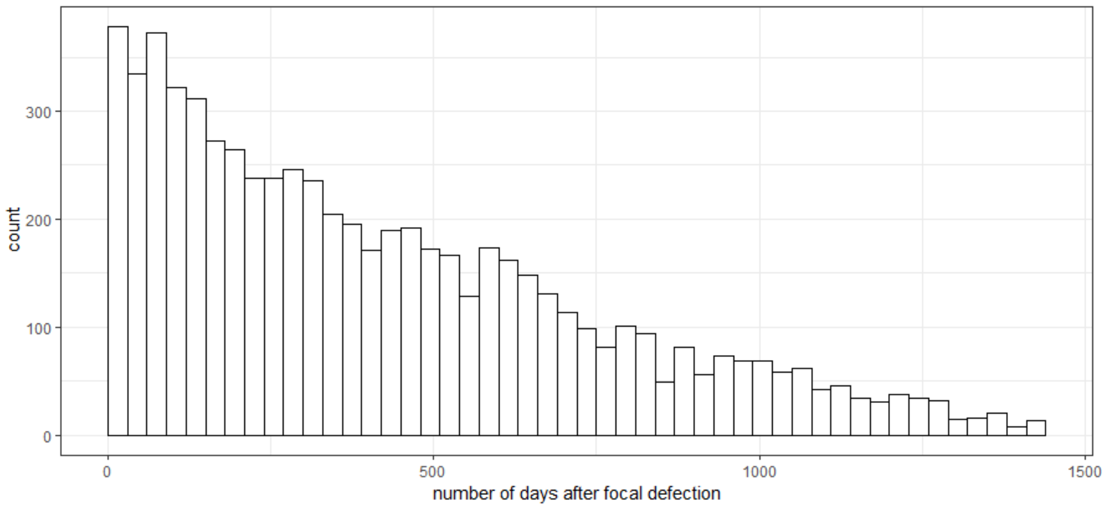
Source: Own analysis based on the data provided by telecom company.

Note: Observations where a *"friend"* never adopted the service were excluded.

The figure shows a negative relation between the number of *"friends"* adopting the service and the time from the moment of *"focal"* adopting the service. This relation seems to persist in short and medium terms (until 240 days after the adoption by the *"focal"*), since in the longer time period the number of *"friends"* adopting the service is relatively stable. As it seems, this observation can illustrate that the fact of *"focal"* adopting the service can indeed influence the *"friend's"* decision and make the latter to conform. As the strength of the social influence weakens with time, this research investigates different limits to the moment of adoption by the friend: 120 days, 240 days and 480 days, corresponding to the short, medium and long adoption horizons, accordingly.

The average time for *"friends"* to defect the service after the *"focal"* is about 420 days, about twice longer than that related to the adoption of the service. Then the number of friends who defected the service within a specific number of days after the "focal" is plotted.

Figure 2: Histogram of the number of days of *"friend"* defection after the *"focal"*



Source: Own analysis based on the data provided by telecommunication company.
Note: Observations where a *"friend"* defected before the *"focal's"* defection, never adopted or defected the service were excluded.

Based on the finding that on average *"friends"* take twice as much time to defect the service than to adopt it, and that in the long run all service adopters, at some point, will defect the service (due to e.g., termination of the service by the telecom company) the following time thresholds regarding the service defection are investigated: 240 days, 480 days and 960 days, corresponding to short, medium and long defection horizons, respectively.

Explanatory variables

Demographic characteristics: Four demographic variables, which describe both the *"focal"* and the *"friend"* are investigated in the models: i) "*age"*; ii) "*social class"*; iii) "*gender"*; and iv) *"segment"* (including indicators of "*young"*, "*student"* and four additional socio-demographic segments, defined by the telecommunication company). "*Age"* and "*social class*" are continuous variables. The second of these attributes can take a value from 1 to 10 and it approximates the individual's social status based on her/his zip code.[9] "*Gender"*, "*young"*, "*student*" and the remaining four *"segments"*

[9] The attribute of social class was defined by the telecommunication company and the exact methodology of its estimation is a proprietary information of the company.

are categorical variables, and they indicate respectively, the individual's gender[10], age less than 18 years old, the fact of being a student, and if she/he belongs to one of the four additional customer segments. For both the *"friend"* and the *"focal"* it is possible to belong to more than one, or none of the six *"segments"*.[11]

Volume of communication: The attribute of "*volume of communication*" describes the degree of communication of an individual within his entire social network, in the period of 12 months of the year 2008. The variable is created based on the duration of calls and the number of text messages that dyads have exchanged. Each message is considered as an equivalent of one minute of phone call and it is added to the total duration of calls. Due to the fact that the "*volume of communication"* is a continuous variable with frequent outliers[12], a logarithmic transformation of this attribute is performed.

Social Hub: In the dataset it is given the number of members (degree) in the social network of each *"focal"* and *"friend"*; hence, it is known with how many people a person interacted during the whole year. The literature (Goldenberg et al., 2009) gives the definition that as "*social hub"* is considered the actor, whose number of the social links exceeds three standard deviations of such a number in the social network. The variable "*social hub*" is created based on this definition.

Volume of intercommunication: Similarly, to the attribute of *"volume of communication"*, the *"volume of intercommunication"* is constructed based on the attributes of the "*amount of calls"* and "*number of text messages"* dyads, with the difference that this attribute corresponds to a specific *"focal"*-*"friend"* pair only. Hence, the variable measures the degree of communication within a dyad.

Homophily: The *"homophily"* variable provides information regarding how similar a dyad of people is. Homophily is able to provide more profound description about the relationship of the dyad. For the purpose of the assessment of "*homophily*" in the dyad, the demographic information is used. The information used includes four attributes: (i)

---

[10] The value of "0"/"1" indicates that an individual is a female/man. For c. 20% of the sample gender is a missing variable. In order to correct the lack of information, a random number (0/1) is attributed to those records.
[11] In particular, in the dataset c. 42.1%, 54.4%, and 3.4% of individuals belong to respectively none, one and more than one socio-demographic *"segments"*.
[12] High degree of outliers is evidenced by e.g., high skewness of c. 16.41 (based on the adoption dataset).

"*gender*"; (ii) "*age"*; (iii*) "consumer segment"*; and (iv) "*social class"*. In line with previous research (Landsman and Nitzan, 2019), that had defined "*homophily*" variable, scores of *"0.25"* are allocated every time when one of the four above-mentioned demographic attributes was the same for the *"focal"* and the *"friend"*. For example, the score is *"0"* for a dissimilar pair and *"1"* for dyads with the same four demographic characteristics.

Regarding the "*age*" attribute, the "*age score"* was considered identical if the actors had age difference of maximum two years[13]. By using an absolute age similarity, a lot of information is hold back. The auxiliary attribute of "*segment*" classification was provided by the telecom company and the customer could belong to one of six segments (see description of the demographic characteristics above). The auxiliary attribute of "*social class"* has been assigned based on the customer's postal code and takes values from 1 to 10. The "*homophily*" variable can have values: *"0"*, *"0.25"*, *"0.5"* or *"1"*.

Social superiority: "*Social superiority"* consists of two attributes as its components: i) "*age"*; and ii) "*social class"*. The criteria for its construction are as follows, whether the *"focal"* (*"friend"*) is older than the *"friend"* (*"focal"*) and whether he belongs to a higher social class than his *"friend"* (*"focal"*). In line with the definition of the "*homophily*" attribute, an individual who is older more than two years, than the person with whom she/ he communicates, is considered to be socially superior. Regarding the social class attribute, an individual is considered to be socially superior if his "*social class*" score is higher than that of the other person. If one such a criterion is met within the dyad, then the "s*ocial superiority's*" value will be *"0.5"* and if both of them are met, then the value will be *"1"*.

Strong tie: A strong tie is defined as frequent interaction and long-lasting between two people who in real life share a strong bond, like a friendship or a family bond. In the dataset the information on the kind of relationship a dyad shares is not identifiable. However, the amount of "*intercommunication"* is used as an approximation of real-life relationship between the social actors. In order to identify the existence of more frequent communication, it is necessary to define how much every actor communicates

[13] The choice of two years is supported by the research on age homophily of Fischer (1977), who indicated that a big proportion of individuals' closest friends (i.e., 38%) is within two years of their age. For more details on age homophily refer also to McPherson et al. (2001).

on average with all his friends, and to compare these numbers with the amount of communication within each tie. More specifically, if a dyad communicates during the 12-month period more than the average time the *"focal"* communicates with all the members in his social network, this bond will be considered as a strong tie.

3.2.Univariate and multivariate data analysis

Adoption dataset

The final dataset contains 39,083 records for 27 variables (3 response and 24 explanatory variables). The sample is constructed by excluding all *"friends"* who adopted the service prior to the *"focal"* (14,577 records). Out of total, for 7,271 (c. 18.6%), 9,894 (c. 25.3%) and 11,446 (c. 29.3%) observations describe the situation when a *"friend"* adopted the service after the *"focal"* in 120, 240 and 480 days, respectively.

In Table 1, summary statistics of the response and explanatory variables are presented. The table provides information on each variable's type, range, mean, standard deviation, (excess) kurtosis and skewness.[14] Variables are classified into three types: categorical, discrete and continuous. The mean is the simple average of all observations in the dataset. The standard deviation demonstrates how the data spread around the mean for every variable. Lower values of standard deviations demonstrate that the data of this attribute are distributed close to the mean value and higher numbers show a wider spread. The kurtosis and skewness give further insights on how the data are distributed.[15]

[14] Note that, for variables with binomial outcomes (the majority in the dataset) the measures of standard deviation, kurtosis and skewness are outcomes of the sample mean and the number of observations (i.e., $SD(p) = \sqrt{p * (1-p)}$, $K(p,n) = \frac{1-6p(1-p)}{np(1-p)}$, $Sk(p,n) = \frac{1-2p}{\sqrt{np(1-p)}}$). It follows that, the interpretation of reported outcomes should be limited to the analysis of the mean, since e.g., in comparison to continuous variables, kurtosis for binomial attributes (i.e., variables which are asymmetrically distributed) describes not only the level of tail-weight, like in the case of normally distributed variables, but also asymmetry.

[15] On the one hand, skewness measures symmetry in the data and more specifically, skewness of zero demonstrates perfectly symmetrical data. Negative (positive) skewness is evidence of longer tail on the left(right) side. On the other hand, kurtosis describes the shape of the tails that a variable's distribution has. Kurtosis higher(lower) than zero indicates leptokurtic (platykurtic) distribution.

