ARTICLE IN PRESS

Journal of Business Research xxx (xxxx) xxx–xxx

Customer segmentation in e-commerce: Applications to the cashback business model

María Teresa Ballestar*, Pilar Grau-Carles, Jorge Sainz

Universidad Rey Juan Carlos, Spain https://doi.org/10.1016/j.jbusres.2017.11.047

A R T I C L E I N F O

Keywords:
Cashback
Social network
E-commerce
Customer behavior
Loyalty
Affiliate marketing

A B S T R A C T

This paper presents a segmentation of cashback website customers. The segmentation is based on customers' commercial activity and role within the site's social network. In this social network, customers profit from the transactions they make on affiliate websites. Mixing traditional marketing strategies with word-of-mouth recommendations is crucial for the success of this business model because these recommendations boost new customer acquisitions and strengthen the loyalty of existing customers. This study shows how the customer's role within the cashback website's social network determines the customer's behavior and commercial activity on the website. The segmentation presented describes the customer journey in terms of customer profitability and seniority. The findings explain customer behavior in e-commerce and the value of applying personalized retention strategies to each cluster rather than generic strategies or customer acquisition strategies. This paper describes how customers move between clusters, enabling practitioners to increase customer loyalty and long-term profitability.

1. Introduction

With millions of competing commercial websites, attracting and retaining high-quality customers is critical for success. To gain an edge over competitors, these websites adopt different marketing methods, many of which are based on updated, Internet-oriented versions of traditional marketing strategies (Libai, Biyalogorsky, & Gerstner, 2003; Rapp, Trainor, & Agnihotri, 2010). Hoffman and Novak (2000) defines working out what customers' need and how to satisfy them as a "work in process." New theoretical and empirical applications continually enrich the literature. Zhang, Hu, Guo, and Liu (2017) and Hair, Hult, Ringle, Sarstedt, and Thiele (2017) focused on applying traditional methods to e-commerce sites and extending these methods. In studies of cashback sites, such applications are yielding promising results.

Cashback websites are based on a specific type of affiliate marketing, which is "…a web-based marketing practice in which a business rewards one or more affiliates for each visitor or customer brought about by the affiliate's marketing efforts" (Ryan, 2016, p. 206). In the case of cashback websites, affiliate websites attract customers in exchange for a reward per visit, transaction, or use of a third party's website. The largest cashback website in the US, Ebates, hosts more than 2000 shopping sites and has more than 2.5 million customers, who spent more than 2.3 billion USD in 2013. TopCashback, the largest UK cashback store, lets customers interact with more than 4000 stores, including some of the country's principal stores.

As with all other aspects of Internet marketing, activities related to community building and constructing social networks are crucial for success. A small core of engaged individuals can increase the returns of any strategy, especially in affiliate marketing, where trust in recommendations is a key issue (Weber, 2009). As Ashley and Tuten (2015), Stephen and Galak (2012), and Naylor, Lamberton, and West (2012) have shown, social media is important for sales, as is the ability to attract new customers and keep them engaged with the brand. Analyzing these interactions is a growing research area because of its academic and practical relevance.

Academic analysis of cashback websites began only recently. The earliest relevant studies are those of Ballestar, Grau-Carles, and Sainz (2016) and Ballestar, Sainz, and Torrent-Sellens (2016), who focused on how size and make-up of a user's network relate to that user's online behavior, brand loyalty, and profitability. Vana, Lambrecht, and Bertini (2017) empirically studied the effect of cashback offers on customers, showing that financial rewards induce customers to purchase and then make larger acquisitions. Ho, Ho, and Tan (2017) developed a game theory model that analyzes pricing strategies of cashback firms depending on their brand value, providing insights for practitioners.

This study enriches the literature by showing how different customer roles within social networks determine customers' behavior and commercial activity on the cashback website. The empirical analysis

* Corresponding author.
E-mail address: mt.ballestar@alumnos.urjc.es (M.T. Ballestar).

https://doi.org/10.1016/j.jbusres.2017.11.047
Received 18 June 2017; Received in revised form 27 November 2017; Accepted 29 November 2017
0148-2963/

reveals eight clusters of customers who differ in terms of their profitability and seniority. The method helps practitioners understand customers' e-commerce behavior. The study shows the value of applying personalized digital marketing strategies to retain customers.

2. Theoretical background

Cashback websites encompass a wide range of merchants that offer rewards to customers for transactions or visits. These sites thereby facilitate the purchase process and provide financial rewards to customers (Ballestar et al., 2016). Understanding digital marketing: marketing strategies for engaging the digital generation. The success of this business model lies not only in the structure and marketing strategy, but also in the mix of customer-centric strategies, including word-of-mouth as a tool for developing customer's social networks through recommendations attached to financial incentives.