Table 1: Response and explanatory variables – summary statistics (service adoption)

Variable name	Type	Min	Mean	Max	Std.Dev.	Kurtosis	Skew.
Response variables	-	-	-	-	-	-	-
Adoption (120 days)	Cat.	0.000	0.186	1.000	0.389	0.604	1.614
Adoption (240 days)	Cat.	0.000	0.253	1.000	0.435	-0.711	1.135
Adoption (480 days)	Cat.	0.000	0.293	1.000	0.455	-1.171	0.910
Explanatory Var. (focal)	-	-	-	-	-	-	-
Age	Cont.	4.000	32.057	85.000	11.395	0.658	0.742
Social Class	Cont.	1.000	4.178	10.000	2.603	-0.941	0.438
Gender	Cat.	0.000	0.446	1.000	0.497	-1.952	0.219
Young	Cat.	0.000	0.053	1.000	0.224	13.863	3.983
Student	Cat.	0.000	0.036	1.000	0.185	23.112	5.011
Segment (1)	Cat.	0.000	0.034	1.000	0.182	24.157	5.114
Segment (2)	Cat.	0.000	0.110	1.000	0.313	4.204	2.491
Segment (3)	Cat.	0.000	0.300	1.000	0.458	-1.241	0.871
Segment (4)	Cat.	0.000	0.077	1.000	0.266	8.115	3.180
Log(Volume com.)	Cont.	-2.120	7.681	10.913	1.185	2.859	-0.853
Social hub	Cat.	0.000	0.032	1.000	0.177	25.903	5.282
Social superiority	Disc.	0.000	0.350	1.000	0.347	-0.864	0.481
Explanatory Var. (friend)	-	-	-	-	-	-	-
Age	Cont.	1.750	32.121	88.000	10.374	1.779	0.825
Social Class	Cont.	0.041	4.276	10.000	2.599	-0.925	0.390
Gender	Cat.	0.000	0.448	1.000	0.497	-1.957	0.208
Young	Cat.	0.000	0.028	1.000	0.166	30.206	5.675
Student	Cat.	0.000	0.037	1.000	0.189	22.065	4.906
Segment (1)	Cat.	0.000	0.054	1.000	0.226	13.623	3.953
Segment (2)	Cat.	0.000	0.106	1.000	0.308	4.558	2.561
Segment (3)	Cat.	0.000	0.276	1.000	0.447	-0.992	1.004
Segment (4)	Cat.	0.000	0.091	1.000	0.288	6.042	2.836
Log(Volume com.)	Cont.	-3.912	7.269	11.456	1.496	4.448	-1.390
Social hub	Cat.	0.000	0.024	1.000	0.152	37.411	6.278
Social superiority	Disc.	0.000	0.356	1.000	0.338	-0.812	0.423
Explanatory Var. (dyad)		-	-	-	-	-	-
Log(Volume intercom.)	Cont.	-3.912	2.371	10.691	2.329	-0.250	0.358
Homophily	Disc.	0.000	0.399	1.000	0.281	-0.683	0.308
Strong tie	Cat.	0.000	0.247	1.000	0.431	-0.617	1.176

Source: Own analysis based on the data provided by telecommunication company.

Regarding the variables describing the demographic characteristics, above presented table shows the average "*age*" of individuals in the sample is c. 32 years, the *"social class"* score is about 4.2 and is highly variable, with the standard deviation of c. 2.6, there are slightly more women than men in the sample, and *"segment 3"* seems to be the most numerous out of the segments defined by the telecom company. The "*log(volume of communication)*" of *"focals"* is higher than for *"friends"* (however, both variables seem to be highly variable); thenceforth, *"focals"* are more frequently categorized as "*social hubs"*. Finally, the average *"social superiority"* score value for both *"focals"* and *"friends"* is around 0.35. Regarding the variables describing the dyads, on average the "log(*volume of intercommunication)*" amounts 2.371, the average *"homophily"* score is c. 0.40 and c. 25% of dyads are characterized as strong ties.

In Table 2 the correlation matrix for the explanatory variables is presented. The Pearson's correlation coefficients measure the strength and the direction of a linear relation between two attributes.

Table 2: Explanatory variables – correlation matrix (service adoption)

	1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
Variables (focal)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1. Age	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2. Social Class	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3. Gender	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4. Young	-0.3	0.1	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5. Student	-0.1	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6. Segment (1)	0.0	0.0	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7. Segment (2)	-0.3	0.1	0.0	0.1	0.0	-0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8. Segment (3)	-0.1	-0.6	-0.1	-0.1	0.0	-0.1	-0.2	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9. Segment (4)	0.1	0.1	0.0	0.0	0.0	0.0	-0.1	-0.2	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10 log(Vol. com.)	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11 Social hub	0.0	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12 Social sup.	0.4	0.3	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Variables (friend)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13 Age	0.2	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	0.1	0.0	0.0	-0.3	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14 Social Class	0.1	0.6	0.1	0.1	0.0	0.0	0.1	-0.5	0.1	0.0	0.0	-0.1	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-
15 Gender	0.0	0.1	0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-
16 Young	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.1	-0.2	0.1	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-
17 Student	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.1	-0.1	0.1	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-
18 Segment (1)	0.0	0.0	0.0	0.0	0.0	0.3	0.0	-0.1	0.0	0.0	0.2	0.1	0.0	-0.1	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-
19 Segment (2)	-0.1	0.1	0.0	0.1	0.0	0.0	0.2	-0.2	0.0	0.1	0.0	0.1	-0.3	0.1	0.0	0.1	0.0	-0.1	1.0	-	-	-	-	-	-	-	-
20 Segment (3)	-0.1	-0.5	-0.1	-0.1	0.0	-0.1	-0.2	0.7	-0.2	-0.1	0.0	0.0	-0.2	-0.6	-0.1	-0.1	0.0	-0.1	-0.2	1.0	-	-	-	-	-	-	-
21 Segment (4)	0.1	0.1	0.0	0.0	0.0	0.0	0.0	-0.2	0.5	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	-0.1	-0.2	1.0	-	-	-	-	-	-
22 log(Vol. com.)	0.0	0.0	0.0	0.0	0.0	0.0	0.1	-0.1	0.0	0.2	0.0	0.1	-0.1	0.0	0.0	0.0	0.1	0.1	0.1	0.0	0.0	1.0	-	-	-	-	-
23 Social hub	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	1.0	-	-	-	-
24 Social sup.	0.4	-0.1	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.3	-0.2	0.3	0.0	0.1	0.1	0.0	0.1	-0.1	0.0	0.1	0.0	1.0	-	-	-
Variables (dyad)	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25 log(Vol. inter.)	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.2	0.0	-0.1	1.0	-	-
26 Homophily	-0.1	-0.3	-0.1	-0.1	0.0	0.0	0.0	0.4	0.1	0.0	0.0	-0.3	-0.1	-0.3	0.0	0.0	0.0	0.0	0.0	0.5	0.0	0.0	0.0	-0.3	0.2	1.0	-
27 Strong tie	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	-0.1	0.6	0.1	1.0

Source: Own analysis based on the data provided by telecommunication company.

Variables moderately to highly correlated[16] with other attributes are, for both the focal and the friend: *"social class"*, *"segment (3)"*, *"segment (4)"*, for the dyad: *"homophily"*, *"log(volume of intercommunication)"* and *"strong tie"*. The attributes of *"social class"* and *"segment (3)"* are highly negatively correlated, which can be explained by the possible segmentation of customers by using their social class. Another pair of correlated variables is the *"segment (3)"* for the *"focal"* and the *"segment (3)"* for the *"friend"*. The correlation can be explained through the phenomenon of homophily, where social neighbors are member of the same social class. This explanation seems to be confirmed through a high correlation of *"homophily"* with those attributes. For the same reason, the variables of *"segment (3)"* and *"segment (4)"* for the focal can be correlated with, respectively, *"segment (3)"* and *"segment (4)"*, for the *"friend"*. Finally, the variable *"log(volume of communication)"* and *"strong tie"* are moderately correlated, as the second attribute is created based on the first.

Defection dataset

Similarly, the defection decision making is investigated, with response variable the categorical attribute of defection, in 240, 480 and 960 days after the *"focal's"* defection. In contrast to the adoption, the observations will consist of different dyads of people, as it is necessary for the *"friend"* to adopt the service first, in order to be able to defect it. For this reason, 25,290 records were excluded from the dataset. Additionally, *"friends"* who defected before the *"focal"* were excluded, in order to analyze the influence of a *"focal"* on the decision making of his friends and to keep the study consisted with the part of the paper discussing the adoption of the service. The final dataset contains 14,022 records, among which for 4,666 (c. 33.3%), 7,573 (c. 54.0%) and 10,658 (c. 76.0%) a *"friend"* defected the service in 240, 480 and 960 days, respectively.

In Table 3, the summary statistics of the response and explanatory variables for the defection dataset are presented.

[16] Moderate and high correlation is referred to as the positive, or negative, correlation between attributes, which is higher than respectively 0.5 and 0.7.

Table 3: Response and explanatory variables – summary statistics (service defection)

Variable name	Type	Min	Mean	Max	Std.Dev.	Kurtosis	Skew.
Response variables	-	-	-	-	-	-	-
Defection (240 days)	Cat.	0.000	0.333	1.000	0.471	-1.496	0.710
Defection (480 days)	Cat.	0.000	0.540	1.000	0.498	-1.974	-0.161
Defection (960 days)	Cat.	0.000	0.760	1.000	0.427	-0.516	-1.218
Explanatory Var. (focal)	-	-	-	-	-	-	-
Age	Cont.	4.000	31.168	80.000	10.779	0.630	0.759
Social Class	Cont.	1.000	3.650	10.000	2.543	-0.694	0.690
Gender	Cat.	0.000	0.423	1.000	0.494	-1.904	0.310
Young	Cat.	0.000	0.047	1.000	0.212	16.196	4.266
Student	Cat.	0.000	0.033	1.000	0.179	25.120	5.208
Segment (1)	Cat.	0.000	0.014	1.000	0.117	67.282	8.323
Segment (2)	Cat.	0.000	0.101	1.000	0.301	5.014	2.648
Segment (3)	Cat.	0.000	0.427	1.000	0.495	-1.913	0.296
Segment (4)	Cat.	0.000	0.060	1.000	0.237	11.754	3.709
Log(Volume com.)	Cont.	-1.609	7.599	10.913	1.251	1.972	-0.684
Social hub	Cat.	0.000	0.023	1.000	0.151	37.638	6.296
Social superiority	Disc.	0.000	0.369	1.000	0.350	-0.915	0.411
Explanatory Var. (friend)	-	-	-	-	-	-	-
Age	Cont.	4.000	30.149	86.000	9.563	1.556	0.788
Social Class	Cont.	0.306	3.619	10.000	2.486	-0.641	0.682
Gender	Cat.	0.000	0.453	1.000	0.498	-1.965	0.187
Young	Cat.	0.000	0.032	1.000	0.175	26.752	5.362
Student	Cat.	0.000	0.033	1.000	0.179	25.056	5.201
Segment (1)	Cat.	0.000	0.012	1.000	0.108	79.982	9.054
Segment (2)	Cat.	0.000	0.110	1.000	0.313	4.222	2.494
Segment (3)	Cat.	0.000	0.443	1.000	0.497	-1.948	0.229
Segment (4)	Cat.	0.000	0.069	1.000	0.253	9.618	3.408
Log(Volume com.)	Cont.	-0.916	7.612	11.456	1.268	2.197	-0.706
Social hub	Cat.	0.000	0.020	1.000	0.141	44.562	6.823
Social superiority	Disc.	0.000	0.356	1.000	0.333	-0.783	0.405
Explanatory Var. (dyad)	-	-	-	-	-	-	-
Log(Volume intercom.)	Cont.	-3.912	2.365	10.644	2.339	-0.304	0.332
Homophily	Disc.	0.000	0.440	1.000	0.283	-0.740	0.159
Strong tie	Cat.	0.000	0.266	1.000	0.442	-0.881	1.058

Source: Own analysis based on the data provided by telecommunication company.

By comparing the above table with the results obtained for the service adoption dataset, it is demonstrated that dyads are more homophilic and are more often characterized as *"strong ties"*. Additionally, the *"focals"* and *"friends"* usually belong to a lower *"social class"* and *"segment (3)"*. These observations can indicate the likely characteristics of the service adopters, as the defection dataset excludes all observations for which the individuals have not adopted. This fact will be further investigated in the following sections of the research.

Table 4 provides values of correlation among the variables for the defection dataset. Similarly, to the correlation matrix for the adoption dataset (see Table 2), the most noteworthy values of correlation are observed for *"social class"*, *"segment (3)"*, *"segment (4)"*, *"homophily"*, *"log(volume of intercommunication)"* and *"strong tie"*. As it has been discussed above, in the section describing the dataset of service adoption, the reason for observing the correlation between these variables can be attributed to the methodology of their construction. Fluctuations in these values of correlation between the two datasets are negligible and the directions of the relationships are aligned.

Table 4: Explanatory variables – correlation matrix (service defection)

		1.	2.	3.	4.	5.	6.	7.	8.	9.	10.	11.	12.	13.	14.	15.	16.	17.	18.	19.	20.	21.	22.	23.	24.	25.	26.	27.
Variables (focal)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
1.	**Age**	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
2.	**Social Class**	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
3.	**Gender**	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
4.	**Young**	-0.3	0.1	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
5.	**Student**	-0.1	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
6.	**Segment (1)**	0.0	0.0	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
7.	**Segment (2)**	-0.3	0.2	0.0	0.2	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
8.	**Segment (3)**	-0.1	-0.6	-0.1	-0.1	0.0	-0.1	-0.3	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
9.	**Segment (4)**	0.1	0.1	0.1	0.0	0.0	0.0	0.0	-0.2	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
10	**log(Vol. com.)**	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
11	**Social hub**	0.0	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
12	**Social sup.**	0.4	0.3	0.0	-0.1	-0.1	0.0	-0.1	-0.2	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Variables (friend)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
13	**Age**	0.1	0.1	0.0	0.0	0.0	0.0	-0.1	-0.1	0.1	0.0	0.0	-0.3	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-	-
14	**Social Class**	0.1	0.6	0.1	0.1	0.0	0.0	0.2	-0.5	0.1	0.1	0.0	-0.1	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-	-
15	**Gender**	0.0	0.0	0.1	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.1	1.0	-	-	-	-	-	-	-	-	-	-	-	-
16	**Young**	0.0	0.1	0.0	0.1	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	0.1	-0.2	0.1	0.0	1.0	-	-	-	-	-	-	-	-	-	-	-
17	**Student**	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-	-
18	**Segment (1)**	0.0	0.0	0.0	0.0	0.0	0.2	0.0	-0.1	0.0	0.0	0.1	0.0	0.0	0.0	0.0	0.0	0.0	1.0	-	-	-	-	-	-	-	-	-
19	**Segment (2)**	-0.1	0.2	0.0	0.1	0.0	0.0	0.3	-0.3	0.0	0.1	0.0	0.1	-0.3	0.2	0.0	0.1	0.0	0.0	1.0	-	-	-	-	-	-	-	-
20	**Segment (3)**	-0.1	-0.5	-0.1	-0.1	0.0	-0.1	-0.2	0.7	-0.2	-0.1	0.0	0.0	-0.1	-0.6	-0.1	-0.1	0.0	-0.1	-0.3	1.0	-	-	-	-	-	-	-
21	**Segment (4)**	0.1	0.1	0.0	0.0	0.0	0.0	0.0	-0.2	0.5	0.0	0.0	0.0	0.2	0.1	0.0	0.0	0.0	0.1	-0.1	-0.2	1.0	-	-	-	-	-	-
22	**log(Vol. com.)**	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.2	0.0	0.1	-0.1	0.0	0.0	0.0	0.1	0.0	0.1	0.0	0.0	1.0	-	-	-	-	-
23	**Social hub**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.0	0.1	1.0	-	-	-	-
24	**Social sup.**	0.4	-0.1	0.0	-0.1	0.0	0.0	-0.1	-0.1	0.0	0.0	0.0	0.3	-0.3	0.3	0.0	0.1	0.1	0.0	0.1	-0.1	0.0	0.0	0.0	1.0	-	-	-
Variables (dyad)		-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
25	**log(Vol. inter.)**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.2	0.0	-0.1	0.0	0.1	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	0.2	0.0	-0.1	1.0	-	-
26	**Homophily**	-0.2	-0.4	-0.1	-0.1	0.0	0.0	0.0	0.5	0.0	0.0	0.0	-0.3	-0.1	-0.4	-0.1	0.0	0.0	0.0	-0.1	0.5	0.0	0.0	0.0	-0.3	0.1	1.0	-
27	**Strong tie**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	-0.1	0.0	-0.1	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.1	0.0	0.0	0.6	0.1	1.0

Source: Own analysis based on the data provided by telecommunication company

4. Methodology

The main research topic discussed in this paper is the analysis of the customer characteristics, which influence his decision on adopting and defecting the add-on service. In particular, the analysis will be focus at the identification and then at an assessment of the factors that drive the customer's decision. The aim is to describe the adoption and defection of the add-on service – a categorical variable *y*, (exhibiting the value of "1" when the service is adopted/defected, and the value of "0" when it is not), as a function of a set of independent variables *x*. Mathematically this can be described as:

$$y = f(x)\,. \tag{1}$$

In the context of machine learning and statistics this problem is known as statistical classification. In this research two widely used classification models are used: (i) logistic regression; and (ii) classification decision trees. Apart from the popularity in both research industry and professional sector, the advantage of these methods lays within the interpretability of their outputs, as statistical inference is the main purpose of this research.

4.1. Logistic Regression

Logistic regression is a classification technique which is used when the response variable is categorical or qualitative. It helps to find and measure the relationship among this response variable and a number of predictors, by using the logistic function to estimate probabilities. In logistic regression to fit the model the maximum likelihood method is used, a general approach for parameter estimating, which description is considered to be outside of the scope of this research paper[17]. Its goal is to maximize the likelihood of correct classification of each data record.

As a matter of fact, logistic regression models the probability of the response variable belonging to a specific category. That is how the prediction is made, with the help of sigmoid function, a S-shaped curve which transforms and maps continuous values to values between 0 and 1. Later on, these values will be transformed to "0" or "1", based on a set-up threshold. For probabilities higher than "0.5" or another set up threshold,

[17] For further details refer to for example (James et al., 2017).

the response variable will have the label “1”, and for probabilities lower than “0.5” it will be labeled as “0”.

Interpretation of logistic regression results

To interpret logistic regression, a linear relationship between the independent variables and log-odds of the event of adoption/defection of the add-on service is assumed. Mathematically this can be written as:

$$\log\left(\frac{p}{1-p}\right) = \beta_0 + \beta_1 x_1 + \beta_2 x_2 \,, \qquad (2)$$

where p denotes the probability of the event happening (i.e., y=1) and β_i describes parameters (coefficients) of the model.

Logistic regression provides coefficients for the predictors based on the maximum likelihood method. The value of the coefficient corresponds to an increase of the log odds of the event happening based on a variable increase by one unit. The sign of the coefficient shows if the effect of the attribute on the response variable is positive or negative.

A large (absolute) value of the z-statistic associated with the attribute’s coefficient value gives evidence against the null hypothesis, which states that the parameter equals “0” and this predictor has no effect on the response variable. In the context of logistic regression coefficient attributes, z-value is estimated by dividing the coefficient value by its standard error. If a coefficient has a large standard deviation, it will have a higher standard error, and a smaller z-value.

Highly related concept to the z-statistics are p-values, which demonstrate the existence or absence of a relationship between the response variable and the predictor. P-values describe the probability of observing test results (in this case, z-statistics) at least as extreme, as the results obtained when performing statistical testing, under the assumption of the null hypothesis being correct.

Another diagnostic for the analysis of logistic regression results is the receiver operating characteristic curve (ROC curve). It is a graphical depiction of the true positive rate and the false positive rate of a binary response variable. At this point it is important to define true positive rate as the proportion of actual positives correctly identified as positives. Accordingly, false positive rate is a test result, which indicates how many negatives were classified as positives. The area under the (ROC) curve (AUC) gives insights about the model's predictive ability. A value near to "1" demonstrates a good predictive ability and a value equal to "1" demonstrates that the model classifies observations perfectly. On the contrary, a value of "0.5" demonstrates that the model classifies the records randomly.

Advantages and disadvantages of the logistic regression

Logistic regression is a method widely used in both scientific research and within the applied sector. The method does not require an extensive computation power and it offers an easy interpretation. In the method it is easy to regularize and to put penalties on coefficients. It gives the opportunity to remove less significant and irrelevant attributes for the prediction of the response variable, to make the model more efficient. Finally, it is broadly used as a benchmark for more complex and less interpretable algorithms.