Ballestar et al. (2016) confirmed the success of this strategy mix by showing that customers' revenues grow as customers become more engaged with the cashback website and their social network of referees grows. Customers behave significantly differently depending on the nature of their social network because customers enrolled in the cashback website make more transactions and are more multi-transactional. The marketing strategy of instant financial rewards for social network activity offers an incentive for customers to join and develop their social networks (Swamynathan, Wilson, Boe, Almeroth, & Zhao, 2008).

Not all enrolled customers are equally engaged. Their role within the social network is the key to understanding their purchase behavior and commercial activity on the website, as well as their loyalty. Social network analysis offers insight into relationships among users and the complex structure of business organizations and markets. Social network analysis enables analysis and quantification of these interactions (Monaghan, Lavelle, & Gunnigle, 2017). The advantage of social network analysis is that it uses direct data on behavior, which avoids problems of surveys and thereby increases accuracy (Conway, 2014). In this framework, the research hypothesis is as follows:

H1. The purchase behavior of customers depends on their role within the social network.

As an application of word-of-mouth to e-commerce, online communities offer an effective way for social interactions and information sharing to be translated into purchasing tools (Dennison, Bourdage-Braun, & Chetuparambil, 2009). Sharing knowledge allows for more informed decisions, shifting the power from retailers to customers because a more collaborative and social approach allows for better-informed marketplaces (Stephen & Toubia, 2009). Zhang, Lu, Gupta, and Zhao (2014) developed a research model according to which customers' interest in participating in social networks is given by how they interact or socialize within the network.

H2. Customers who are more involved in the social network make more transactions that require payouts. These customers are more profitable to the brand because they generate a greater volume of cashback.

Electronic commerce has undergone a fundamental change. The rise of social commerce has allowed online customers to share experiences, share advice, and examine and buy goods and services (Mardsen, 2010). Pavlou, Liang, and Xue (2007) showed that uncertainty in online markets has decreased because of aspects such as trust and social presence, which encourage transactions. Qiu and Rao (2017) showed that involvement and profitability are related to social applications and cashback services.

H3. Customers who are more involved in the social network are more multi-transactional. These customers make more kinds of transactions, indicating that they are more engaged and thus more loyal to the brand.

Aral and Walker (2011) showed that the characteristics of online social networks tend to remain constant over time, with strong peer effects and homophily. Bapna and Umyarov (2015) found that peer effects diminish with the number of connections, concluding that in social networks, greater involvement means a greater effect on co-users. We extended this empirical analysis and drew upon Aral's (2011) study by examining the participation of highly involved users in the social network.

3. Data collection

We used data from one of the largest cashback websites in continental Europe. This website operates in 14 countries, has more than two million customers, and has annual sales of more than 20 million Euros. Observations were gathered from January to March 2015. The data collection procedure enabled simplification without losing representativeness. Data on all transactions by customers and their roles within the social network were collected. To avoid sample bias, new customers who enrolled in the website during this period were excluded because these customers had insufficient time to develop their social networks.

Customers can join the cashback website social network proactively or as a referee following existing customers' recommendations in a word-of-mouth marketing strategy. The cashback policy stipulates that customers receive direct cashback for all transactions they make and network cashback for click and visit transactions by users in their network of referees up to second level. The level in the social network reflects the degree of the relationship between referrer and referee. First-level referees joined following recommendation from the referrer. Second-level referees joined following recommendation from users in the first level. The average size of the network for customers in the sample was 32.8 referees per customer (13 in the first level and 19.8 in the second).

The 12,548 customers in the sample had each made at least one transaction. The sampled customers made 687,682 transactions altogether, receiving 90,876 Euros in direct cashback and 4868 Euros in network cashback from referees' transactions. Men accounted for 59.55% of the sample (mean age 35.3 years). Women accounted for 40.44% of the sample (mean age 36.4 years).

The data were aggregated at the customer level in a single table with 12,548 records, and 60 variables were calculated to measure customers' commercial activity and role within the social network. Of these 60 variables, 8 were relevant for the empirical analysis. Data mining was performed in MySQL and SAS. Descriptive analysis is detailed in Subsections 3.1 and 3.2.

3.1. Customer transactionality

Customers made an average of 55 transactions during the period of observation. These transactions can be divided into five types. For each customer, the number of each type of transaction was captured in a numerical variable. The five types of transactions are as follows:

3.1.1. One-click or visit transactions

These transactions do not require payouts. This type of transaction consists of activities such as watching videos, becoming a fan in social networks, and so forth. In the sample, 80.90% of customers had made at least one transaction of this kind. These transactions represented 97.76% of total transactions and generated 13.23% of total direct cashback (0.02 Euros average cashback per transaction).