A statistical method never comes with no disadvantages. Logistic regression brings with it a vulnerability to overfitting, the phenomenon of much better fitting of the train set in comparison to the test set. Moreover, the method is not applicable to solving non-linear problems, since it describes the linear relationship between the log of odds ratio and a set of dependent variables.

4.2. Classification Decision Trees

Classification decision trees is a widely used supervised machine learning algorithm, which is used to solve classification problems. The algorithm can solve both prediction and description problems; however, prediction seems to be the most common application of decision trees methodology (Song and Ying, 2015). The objective of the algorithm is to classify the observations by posing ordered questions, which are related to the attributes of the dataset. The form of questions (and their order) is decided by the algorithm by minimizing the so-called impurity measures. At the bottom of the tree, the method performs classification based on the frequency of class occurrence (i.e., the label that has the majority of votes is considered to be the final label of the node) (Tan et al. 2013).

Structure of the tree

The decision tree has three types of nodes. The first type refers to the node on the top of the tree with no incoming links and is called the root node. The second type of nodes is the internal node, with one incoming arrow-link and two or more outcoming links. The last type of nodes is the leaf node, which is the last node of the tree and it gives the label or class for each record of the data as it does not have any outcoming links (Tan et al. 2013; Song and Ying, 2015). Another term used for trees is the parent node, which is followed by a child node, and it is used to demonstrate the node order in the tree. The tree consists not only of nodes, but also of branches which are the links among nodes. Each branch carries an answer to if-condition (*"yes"* or *"no"*) that leads to the following node, based on an attribute's value (Song and Ying, 2015).

Splitting

The decision tree algorithm progressively splits the dataset into smaller and smaller subsets with a higher degree of purity (i.e., the subsets where the response variable has the same class label). The binary splitting, with two outcoming links is the most common. On each node a variable or its level is provided, which can be expressed as an "if-condition". If the condition is satisfied then the tree leads to the right branch and if not, to the left branch.

The splitting is based on the impurity measures, e.g., Gini index, Entropy and Classification error, among which Gini index is the most commonly used measure and as such it is used in this research. The goal is to minimize the impurity value, as the purest class is the preferable. Below, the formula of Gini index, for a classification problem with *c* labels, is provided:

$$Gini = 1 - \sum_{i=1}^{c} p_i^2, \qquad (3)$$

where p_i corresponds to the proportion of the sample (probability) that belong to label i.

Because the number of observations in each of the nodes might be different, the algorithm calculates both the Gini index for the left and for the right branch and then it calculates the weighted average of those impurities. The process continues similarly for all of the nodes. The process is repeated until it finds the purest class and it puts that attribute in the root node. The splitting varies based on the type of the variable. If the variable has two (binary) values the right branch relates to the occurrence of the one value and the left branch to the other. If the variable is continuous or categorical the algorithm finds the most efficient splitting, which minimizes the Gini index.

Stopping criterion

The pruning technique is applied in order to decrease the growing of the tree and to keep the high accuracy on the model by avoiding overfitting. The complexity parameter (cp) can be used for this purpose, which defines the threshold for the cost of adding one more variable to the decision tree. The measure is based on the misclassification rate and it adds a penalty term multiplied by the size of the tree. The desired level of the complexity parameter can be selected through the optimization of the accuracy measured by cross-validation, which trains the tree in a different sample of data (train set) and then it tests its accuracy in the test set.

Advantages and disadvantages of the decision trees

Decision trees give the advantage of no need of prior assumptions, easy interpretation, possibility of graphical visualization and ease to be understood, because they simulate the human decision making (James et al., 2017). The useful advantage is that trees demonstrate which are the most important variables, as these are used in the nodes and the splitting. It is, also, a robust technique in terms of dealing with outliers (Song and Ying, 2015).

Decision trees have the advantage of dealing with missing values (Vaughn and Wang, 2008). There are different ways to deal with missing data. Firstly, the tree can create a separate category of the missing one or it can assign probabilities to the possible categories or it can fill the missing value with surrogate values based on other similar records.

On the other hand, decision trees do not offer as high accuracy as other methods do. Its disadvantage is that it is not a robust technique, because small perturbation to the data will affect significantly the estimated tree (James et al., 2017). Lastly, there is danger of overfitting which is the phenomenon of low error rate in the train set and high error rate in the test set. As a matter of fact, Tan et al. (2013) suggest creating a larger training set in order to avoid overfitting.

5. Results

This research is dedicated to the investigation of factors responsible for consumer's motivation of adopting or defecting a specific service and conforming to her/his social neighbor. Two different scenarios are investigated, the first is the conformity of the *"friend"* with *"focal"* adopter of the service, and second is the conformity of the *"friend"* with *"focal"* defector. For this purpose, different attributes which distinguish individuals, such as *"age"*, *"gender"*, *"social class"*, as well as characteristics among dyads of people are put under the microscope.

Two types of models are applied. The first is the logistic regression model, which illustrates if there is a relationship between the response variable and each one of the input predictors. Additionally, it illustrates if this relationship is significant and if there is a positive or negative influence of the predictor on the response variable. The second model is the classification decision tree model. Decision trees have the ability to provide information on which variables are the most important, by showing them closer to the root node. Additionally, with this method it is possible to discover the relationship of a predictor with the response variable, by considering at the same time the effect of another predictor. This is an advantage that logistic regression does not provide, as long as additional variables, which describe the interaction effect of attributes, are not introduced to the model.

5.1. Adoption of the service

Table 5 provides results of the logistic regression model, where the response variable is service adoption by the *"friend"* 120, 240 and 480 days after the *"focal"*. The table presents the coefficient of each predictor included in the model, the corresponding z-values with an indication if the coefficient is statistically significant at the 1%, 5%, or 10% level, and measures of (generalized) variance inflation factor[18] (GVIF). The table

[18] A variance inflation factor (VIF) for i^{th} explanatory variable is measured as $VIF_i = 1/(1 - R_i^2)$, where R_i^2 is the R^2 measure (McFallen's R^2 for binomial logistic regression problems) obtained by regressing the i^{th} predictor against the remaining explanatory variables. VIF is a frequently used measure, which helps in quantifying multicollinearity in regression problems. As a rule of thumb, a problem of multicollinearity is evidenced when VIF exceeds 10 (see, e.g., Kleinbaum et al.,1988). As pointed by Fox and Monette (1992) the VIF measure is not fully applicable to models that include regressors which are internally related, such as dummy variables. For this reason, the authors generalized the notion of

is supplemented with the regression diagnostics measured by Area Under Roc Curve (AUC), and measures of information criteria: Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC).

Table 5: Logistic regression model – results (service adoption)

	Adoption (120 days)			Adoption (240 days)			Adoption (480 days)		
Variable name	**Coef.**	**Z-val.**	**GVIF**	**Coef.**	**Z-val.**	**GVIF**	**Coef.**	**Z-val.**	**GVIF**
Intercept	-0.097	-0.6	-	-0.434	-3.2***	-	-0.621	-4.7***	-
Var. (focal)	-	-	-	-	-	-	-	-	-
Age	-0.024	-12***	1.857	-0.018	-10***	1.887	-0.015	-9.6***	1.904
Social Class	-0.061	-5.4***	3.689	-0.048	-4.8***	3.747	-0.042	-4.4***	3.772
Gender	0.091	3.2***	1.030	0.076	2.9***	1.033	0.069	2.7***	1.034
Young	-0.747	-8.6***	1.090	-0.556	-7.8***	1.103	-0.438	-6.8***	1.112
Student	-0.380	-4.5***	1.024	-0.324	-4.3***	1.027	-0.301	-4.2***	1.028
Segment (1)	-1.005	-6.3***	1.033	-0.931	-7.1***	1.032	-0.743	-6.8***	1.049
Segment (2)	-0.066	-1.2	1.293	-0.022	-0.5	1.286	-0.018	-0.4	1.280
Segment (3)	0.342	7.4***	2.753	0.359	8.5***	2.695	0.349	8.5***	2.666
Segment (4)	0.298	4.4***	1.524	0.249	4.1***	1.473	0.193	3.4***	1.469
log(vol. com.)	-0.045	-3.7***	1.162	-0.041	-3.6***	1.162	-0.044	-4.1***	1.162
Social hub	-0.427	-4.4***	1.026	-0.366	-4.3***	1.029	-0.342	-4.3***	1.033
Social super.	0.035	0.5	2.627	-0.017	-0.3	2.680	0.001	0.0	2.709
Var. (friend)	-	-	-	-	-	-	-	-	-
Age	-0.020	-9.8***	1.714	-0.019	-10***	1.707	-0.018	-10***	1.711
Social Class	-0.107	-9.5***	3.643	-0.098	-9.8***	3.658	-0.090	-9.5***	3.661
Gender	0.138	4.8***	1.017	0.123	4.7***	1.018	0.119	4.8***	1.018
Young	-0.550	-5.1***	1.064	-0.595	-6.3***	1.063	-0.392	-4.7***	1.073
Student	-0.803	-8.6***	1.029	-0.769	-9.6***	1.033	-0.719	-9.7***	1.036
Segment (1)	-1.848	-12***	1.048	-2.019	-15***	1.048	-1.663	-16***	1.078
Segment (2)	-0.098	-1.7*	1.378	-0.018	-0.4	1.380	-0.007	-0.1	1.374
Segment (3)	0.676	14.1***	2.927	0.867	19.8***	2.836	0.912	21.6***	2.793
Segment (4)	0.108	1.7*	1.530	-0.023	-0.4	1.476	-0.002	0.0	1.472
log(vol. com.)	0.076	7.7***	1.119	0.143	15.1***	1.123	0.150	19.2***	1.129
Social hub	-0.556	-4.7***	1.019	-0.432	-4.3***	1.022	-0.322	-3.5***	1.028
Social super.	0.088	1.3	2.476	-0.019	-0.3	2.536	-0.048	-0.8	2.573
Var. (dyad)	-	-	-	-	-	-	-	-	-
log(vol. inter.)	0.001	0.1	1.888	-0.009	-1.2	1.888	-0.008	-1.1	1.892
Homophily	-0.083	-1.2	1.984	-0.185	-3.0***	1.948	-0.124	-2.1**	1.908
Strong tie	0.205	5.0***	1.708	0.216	5.7***	1.715	0.217	5.9***	1.721
Diagnostics		-			-			-	

variance inflation to models with related explanatory variables and introduced the measure of generalized variance inflation factor (GVIF) – refer to their paper for more details.