3.1.2. Registration

Registration occurs when a customer sets up a user account with an affiliate merchant. This transaction does not require payouts. In the sample, 6.69% of customers had made at least one transaction of this kind, which represented 0.15% of all transactions and generated 13.78% of total direct cashback (12.48 Euros average cashback per transaction).

3.1.3. Purchasing products and services

This type of transaction consists of making purchases from an affiliate merchant. These transactions require payouts. In the sample, 27.14% of customers had made at least one transaction of this kind, which represented 0.98% of all transactions and generated 70.21% of total direct cashback (9.44 Euros average cashback per transaction).

3.1.4. Qualified lead transactions

Becoming a qualified lead means that customers show an intention to purchase a product or service by completing the purchase process with an affiliate merchant. This type of transaction consists of providing information about a loan, mobile phone company, or similar. The customer receives cashback once the service is activated. In the sample, 20.70% of customers had made at least one transaction of this kind, which represented 1.50% of all transactions and generated 2.43% of total direct cashback (0.21 Euros average cashback per transaction).

3.1.5. Manually processed transactions

When the transaction is not automatically processed by the system for any reason, it is processed manually. This type of transaction accounted for 0.005% of the transactions and generated 0.35% of total direct cashback (10.11 Euros average cashback per transaction).

3.2. Customer characteristics

The customer characteristic variables that were pertinent to our research were the customer's role within the social network, customer seniority, and customer profitability.

3.2.1. The customer's role within the social network

Two factors determine the customer's role within the social network. The first refers to how the customer joins the social network cashback website (i.e., whether the customer has a referrer). This factor cannot change over time. The second refers to whether the customer develops his or her social network. Customers only receive cashback from their network of referees up to the second level (i.e., from direct referees and from referees' referees). There is no limit to the number of members at these two levels. Accordingly, the role of each customer within the sample was captured by a categorical variable with the following four categories:

Lonely users (13% of the sample) have no social network. Transactions by these customers generate cashback only for themselves. *Users without a referrer but with a network of referees* (11.9% of the sample) join the cashback website without the recommendation of a referrer and develop their own network. There are two types of these users: customers with referees up to first level (7.7% of the sample) and customers with referees up to second level (4.2% of the sample). Both groups benefit from cashback from their own commercial activity and referees' transactions. *Users with a referrer but without a network of referees* (36% of the sample) join the cashback website following recommendation from a referrer but do not develop their own network. They generate cashback for themselves and their referrers only. *Users with a referrer and a network of referees* (39.1% of the sample) join the cashback website following recommendation from a referrer and develop a network of referees. There are two types of these users: customers with referees up to first level (24.8% of the sample) and customers with referees up to second level (14.2% of the sample). These groups generate cashback for themselves and their referrers while benefiting from referees' transactions.

3.2.2. The customer's seniority

Customer seniority refers to the number of years since the customer joined the cashback website. The mean number of years as member was 3.1 years. This information was stored in a numerical variable.

3.2.3. The customer's profitability

Customer profitability refers to the total cashback generated by the user. The mean was 7.63 Euros, of which 5.4% came from the network of referees. This information was stored in a numerical variable.

4. Empirical analysis and results

Two-step cluster analysis in SPSS 24 was used to group customers based on their commercial activity and role within the social network. The main advantages of this method are the ability to manage both categorical and continuous variables, automatic selection of the number of clusters, and the ability to analyze large datasets. Two-step cluster analysis yields the best results when all variables are independent, categorical variables follow a multinomial distribution, and continuous variables follow a normal distribution. Nevertheless, empirical internal testing indicates that the procedure is robust to violations of both assumptions. Hence, "because cluster analysis does not involve hypothesis testing and calculation of observed significance levels, other than for descriptive follow-up, it's perfectly acceptable to cluster data that may not meet the assumptions for best performance" (Norušis, 2011).

Our data met the criteria described by Norušis (2014). We had one categorical variable (customer's role) and five continuous variables. The number of transactions on the cashback website by type was aggregated at the customer level, yielding 12,548 records (Table 1).

The first step of two-step cluster analysis consisted of pre-clustering the raw data using the log-likelihood distance as the similarity criterion. Standardized data records were merged in a sequential process using an

Table 1
List of variables included in the two-step cluster analysis.

Variables	Description
Role_in_social_network	Categorical variable capturing the customer's role within the social network of the cashback website: role 1: lonely user role 2: user without referrer but with a network of referees (up to second level) role 3: user without referrer but with a network of referees (up to first level) role 4: user with referrer but without a network of referees role 5: user with referrer and with a network of referees (up to second level) role 6: user with referrer and with a network of referees (up to first level)
Click_direct_transactions	Numerical variable that captures the number of one-click interactions or visit transactions the customer makes on the platform. Payout by the customer is not required.
Manual_direct_transactions	Numerical variable that captures the number of manually processed transactions the customer makes.
Registration_direct_transactions	Numerical variable that captures the number of registration transactions the customer makes on the platform. Payout by the customer is not required.
Lead_direct_transactions	Numerical variable that captures the number of transactions where the customer becomes a qualified lead. These transactions do not require instant payout, but later payout is required to activate the cashback process.
Purchase_direct_transactions	Numerical variable that captures the number of times a customer purchases a product or service on the platform. Payout by the customer is required.

existing pre-cluster or a new pre-cluster that led to the largest log-likelihood. In the second step, the pre-clusters were combined using agglomerative hierarchical clustering under the Schwarz criterion (BIC), yielding eight clusters.