AUC	0.7608	0.7643	0.7582
AIC	32314.18	37473.94	40233.29
BIC	32554.23	37714.00	40473.35

Source: Own analysis based on the data provided by telecom company.

The metric of AUC for all three models have a high value of around 0.76 and proves that the classification of the records does not happen randomly. The values of GVIF are not high for any of the variables; hence, the problem of multicollinearity in the models may not be severe, despite some of the attributes being correlated (see Section 3.2 on univariate and multivariate data analysis).

Demographic characteristics of both the *"friend"* and *"focal"* are proven to be important factors for the first to adopt the service. Hence, it is evidenced that not only the characteristics of the individual who adopts the service, but also the attributes of the person who already adopted the service (*"focal"*) and with whom the *"friend"* communicates, are important factors for the decision making. However, while the (positive versus negative) impact direction are aligned for different attributes, demographic characteristics of the *"friend"* usually have a stronger effect on his service adoption. These observations are in line with the research of Katona et al.(2011), who suggested that information about demographic characteristics and social network are useful factors to identify both influential people and profiles of people who are likely to adopt.

It is found that being a man has a positive impact on service adoption for both *"focal"* and *"friend"*. This observation supports the research of Depret and Fiske (1993), who showed that men were proved to be more influential. However, at the same time Katona et al. (2011), studying online social networks, found a contradictory relationship in their data, where women seemed to be more influential than men.

Similarly, younger *"focals"* have higher influential power regarding *"friend's"* decision to adopt the service, since the coefficient of *"age"* for the *"focal"* is negative. This finding is in line with the results of Katona et al. (2011), who showed that younger individuals are more influential. However, too young individuals (i.e., less than 18 years old or students, who are usually in a lower age group) are also not likely to adopt the service, as both the coefficients of *"young"* and *"student"* are negative. The strong negative relationship between the response variable and the "*young*" predictor can be associated with the fact that younger people are not financially independent and their decision making depends on their parents. Another reason for this can be related to the theoretical concept by Mourali and Yang (2013), "resistance to social influence", where people might receive influence as a threat for their freedom and they are more resistant to the adoption.

Moreover, *"focals"* classified into a higher *"social class"* have a lower influencing power than individuals from the lower social classes. This finding is contradictory to the previous research on this topic; e.g., Weimann (1991) proved that people of a higher social class are more influential and because of their upper financial situation are becoming opinion leaders. One of the potential explanations of this finding is the fact that this particular *"fun tune"* service, which adoption is predicted by the model, may not be a desired service within the highest social class. This argumentation seems to be supported by the fact of a negative impact of *"friends' social class"* on their adoption decision making. Alternatively, the fact that *"social class"* of the *"friend"* has a negative impact on service adoption can be explained by his resistance to be influenced. In particular, as explained by Iyengar et al. (2011), people of a higher social status are potentially opinion leaders, whose decision making depends on their own will.

Additional segmentation groups, defined by the telecom company, are proven to be strong discriminatory factors for the models. In particular, individuals classified into *"segment (1)"* are substantially less likely to adopt the service. On the other hand, the fact of categorizing the person into *"segment (3)"* has a positive effect.

The level of communication of the *"friend"*, described by the *"log(volume of communication)"* variable has a positive impact on her/his service adoption. On the other hand, the *"focals"* with the high level of communication seem to be less influential, as the *"log(volume of communication)"* attribute for *"focal"* has a negative impact on her/his *"friend's"* adoption. Additionally, the number of people with whom both individuals communicate has a negative impact on the service adoption, which is demonstrated by a negative coefficient of the *"social hub"* attribute for both the "*focal"* and the *"friend"*. This observation can be referred to the finding of Katona et al (2011), who showed that individuals with a higher number of social ties have a lower influencing power.

It has been indicated that information related to the dyad of individuals is important for the predication of a *"friend"* adopting the service. However, not all attributes were found to be significant, contrary to the level of communication of the *"friend"* and *"focal"* the information about the volume of communication within the dyad was proven to be not statistically significant. Likewise, the impact of *"homophily"* is not unequivocal. For the model where the response variable is the service adoption 120 days after the *"focal"*, the attribute is not statistically significant. However, in the models where the adoption time threshold is extended to 240 and 480 days, the effect of *"homophily"* is statistically significant and negative. Nevertheless, in comparison to other variables, the overall impact of the attribute is relatively low. This finding is contradictory with the previous research on this topic. For example, Ma et al. (2010) found that *"homophily"* influences positively product adoption, because of the similarity of tastes between the individuals.

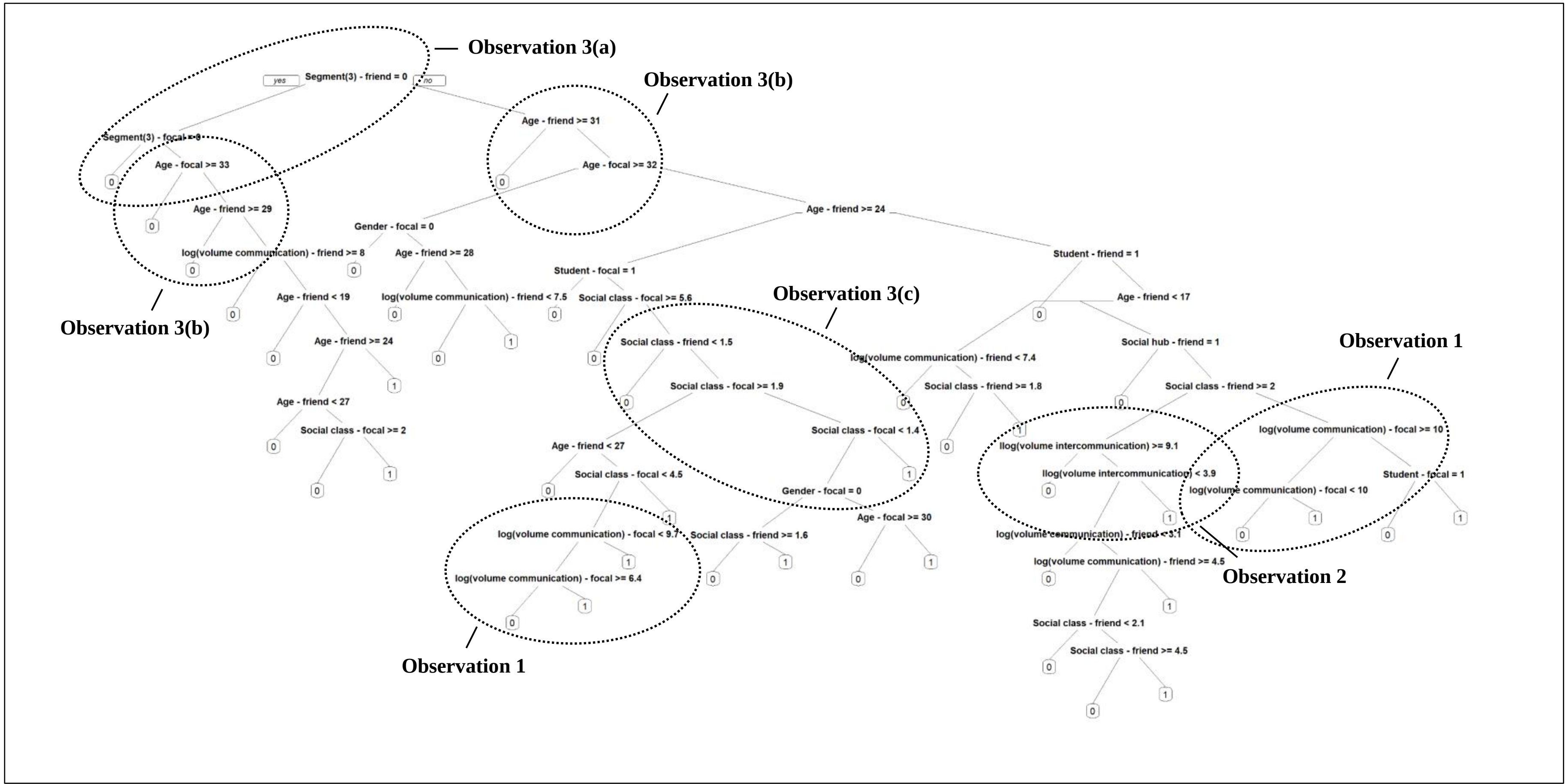
On the other hand, *"Strong tie"* has a statistically significant positive impact on the adoption; however, the effect is not so strong, since the value of its coefficient is relatively low. This finding supplements the study by Nitzan and Libai (2011) of the reasons for service defection. Similarly to this research, they found that the tie strength has a weak, positive impact on decision making.

As a supplementary analysis to the logistic regression models, in order to describe the relationship between the adoption of the service by a *"friend"* and the set of explanatory variables, the classification decision tree methodology is employed. Due to the similarities between the forms of classification trees for service adoption in different time horizons, only results of the model describing service adoption by the *"friend"* within 120 days after the *"focal"* are reported.[19] Figure 3 presents results of the pruned classification decision tree, where the desired size of the tree was selected by pruning back the tree, through minimizing the cross-validation error (refer to Section 4.2) for more details about pruning methods). Discriminatory power of the model, measured with the AUC metric, amounts 0.7205, which is an indicator of a good model performance. The most interesting results of the decision tree are reported in form of

[19] The shortest time horizon was chosen as it corresponds to the period when the degree of *"focal"* influence is the strongest, what is also evidenced by studying the moment in time of *"friend's"* adoption (refer to Section 3).

"Observations", which supplement the analysis of the logistic regression models, by identifying non-linear relationships and cross-effects between variables.

Figure 3: Decision classification tree model – results (adoption within 120 days)



Source: Own analysis based on the data provided by telecom company.

The first *"Observation"* refers to the *"log(volume of communication)"* attribute for the *"focal"*. The results of the logistic regression models showed a negative relation between the volume of communication of the *"focal"* and his influencing power. This observation remains accurate, based on the results of the decision tree model, since *"focals"* who communicate less often have a higher influencing power. However, it is also observed that individuals who communicate exceptionally often, in comparison with other members of their social network, have a high influencing power.

The second *"Observation"* illustrates the importance in the process of decisions making of intercommunication within the dyad. The *"log(volume of intercommunication)"* variable was proven to have a non-linear impact on service adoption by the *"friend"*. In particular, if the communication within a dyad is frequent, but not excessive, the *"friend"* is more likely to adopt the service.

The third *"Observation"* relates to the impact of homophily on service adoption. It is evidenced that the cross-effects for *"friend"* and *"focal"* of socio-demographic segmentation (refer to: *"Observation 3(a)"*), *"age"* (refer to: *"Observation 3(b)"*) and *"social class"* (refer to: *"Observation 3(c)"*), are important factors for the decisions making. However, in comparison to the *"homophily"* attribute which assigns a score based on how similar two individual are, it is evidenced that it may be more important if e.g., both individuals belong to the youngest age group, or to the lowest social class.

5.2. Defection of the service

In order to further understand influences on consumer behavior in the social network, it is useful to investigate what motivates consumers who have already adopted the service, to defect it. This research employs similar methodology for this investigation, as in the case of describing the service adoption criteria. Research conclusions are based on the results of the logistic regression model and supplemented with the analysis of the classification decision tree model.

Results of the logistic regression model are presented below in Table 6. The AUC value equals from 0.6003 to 0.6101, depending on the model specification, which is remarkably less than for the models predicting the service adoption. Nevertheless, the model performs significantly better than a random model.

Table 6: Logistic regression model – results (service defection)

	Defection (240 days)			Defection (480 days)			Defection (960 days)		
Variable name	**Coef.**	**Z-val.**	**GVIF**	**Coef.**	**Z-val.**	**GVIF**	**Coef.**	**Z-val.**	**GVIF**
Intercept	1.554	7.7***	-	1.772	9.2***	-	2.886	12.8***	-
Var. (focal)	-	-	-	-	-	-	-	-	-
Age	-0.005	-2.2**	1.941	-0.005	-2.3**	1.962	-0.013	-5.1***	1.989
Social Class	-0.013	-0.9	3.901	-0.012	-0.9	3.913	-0.062	-4.1***	3.962
Gender	0.038	1.0	1.035	0.044	1.2	1.037	0.093	2.2**	1.042
Young	-0.482	-4.8***	1.112	-0.403	-4.7***	1.131	-0.579	-6.3***	1.160
Student	0.000	0.0	1.037	0.059	0.6	1.036	-0.024	-0.2	1.035
Segment (1)	-0.364	-2.0**	1.071	-0.393	-2.5**	1.081	-0.448	-2.8***	1.100
Segment (2)	-0.003	0.0	1.322	0.014	0.2	1.312	0.103	1.4	1.278
Segment (3)	-0.050	-0.8	2.682	-0.024	-0.4	2.693	-0.184	-2.7***	2.677
Segment (4)	-0.015	-0.2	1.433	0.032	0.4	1.429	0.010	0.1	1.426
log(vol.com.)	-0.051	-3.3***	1.156	-0.012	-1.3	1.161	-0.078	-4.4***	1.152
Social hub	0.049	0.4	1.039	-0.274	-4.0***	1.029	-0.299	-2.4**	1.043
Social super.	-0.259	-3***	2.729	-0.123	-2.4**	2.677	0.005	0.1	2.792
Var. (friend)	-	-	-	-	-	-	-	-	-
Age	-0.014	-5.4***	1.755	-0.014	-5.7***	1.757	-0.008	-3.0***	1.761
Social Class	-0.070	-4.8***	3.752	-0.065	-4.9***	3.756	-0.033	-2.2**	3.766
Gender	0.090	2.4**	1.016	0.043	1.2	1.016	-0.022	-0.5	1.017
Young	-0.297	-2.6**	1.077	-0.337	-3.3***	1.080	-0.261	-2.4**	1.087
Student	-0.173	-1.6	1.032	-0.168	-1.7*	1.034	-0.082	-0.7	1.035
Segment (1)	-0.397	-2**	1.074	-0.365	-2.2**	1.085	-0.393	-2.3**	1.104
Segment (2)	-0.207	-2.9***	1.375	-0.162	-2.5**	1.380	-0.048	-0.7	1.351
Segment (3)	0.072	1.2	2.775	0.199	3.4***	2.789	0.227	3.3***	2.757
Segment (4)	-0.011	-0.1	1.453	0.060	0.7	1.446	-0.048	-0.5	1.438
log(vol.com.)	-0.123	-8.3***	1.096	-0.044	-3.1***	1.095	-0.016	-0.9	1.099
Social hub	-0.252	-1.8*	1.017	-0.106	-0.9	1.020	-0.176	-1.3	1.025
Social super.	-0.027	-0.3	2.513	-0.028	-0.3	2.558	-0.076	-0.8	2.605
Var. (dyad)	-	-	-	-	-	-	-	-	-
log(vol.inter.)	0.006	0.6	1.847	0.002	0.2	1.843	-0.011	-1	1.855
Homophily	0.006	0.1	1.955	-0.125	-1.5	1.923	0.105	1.1	1.879
Strong tie	0.128	2.4**	1.708	0.088	1.7*	1.706	0.066	1.1	1.716
Diagnostics		-			-			-	
AUC		0.6093			0.6003			0.6101	
AIC		17452.72			18972.71			15103.10	
BIC		17664.07			19184.07			15314.46	

Source: Own analysis based on the data provided by telecommunication company.

The first interesting observation revealed by the models describing the service defection is the difference between models describing service adoption in the short and medium time horizons (i.e., 240 and 480 days) and the model which predicts defection in the long horizon (i.e., 980 days). In particular, while the socio-demographic characteristics of the *"friend"* seem to be more significant in the short and medium time horizons, the importance of the features of the *"focal"* takes the lead when defection takes place in the long time period.

The patterns identified for the models the describing service adoption (refer to Table 5 and the corresponding discussion) and models predicting the service defection are aligned for many of the attributes. The *"focal's age"* has a negative non-linear effect on service defection, since both *"age"* and *"young"* attributes have a negative impact on *"friend's"* defection. This observation additionally does not support the previous finding made for service adoption of Katona et al. (2011) by showing that individuals within the youngest age group show considerably higher influencing power than other social actors. Similarly, *"social class"* has a negative impact on service defection. This observation seems to be the most profound for the attribute of *"friend"* and, as in the case of the service adoption, can be explained by a reluctance to influence of the individuals of the highest social class (Iyengar et al., 2011). In comparison to the models describing service adoption, *"gender"* of both the *"focal"* and the *"friend"* has a lower relevance regarding the prediction of defection. However, similarly to the case of adoption, men are proven to be more influential than women, which support the previous research done by Depret and Fiske (1993).

Similarly to the models describing the service adoption, the *"log (volume of communication)"* attribute for the *"focal"* continues to have a negative impact on service defection, however in the medium term (480 days) the attribute does not show a statistical significance. Likewise, the attribute *"log (volume of communication)"* for the *"friend"* has a negative impact on service defection; however, in the long term (960 days) the impact is not statistically significant. This is an interesting finding, since the models which described service adoption, showed a positive relationship between the level of communication and *"friend's"* willingness to adopt the service. This observation can be supported by the argument that individuals, who communicate more often are more "frequent" users of the service.

Similarly to the case of service adoption, the attribute *"social hub"* for the *"focal"* has a negative effect on the defection of the service by the *"friend"*. This finding is in line with the literature (Katona et al., 2011), where it has been proven that the more social ties an individual has, the lower is her/his influencing power over his friend's decisions. The authors explained that individuals with an extended social network have a decreased impact on decisions of each one of their social ties, due to the split of their attention to many individuals.

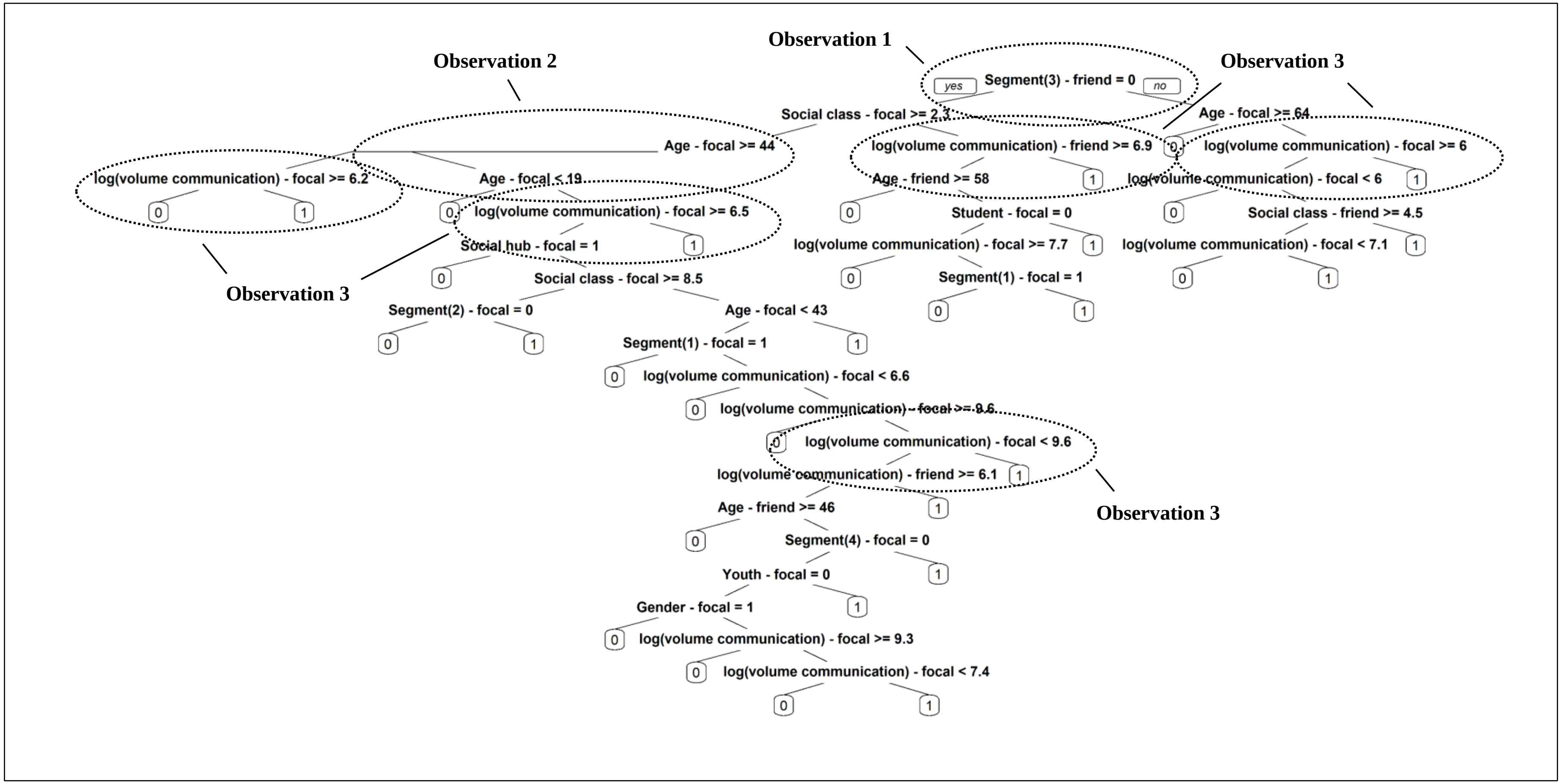
"Focals" characterized by *"social superiority"* have a negative impact on their *"friends'"* decision to defect the service. This finding is somewhat surprising, as the previous literature on social network has demonstrated that older people (Katona et al. 2011) and people of a higher social layer (Strodtbeck et al. 1957) have a strong influence on their social neighbours. One argument that can potentially explain this finding is that both *"age"* and *"social class"* of *"focal"* have a negative impact on defection, and as the variable is created based on these two components, the impact will be aligned with the effect of those attributes.