The consistency of the clustering structure was evaluated using silhouette validation (Rousseeuw, 1987), which measures cohesion between elements within a cluster and separation between clusters. The silhouette coefficient ranges from − 1 to 1, where − 1 means that the model is poor and 1 means that the model is optimal. Values greater than 0.5 indicate good model quality (Kaufman & Rousseeuw, 1990). Here, the silhouette coefficient was 0.6, so the model was robust.

The importance predictor measures the relevance rather than the accuracy of the model. All variables had an importance of 1, except the number of purchase transactions per customer, which had an importance of 0.17. This variable nonetheless remained in the model. Number of purchase transactions per customer was of strategic value to explain the cashback website business model and customer journey (Voorhees et al., 2017) because this variable accounted for a high proportion of total direct cashback (70.21%) and did not impair the model's performance.

4.1. Clustering structure of cashback website users

Customers were grouped into eight clusters based on their commercial activity and role within the cashback website social network (Fig. 1). The percentage of the sample in each cluster was as follows: Cluster 1 (29.7%), Cluster 2 (9.9%), Cluster 3 (0.6%), Cluster 4 (6.5%), Cluster 5 (11.4%), Cluster 6 (19.5%), Cluster 7 (10.6%), and Cluster 8 (11.8%). The smallest cluster (Cluster 3) had 72 customers, and the largest cluster (Cluster 1) had 3722 customers. Cluster profiles appear in Table 2. The centroids for the continuous variables appear in the top part, and the frequencies for the categorical variables appear in the bottom part.

Once customers had been sorted into clusters, the clusters were characterized in terms of customers' profitability and seniority, providing additional information about the segmentation. Profitability represents average cashback (direct plus network) per customer, and seniority represents the average number of years since the customer enrolled.

Two one-way ANOVA tests were conducted to confirm significant differences between clusters in terms of seniority and profitability means. The goal of the ANOVA testing was to confirm the suitability of these variables for characterizing the clusters (Qiu & Rao, 2017). The ANOVA tests revealed significant differences among clusters in terms of average seniority and average profitability. For both variables, the p-value was 0.000. These findings support H1 by showing that the purchase behavior of customers depends on their role within the cashback website social network.

5. Discussion and implications

The eight clusters and the customer journey, as defined by Voorhees et al. (2017), are now described.

5.1. Clusters of convenience buyers

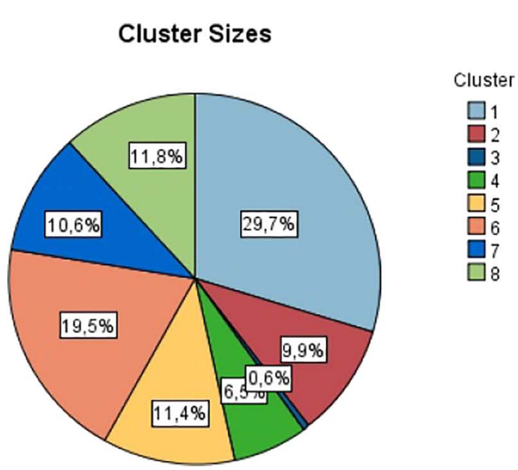
Customers from Clusters 5 and 7 are convenience buyers; they do not have a referrer, and they enrolled on the cashback website to benefit financially. They are proactive users who may start in Cluster 5 and move to Cluster 7 as they become more engaged with the social network and make more transactions, although their main driver remains convenience. These clusters account for 26.9% of total purchase transactions.

5.1.1. Cluster 5: convenience buyers with potential

This cluster comprises lonely users (1435 customers; 11.4% of the sample) (Fig. 1). These users are the newest customers (average seniority 2.16 years) (Fig. 2). Customers in this cluster are the least engaged with the social network and transactions. These customers have not developed relationships. They have the fewest direct transactions, with 20.0 direct transactions per customer (63.6% lower than the overall mean) and an average cashback per customer of 6.34 Euros (Fig. 2). These customers make 13.8% of all purchase transactions. The number of transactions they make is 21% higher than the overall mean. They make 4.1% of one-click interactions or visits and 3.1% of qualified lead transactions (Fig. 3). These customers' behavior is driven by the convenience of an instant discount when buying a product or service and is not linked to impulsive purchases or intensive use. This cluster is therefore the third least profitable cluster in terms of profitability per customer.

5.1.2. Cluster 7: engaged convenience buyers

This cluster (1328 customers) is an evolved form of Cluster 5. Customers in Cluster 7 are more engaged than those in Cluster 5. They interact more with the network and make more transactions. This cluster constitutes 10.6% of the sample. Average seniority is 3.36 years (Fig. 1). These customers have developed a network of referees (66% have first-level referees, and 34% have second-level referees). These customers make more transactions than those in Cluster 5. The average number of direct transactions per customer is 44.4, 122.1% higher than for customers in Cluster 5 yet still 19% below the overall mean. Average cashback per customer is 8.29 Euros (Fig. 2). These customers are 23% above the overall mean in terms of number of purchase transactions. They make 13.1% of all purchase transactions (Fig. 3). This cluster comprises engaged convenience buyers who like instant financial rewards when buying products. They are 31% more profitable, make more diverse transactions, and have better networks than customers in



Size of Smallest Cluster	72 (0,6%)
Size of Largest Cluster	3722 (29,7%)
Ratio of Sizes: Largest Cluster to Smallest Cluster	51,69

Fig. 1. Cluster distribution chart.

Table 2
Cluster profiles: centroids of continuous variables and frequencies of categorical variables.

	Centroids									
	Click_direct_transactions		Manual_direct_transactions		Registration_direct_transactions		Lead_direct_transactions		Purchase_direct_transactions	
Cluster	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
1	20.26	35.65	0.00	0.00	0.00	0.00	0.25	0.77	0.39	1.07
2	273.67	74.63	0.00	0.00	0.00	0.06	4.52	4.66	0.20	0.78
3	154.60	170.95	0.43	0.50	0.57	1.27	15.22	25.52	11.54	76.21
4	50.82	90.59	0.00	0.00	1.17	0.48	0.78	1.77	0.59	1.19
5	19.10	42.77	0.00	0.00	0.00	0.00	0.23	0.84	0.65	1.46
6	24.51	37.18	0.00	0.00	0.00	0.00	0.24	0.75	0.48	1.17
7	43.21	74.75	0.00	0.00	0.00	0.00	0.50	1.37	0.66	1.28
8	38.45	59.03	0.00	0.00	0.00	0.00	0.31	0.95	0.50	1.10
Total	53.36	92.88	0.00	0.05	0.08	0.33	0.82	3.08	0.54	5.91

Cluster	Frequencies											
	Role_in_social_network											
	Role 1		Role 2		Role 3		Role 4		Role 5		Role 6	
	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage	Frequency	Percentage
1	0	0.00%	0	0.00%	0	0.00%	3722	82.29%	0	0.00%	0	0.00%
2	75	4.61%	42	7.88%	4	0.42%	546	12.07%	151	8.46%	423	13.57%
3	12	0.74%	2	0.38%	9	0.93%	17	0.38%	15	0.84%	17	0.55%
4	105	6.45%	38	7.13%	73	7.58%	238	5.26%	134	7.51%	229	7.35%
5	1435	88.20%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%
6	0	0.00%	0	0.00%	0	0.00%	0	0.00%	0	0.00%	2448	78.54%
7	0	0.00%	451	84.62%	877	91.07%	0	0.00%	0	0.00%	0	0.00%
8	0	0.00%	0	0.00%	0	0.00%	0	0.00%	1485	83.19%	0	0.00%
Combined	1627	100%	533	100%	963	100%	4523	100%	1785	100%	3117	100%