"Homophily" does not play a significant role in the models predicting *"friend's"* defection, in comparison to the models describing service adoption, where the attribute had a negative, yet small impact on *"friend's"* decision making. This observation is contradictory to the research by Nitzan and Libai (2011), who showed that *"homophily"* is indeed an important factor regarding the service defection. In this research, after controlling for additional variables, such as demographic characteristics of both the *"focal"* and the *"friend"*, the overall impact of *"homophily"* on service defection was proven to be not statistically significant.

Finally, the variable *"strong tie"* has a statistically significant impact on defection. However, the impact of the attribute is only significant at the 5% and 10% level for the models predicting service defection in the short and medium time horizons, respectively. The models show that the *"strong tie"* attribute for the dyad of individuals has a positive impact on *"friend's"* defection. This finding is in line with the results of Nitzan and Libai's (2011) research, who showed that the stronger the tie within a dyad is, the higher is the probability of *"friend's"* defection.

Figure 4 shows results of the classification decision tree, where the response variable is the service defection by the *"friend"* within 480 days after the *"focal's"* defection. The analysis is limited to the prediction of service defection in the medium time horizon, as pruning of the classification tree for the short time horizon (i.e., 240 days) results in the tree with only one level (i.e., with only the root node), considerably limiting the number of possible conclusions; and results of the model for the long time horizon (i.e., 960 days) are comparable to those of the medium horizon. The AUC measure for the chosen model amounts 0.6219 and is slightly higher than that of the logistic regression model. As the pruned classification decision tree based on cross-validation error is too large for reporting purposes (it contains more than 90 nodes), the tree is limited to 31 nodes closest to the root in Figure 4.

Figure 4: Decision classification tree model – results (defection within 480 days)



Source: Own analysis based on the data provided by telecom company.

The first *"Observation"* refers to the *"segment (3)"* variable for the *"friend"*, which is the root of the tree. Similarly to the decision tree model predicting the service adoption, see Figure 3, this attribute is the most important predictor for the classification. The variable has a positive impact on *"friend's"* decision to defect the service; and similarly to the results of the logistic regression model, it shows the importance of socio-demographic segmentation in decision making.

The second *"Observation"* illustrates the non-linearity of the impact of *"focal's"* age on his *"friend's"* willingness to defect the service. the decision tree model shows that *"focals"* at a younger and older age groups (less than 19 and more than 44 years) are significantly less influential, than individuals between those age groups. This observation is aligned with the precious results of the logistic regression model, which showed a negative impact of *"age"* and *"young"* category on the *"friend's"* probability to defect the service.

In the third *"Observation"* it is evidenced that the *"log(volume of communication)"* attribute for both the *"friend"* and *"focal"* has a negative impact on *"friend's"* decision to defect the service. Individuals who are characterized by a high degree of communication are both less influential and less likely to be influenced by others. Finally, similarly to the results of the decision tree model predicting the service adoption, the impact of *"log(volume of communication)"* on defection is non-linear, since *"focals"* who communicate extremely often are considered to be influential.

In comparison to the decision tree model of the adoption, *"log(volume of intercommunication)"* and dyad homophily (either directly through the *"homophily"* attribute, or indirectly through the cross-effect of socio-demographic characteristics of the *"friend"* and the *"focal"*) are not important in the scenario of defection. This is the finding that indicates, that adoption and defection of the service are driven by different motivators.

6. Conclusions, contribution and limitations

Conclusions

This thesis was aimed at investigating of the most important motivators that direct social actors to conform with the decision making of other participants in their social network. This investigation concerned the so- called call-tune service, which allows the adopter's call connections to listen to a chosen ring tune, while awaiting the call receiver's reply. This service is a low-involvement service, which means that the product is of relatively low importance, low reflection before purchase and risk for the consumer.

It was investigated whether the specific attributes that characterize individuals can have an impact on the decision-making process (i.e., service adoption and defection) of other social actors. Twenty-seven such candidate attributes were investigated, among which twelve referred to the profile of the already adopter (*"focal"*), twelve described his social neighbor's (*"friend's"*) profile, and the remaining three characterize their social tie (*"dyad"*). Attributes which describe both the *"focal"* and the *"friend"* were their socio-demographic characteristics (such as age, gender and social class), variables which described the usage of the telecom services and attributes which summarized their individual social network. Regarding the variables describing the "dyads", they described the degree of homophily and the level of communication between two individuals.

Gender, age and social class all played an important role in *"friend's"* decision making. More specifically, the research has found that men, younger individuals and people of lower social classes were more influential than women, elder and people of the highest social class. The finding, which refers to the effect of the gender supports the previous research of Depret and Fiske (1993); however, at the same time the finding is contradictory to the results of Katona et al. (2011), who studied online social networks and found that women were more influential than men. Regarding the effect of age, the research has shown a non-linear relationship between this attribute and the *"focal's"* influential power. In particular, while the analysis has proven a negative relationship between the age and *"focal's"* influential power, very young people (i.e., younger than 18 years) were not found to be influential. Furthermore, "social class" had a negative influence on the adoption and defection, which demonstrated that a lower social class was more influential for the investigated type of service. These findings, which relate to the effects of age and social class are contrast with other contemporary research. For example, according to Katona et al. (2011) and Strodtbeck et al. (1957), older people and those of a higher social layer have a higher influential power on the decisions of their social neighbours.

Another interesting finding of this research refers to the degree of communication between the individuals. The research has shown a general relationship, where the less the *"focal"* communicated with his social ties the more influential he is. However, it has been also revealed that individuals who communicate extremely often, in comparison with the average level of communication of the entire social network, are influential as well. To continue, it was indicated that *"focals"* characterized as the social hubs (individuals with an extensive social network) exhibit low influence on the *"friend's"* decision to adopt or defect the service. This observation can be related to the results of Goldenberg et al. (2009), who confirm that not all[20] social hubs have a strong degree of influence on their *"friends"* and also to the research of Nitzan and Libai (2011), who proved that the higher the number of social ties is, the less influential is the individual on his social neighbor's decisions.

In this research the homophily effect was also investigated. In comparison to the previous research on this topic (Landsman and Nitzan, 2019) where homophily has been shown to have a high impact on service adoption, in this study, based on the results of the logistic regression model the effect of homophily was weak and negative. However, the decision tree model showed that homophily influenced adoption, but the impact could be indirect. The model has revealed that if both the *"focal"* and *"friend"* are characterized by the same socio-demographic characteristics (e.g., they both belong to the lower social class, or to similar age group), then the *"friend"* is more likely to adopt the service.

Additionally, the aspect of social superiority was investigated. It describes a situation when one individual is superior in terms of age or in terms of social class. *"Focals"*

[20] According to the study of Goldenberg et al. (2009) social hubs can be categorized into "innovative hubs" and "follower hubs", which both have influential power over different aspects of diffusion and adoption processes. Refer also to Section 2 on the literature overview.

characterized by social superiority are proven to be not influential regarding their *"friends'"* decision of adopting the service. However, socially superior *"focals"* can influence negatively their social neighbors to defect the service. This finding can be related to the fact that the service is more popular among people of a lower social layer and lower age. The theory of Strodtbeck et al. (1957) that more influential people in higher social layers and the observation of De Bruyn and Lilien (2008) regarding the stronger influence of actors with dissimilar demographic attributes, have not been found in this study.

In this research, it is demonstrated that a close relationship (strong tie) between the *"friend"* and the *"focal"*, was an important factor that influenced positively the decision of service adoption and defection. This observation is well supported in the literature, since according to Granovetter (1973) a strong tie describes individuals who are potentially more willing to discuss openly about their consumer decisions and they communicate with each other during information seeking. In this research, the attribute was identifiable by the higher degree of communication between a dyad of individuals.

Contribution

This research contributes to the literature by combining investigations of social activity in a large-scale network and consumers' adopting and defecting decisions. This research is using large- scale real call data of thousands of social ties between actors. The large amount of data gives the ability to generalize the findings. Like Hill et al. (2006) have mentioned, communication data of the dyads of individuals provide more profound information than sole demographic or geographic characteristics of people.

As Iyengar, Van de Bulte and Valente (2011) emphasized, there is a need to track consumers' adoption of the service, but also the behavior of others in their social network, in terms of adoption. This research helps to better understand this problem, because the analysis is referred to adoption and defection decisions of every actor. More specifically, in this research both communication data and adoption data were combined to understand the effect of "word-of-mouth", social influence and characteristics of an adopter's/ defector's profile that leads social neighbors' decision making. This research can facilitate the understanding of how people adopt and defect a service, how they are being influenced and how companies can forecast potential adopters and prevent their defections. The value of communication data, that there were used in this research, is associated with the fact that they enable insights of the origin of social influence of the social neighbors.

Based on this research, companies are able to identify the profile of potential adopters and defectors, together with influential individuals by revealing the most important characteristics that make an existing adopter or defector more influential in his social network. Equally important is the illustration of influenceable customers, as both are applicable for company's expansion and acquisition of new customers. All this valuable information helps companies to organize promotional plans and target the most profitable and the most influential actors in the social network, as it is more efficient to target customer's social actors than target randomly.