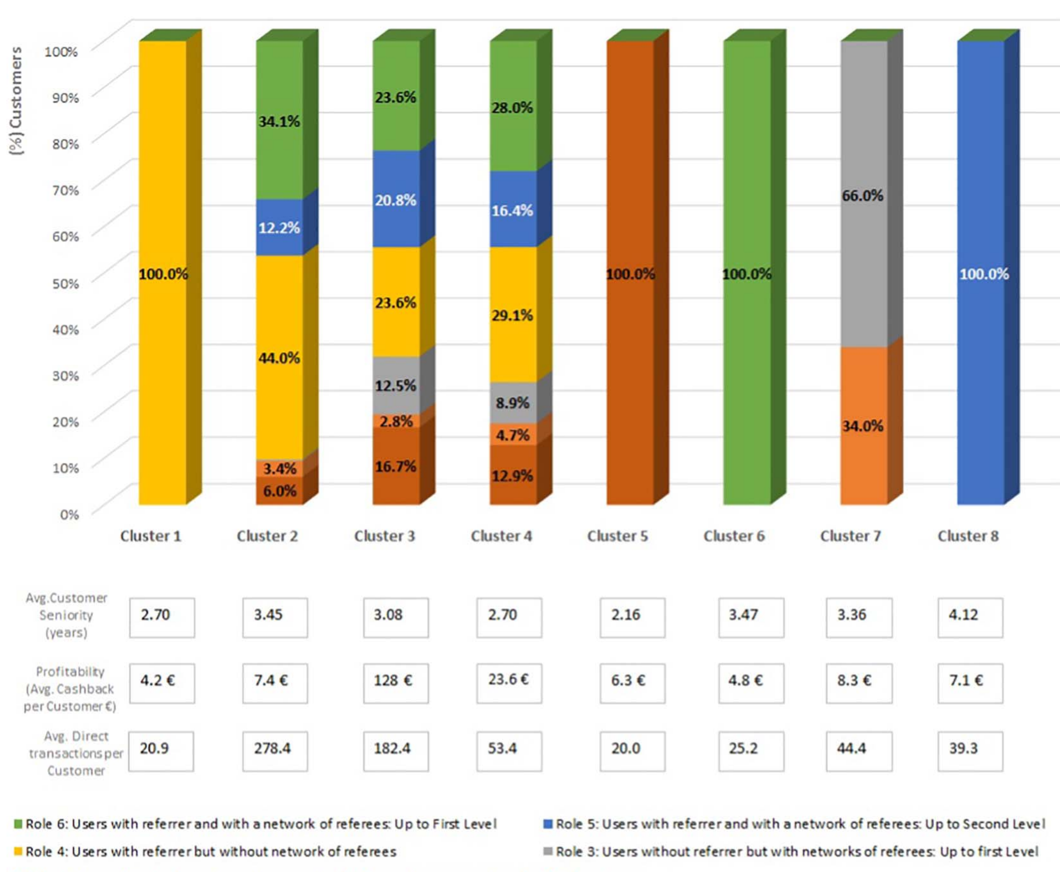


Fig. 2. Cluster characteristics: customer's role within the social network, seniority, profitability, and transactionality.

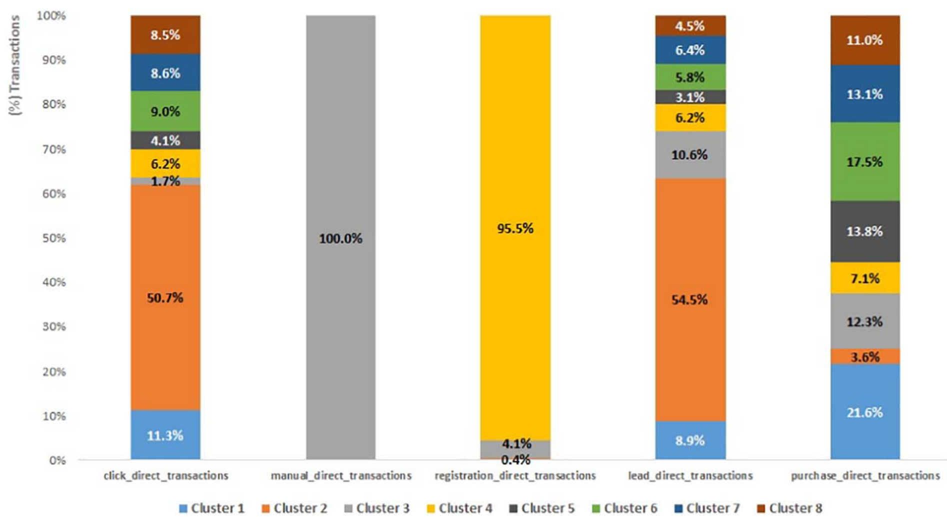


Fig. 3. Cluster transactionality by kind of transaction on the cashback website.

Cluster 5, although they are still below average in this regard.

5.2. *Clusters of referees with medium-low performance*

These customers are light users, but their engagement in terms of social network and transactionality may increase over time as they move from Cluster 1 to Cluster 6 and then to Cluster 8. They joined the cashback website's social network following recommendation by a referrer, and they have medium to low performance in transactions that do not require payouts. They account for 28.8% of one-click interactions or visits, and 19.1% of qualified lead transactions. They also display convenience behavior when they have a specific need. Together, they account for 50.1% of purchase transactions.

5.2.1. *Cluster 1: immature referees*

Cluster 1 is the largest cluster. It comprises users who joined the social network following recommendation from a referrer but who have not yet developed a network of referees (3722 customers; 29.7% of the sample). These customers have average seniority of 2.70 years, the second youngest after Cluster 5. They are poorly engaged, with an average of 20.9 direct transactions per customer. They have the lowest profitability of all clusters, with an average of 4.22 Euros cashback per customer. Purchase transactions are the most appealing type of transactions for these customers. These customers are just 27% below the overall average number of purchase transactions and account for 21.6% of purchase transactions.