Limitations

A limitation of this research is that the nature of this technology service does not stimulate the consumers for counselling or information seeking from their social network, since it is a low involvement and low risk service. Further research should involve the investigation of social influence on adoption of a high involvement product or a service which acquisition is costly and risky.

A very interesting topic for further research would be an analysis, which would identify if every decision inside the social network was caused by social tie's influence. Even though, this research is dedicated to identifying the characteristics of a consumer and the conforming of his social neighbours with his decision making, there is no certainty about the exact motive.

Another important aspect concerns the limitations of data on which this research is based. This limitation is referred to the lack of information about the actual relationship of two social actors. The categorization of each dyad with labels of family, friend, colleague and acquaintance, would potentially help identifying accurately the motivation behind consumer decision making.

Lastly, another limitation of this research, related to the data availability, is the inability to identify mutual friends among focals. Based on literature (Stadtfeld and Pentland, 2015), social actors who share mutual friends, have a higher probability to be connected and be close in the social network. This information would provide an additional evidence for the company to target people of the close social network.

7. References

Ansari, A., Koenigsberg, O., Stahl, F. (2011). Modeling Multiple Relationships in Social Networks. *Journal of Marketing Research,* 48(4), 713- 728.

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

Arndt, J. (1967). Role of product-related conversations in the diffusion of a new product. Journal of marketing Research, 4(3), 291-295.

Assael, H. (1998). Consumer behaviour and marketing action. 6th ed. Cincinnati (Ohio): South-Western college publishing.

Berger, J., & Heath, C. (2007). Where consumers diverge from others: Identity signaling and product domains. *Journal of Consumer Research*, 34(2), 121-134.

Blau, P. M., & Schwartz, J. E. (2018). Crosscutting social circles: Testing a macrostructural theory of intergroup relations. Routledge.

Boorman, S. A. (1974). Island models for takeover by a social trait facing a frequency-dependent selection barrier in a Mendelian population. *Proceedings of the National Academy of Sciences*, 71(5), 2103-2107.

Bristor, Julia M. (1985), "Buying Networks in Organizational Buying Behaviour," in Proceedings of the Fifteenth Annual Albert Haring Symposium. *Bloomington: Graduate School of Business, Indiana University*, 212- 33.

Brown, J., Reingen, P. H. (1987). Social Ties and Word-of- Mouth Referral Behavior. *Journal of Consumer Research*, 14(3), 350-362.

Centola, D. (2015). The Social Origins of Networks and Diffusion. *American Journal of Sociology*, 120 (5), 1295-1338.

Centola, D., & Macy, M. (2007). Complex contagions and the weakness of long ties. *American journal of Sociology*, 113(3), 702-734.

Centola, D., & van de Rijt, A. (2015). Choosing your network: Social preferences in an online health community. *Social science & medicine*, 125, 19-31.

De Bruyn, A., & Lilien, G. L. (2008). A multi-stage model of word-of-mouth influence through viral marketing. *International journal of research in marketing*, 25(3), 151-163.

Depret, E., & Fiske, S. T. (1993). Social cognition and power: Some cognitive consequences of social structure as a source of control deprivation. In *Control motivation and social cognition* (pp. 176-202). Springer, New York, NY.

Eagle, N., Pentland, A. S., & Lazer, D. (2009). Inferring friendship network structure by using mobile phone data. Proceedings of the national academy of sciences, 106(36), 15274-15278.

Fischer C. S. (1977). Networks and Places: Social Relations in the Urban Setting. New York: Free Press.

Fox, J., & Monette, G. (1992). Generalized collinearity diagnostics. *Journal of the American Statistical Association*, 87(417), 178-183.

Frenzen, J., Nakamoto, K. (1993). Structure, Cooperation, and the Flow of Market Information. *Journal of Consumer Research,* Vol.20, No. 3, pp. 360-375.

Friedkin, N. E. (2006). A structural theory of social influence (Vol. 13). Cambridge University Press.

Goldenberg, J., Shavitt, Y., Shir, E., & Solomon, S. (2005). Distributive immunization of networks against viruses using the 'honey-pot' architecture. *Nature Physics*, *1*(3), 184-188.

Goldenberg, J., Han, S., Lehnmann, D. R., Hong, J. W. (2009). The Role of Hubs in the Adoption Process. *Journal of Marketing,* 73 (2), 1-13.

Granovetter, M. (1985). Economic action and social structure: The problem of embeddedness. American journal of sociology, 91(3), 481-510.

Granovetter, M. S. (1973). The Strength of Weak Ties. *American Journal of Sociology,* 78 (6), 1360-1380.

Haenlein, M., Libai, B. (2013). Targeting Revenue Leaders for a New Product. *Journal of Marketing,* 77 (3), 65- 80.

Hansen, M. T. (1999). The search-transfer problem: The role of weak ties in sharing knowledge across organization subunits. *Administrative science quarterly*, *44*(1), 82-111.

Hartmann, W., Manchanda, P., Nair, H., Bothner, M., Dodds, P., Godes, D., Hosenagar, K., Tucker, C. (2008). Modeling social interactions: Identification, empirical methods and policy implications. *Market Lett,* 19, 287-304.

Hill, S., Provost, F., Volinsky, C. (2006). Network-Based Marketing: Identifying Likely Adopters via Consumer Networks. *Statistical Science,* 21 (2), 256- 276.

Iacobucci, D. Salter, J. M., (2009). Social Networks: You've Lost Control. *The Network Challenge.*

Iyengar, R., Van den Bulte, C., Valente, T. W. (2011). Opinion Leadership and Social Contagion in New Product Diffusion. *Market Science,* 30 (2), 195- 212.

James, G., Witten, D., Hastie, T., Tibshirani, R., Hastie, M. T., & Suggests, M. A. S. S. (2017). Package 'ISLR'.

Katona, Z., Pal Zubcsek, P., Sarvary, M. (2011). Network Effects and Personal Influences: The Diffusion of an Online Social Network. *Journal of Marketing Research,* 48 (3), 425-443.

Katz, E., Lazarsfeld, P. F., & Roper, E. (2017). Personal influence: The part played by people in the flow of mass communications. Routledge.

Kleinbaum David, G., Kupper Lawrence, L., & Muller Keith, E. (1988). Applied regression analysis and other multivariate methods.

Kleinberg, J. (2013). Analysis of large- scale social and information networks. *Philosophical Transactions: Mathematical, Physical and Engineering Sciences,* 371 (1987), 1-7.

Landsman, V., & Nitzan, I. (2019). Cross-decision social effects in product adoption and defection decisions. *International Journal of Research in Marketing.*

Lin, N. (1971). Information flow, influence flow and the decision-making process. Journalism Quarterly, 48(1), 33-61.

Ma, L., Krishnan, R., & Montgomery, A. L. (2015). Latent homophily or social influence? An empirical analysis of purchase within a social network. *Management Science*, *61*(2), 454-473. Muller, E., Peres, R. (2019). The Effect of Social Networks Structure on Innovation Performance: A Review and Directions for Research. *International Journal of Research in Marketing* 36, 3-19

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

Merton, R. K. (1949). Patterns of Influence: A Study of Interpersonal Influence and of Communications Behavior in a Local Community. S. 180-219 in: Paul F Lazarsfeld, und Frank N. Stanton (Hg.): Communications Research 1948-1949.

Murray, K. B. (1991). A test of services marketing theory: consumer information acquisition activities. Journal of marketing, 55(1), 10-25.

Mourali, M., Yang, Z. (2013). The Dual Role of Power in Resisting Social Influence. *Journal of Consumer Research*, 40 (3), 539-554.

Myers, J. H., & Robertson, T. S. (1972). Dimensions of opinion leadership. Journal of marketing research, 9(1), 41-46.

Narayan, V., Rao, V. R., & Saunders, C. (2011). How peer influence affects attribute preferences: A Bayesian updating mechanism. *Marketing Science*, *30*(2), 368-384.

Nitzan, I., & Libai, B. (2011). Social effects on customer retention. *Journal of Marketing*, *75*(6), 24-38.

Onnela, J. P., Saramäki, J., Hyvönen, J., Szabó, G., Lazer, D., Kaski, K., ... & Barabási, A. L. (2007). Structure and tie strengths in mobile communication networks. *Proceedings of the national academy of sciences*, *104*(18), 7332-7336.

Peres, R., Muller, E., & Mahajan, V. (2010). Innovation diffusion and new product growth models: A critical review and research directions. *International journal of research in marketing*, 27(2), 91-106.

Precker, J. A. (1952). Similarity of valuings as a factor in selection of peers and near-authority figures. *The Journal of Abnormal and Social Psychology*, 47(2S), 406.

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

Rogers, E. M. (2003). Diffusion of innovations. NY: Simon and Schuster, 576.

Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, *27*(2), 130.

Stadtfeld, C., & Pentland, A. (2015). Partnership ties shape friendship networks: A dynamic social network study. *Social Forces*, *94*(1), 453-477.

Strodtbeck, F. L., James, R. M., & Hawkins, C. (1957). Social status in jury deliberations. *American Sociological Review*, *22*(6), 713-719.

Tan, P. N., Steinbach, M., & Kumar, V. (2013). Data mining cluster analysis: basic concepts and algorithms. *Introduction to data mining*, 487-533.

Valente, T. W., & Davis, R. L. (1999). Accelerating the diffusion of innovations using opinion leaders. The Annals of the American Academy of Political and Social Science, 566(1), 55-67.

Van den Bulte, C., & Joshi, Y. V. (2007). New product diffusion with influentials and imitators. *Marketing science*, *26*(3), 400-421.

Van den Bulte, C., & Wuyts, S. (2007). *Social networks and marketing*. Cambridge, MA: Marketing Science Institute.

Vaughn, B. K., & Wang, Q. (2008). Classification based on tree-structured allocation rules. *The Journal of Experimental Education*, 76(3), 315-340.

Wangenheim, F. V., & Bayón, T. (2004). The effect of word of mouth on services switching. *European Journal of Marketing*.

Weimann, G. (1994). The influentials: People who influence people. SUNY Press.

Weimann, G. (1991). The influentials: back to the concept of opinion leaders? . Public Opinion Quarterly, 55(2), 267-279.

Wilcox, K., & Stephen, A. T. (2013). Are close friends the enemy? Online social networks, self-esteem, and self-control. *Journal of Consumer research*, *40*(1), 90-103.

Wuyts, S. (2010). Connectivity, control, and constraint in business markets. *The connected customer: The changing nature of consumer and business markets*, 77-103.

8. List of tables and figures

8.1. List of tables

8.2. List of figures