5.2.2. *Cluster 6: referees in development*

This cluster consists of referred users who have increased their engagement by building a network of first-level referees (2448 customers; 19.5% of the sample). The development of the customer's network occurs as their seniority increases. These customers have an average seniority of 3.47 years, 12% higher than the overall mean. Customers in Cluster 6 have 25.2 direct transactions per customer, 20.7% higher than customers in Cluster 1. Customers in Cluster 6 have low profitability, however, with an average cashback per customer of 4.81 Euros. Customers in Cluster 6 also perform better than customers in Cluster 1 do, especially in product and service purchasing, where they are only 10% below average. Customers in Cluster 6 make 17.5% of this kind of transaction. Cluster 6 is an evolved form of Cluster 1. In Cluster 6, customers who joined following recommendations become more engaged in terms of social network development and number of transactions.

5.2.3. *Cluster 8: engaged referees*

This cluster consists of users with a referrer and with a network of referees up to second level (1485 customers; 11.8% of the sample). These customers are more engaged with the social network than customers in Cluster 6 are. Their seniority, 4.12 years, is 33% higher than average. This cluster also has higher engagement than Cluster 6 in terms of transactionality. Customers in Cluster 8 make 39.3 direct transactions per customer, 55.6% higher than in Cluster 6. Their profitability is 7.06 Euros cashback per customer. These customers are still below the average in transactions, although they have strong performance, especially in product and services purchasing. In product and services purchasing, they are just 7% below the average, making 11% of this kind of transaction. Cluster 8 is an evolved form of Cluster 6 and Cluster 1. Customers in Cluster 8 have more seniority, have a network of first- and second-level referees, and make more transactions than customers in Clusters 1 and 6, making customers in Cluster 8 the most profitable of the three.

5.3. *Clusters of high performance customers*

Customers from these clusters are high performers who have evolved from Clusters 5 and 1. Some of these customers have evolved from convenience buyers who proactively joined the social network. Others were referred by a referrer.

5.3.1. *Cluster 3: heavy users of all kind of transactions with low sensitivity to payouts (high profitability)*

This is the smallest cluster, with 72 customers (0.6% of the sample). This cluster has the second highest number of transactions per customer and the highest profitability per customer. Customers in Cluster 3 have an average seniority of 3.08 years. Customers in Cluster 3 are heavy users of all kinds of transactions. They are insensitive to payouts. They make an average of 182.4 transactions per customer, 233% above the mean. They are also the most profitable customers, with an average of 127.97 Euros cashback per customer (Fig. 2). They make 12.3% of purchases, 10.6% of qualified lead transactions, 4.1% of registrations, and 100% of manually processed transactions. Their involvement in the social network varies: 16.7% have no kind of relationship, 68.1% have a referrer, and 59.7% have a network of referees. This is the heavy user cluster most prone to active involvement in the social network whereby customers apply word-of-mouth marketing strategies to develop their own networks of referees.

5.3.2. *Cluster 2: heavy user referees who invest time on the website (low profitability)*

This cluster represents 9.9% of the total sample, with 1241 customers. Average seniority is 3.45 years. They are the most active customers in terms of number of transactions, but their profitability is low because of the kind of transactions. The majority have relationships in the social network. Only 6% are lonely users. Of the customers in Cluster 2, 90.2% enrolled following recommendation (evolving from Cluster 1), and 9.8% evolved from Cluster 5. They are less prone to practicing word-of-mouth marketing strategies than customers in the other two heavy user clusters. Only 50% of these customers have developed social networks. These users are interested in non-payout transactions such as one-click interactions or visits and qualified lead transactions. In these transactions, customers in this cluster are 413% and 451% above average, respectively, with 278.4 transactions per customer. These customers make 50.7% of total one-click interaction or visit transactions and 54.5% of total qualified lead transactions. Their profitability in terms of direct Cashback is therefore just 7.43 Euros because these non-payout transactions are the least profitable.

5.3.3. Cluster 4: recommended profitability and convenience-oriented users (medium-high profitability)

This cluster has 817 customers (6.5% of the sample). Average seniority is 2.7 years. These customers have quickly evolved from Cluster 1 and Cluster 5 (73.6% of them have a referrer; 26.4% do not). In addition, 58% of these customers have developed a network of referees. This is the second most profitable cluster, with 23.56 Euros per customer, and the third most active in terms of transactions, with 53.4 transactions per customer. Their consumer behavior is highly efficient. These customers make many high-revenue transactions. In product purchasing, they are 9% above average, making 7.1% of these transactions. In registration, they are 1367% above average, making 95.5% of these transactions.

5.4. Summary of findings

Using the roles defined in Section 3.2.1, Fig. 2 summarizes the cluster characteristics in terms of seniority, average cashback per customer, and transactionality per customer. Fig. 3 summarizes transactionality for each cluster.

The findings support H2 by showing that the route via which customers enroll in the cashback website determines their role and purpose. The way customers enroll significantly affects their behavior and customer journey on the site. Customers make more purchase transactions as they develop their social network and evolve through the clusters over time, becoming more engaged, loyal, and profitable. The exception to this finding is Cluster 2, which comprises customers who enrolled following recommendation and who seek to maximize financial reward without making payouts, even if this means investing considerable time making low-profit transactions.

The way a customer enrolls in the cashback website and behaves on this website offers the strongest evidence in support of H3. Transactionality and diversity of transactions are greater for customers in Clusters 2, 3, and 4 because these customers' common aim is generating financial reward. Nevertheless, Clusters 2, 3, and 4 still have 6% to 16.7% of lonely customers. These customers have evolved faster than others from Clusters 1 and 5, with an average seniority of 3.0 years versus 3.6 years for the medium-low performance and convenience-buyer clusters (Clusters 6, 7, and 8) of customers who have developed their social networks. Nevertheless, comparing these high-performance clusters, whose members pursue their goals by adopting one of three website activity strategies, shows that the number of transactions that these customers make is higher when the social network is more developed.

6. Conclusions

This paper describes the customer journey and the transformation of customers throughout the life cycle of their use of cashback websites. This paper thus addresses a new and promising research area. By applying the social networks literature to marketing, we apply concepts such as loyalty, social networks, and customer evolution and engagement to show that the customer's role depends on the customer's position within the network (Stephen & Toubia, 2009; Zhang et al., 2014). The study also empirically shows that more engaged customers are more transactional, especially in areas where trust is more important (Chen et al., 2014; Pavlou et al., 2007). Finally, we found that in cashback websites, engagement also relates to multi-transactionality, as reported in a more general setting by Bapna and Umyarov (2015).

This analysis has several implications for practitioners, not only in cashback websites, but also in affiliate marketing. The findings show managers how to deal with different customers with different characteristics to strengthen their loyalty and contribution to the brand in a developing area. The findings can yield especially high returns for affiliates in an increasingly competitive environment.

The study has some limitations. The findings show a path among clusters that allows practitioners to design marketing strategies depending on the customer's cluster. Practitioners can thus devise strategies for customer acquisition depending on customers' expected value and evaluate customers' online behavior over time. Future research should further validate this behavior and confirm whether other affiliate and social networks perform in the same way. Also, the use of other Big Data techniques would enable the study of longer periods, which would yield additional insight into this issue.

References

Aral, S. (2011). Commentary—Identifying social influence: A comment on opinion leadership and social contagion in new product diffusion. *Marketing Science, 30*(2), 217–223.

Aral, S., & Walker, D. (2011). Creating social contagion through viral product design: A randomized trial of peer influence in networks. *Management Science, 57*(9), 1623–1639.

Ashley, C., & Tuten, T. (2015). Creative strategies in social media marketing: An exploratory study of branded social content and consumer engagement. *Psychology and Marketing, 32*(1), 15–27.

Ballestar, M. T., Grau-Carles, P., & Sainz, J. (2016). Consumer behavior on cashback websites: Network strategies. *Journal of Business Research, 69*(6), 2101–2107.

Ballestar, M. T., Sainz, J., & Torrent-Sellens, J. (2016). Social networks on cashback websites. *Psychology and Marketing, 33*, 1039–1045.

Bapna, R., & Umyarov, A. (2015). Do your online friends make you pay? A randomized field experiment on peer influence in online social networks. *Management Science, 61*(8), 1902–1920.

Chen, A., Lu, Y., Chau, P. Y., & Gupta, S. (2014). Classifying, measuring, and predicting users' overall active behavior on social networking sites. *Journal of Management Information Systems, 31*(3), 213–253.

Conway, S. (2014). A cautionary note on data inputs and visual outputs in social network analysis. *British Journal of Management, 25*(1), 102–117.

Dennison, G., Bourdage-Braun, S., & Chetuparambil, M. (2009). Social commerce defined. *White paper #23747*. NC, USA: IBM Corporation, Research Triangle Park, Durham.

Hair, J. F., Hult, G. T. M., Ringle, C. M., Sarstedt, M., & Thiele, K. O. (2017). Mirror, mirror on the wall: A comparative evaluation of composite-based structural equation modeling methods. *Journal of the Academy of Marketing Science,* 1–17. http://dx.doi.org/10.1007/s11747-017-0517-x.

Ho, Y.-C., Ho, Y.-J., & Tan, Y. (2017). Online cashback shopping: Implications for consumers and e-businesses. *Information Systems Research, 28*(2), 250–264.

Hoffman, D. L., & Novak, T. P. (2000). How to acquire customers on the web. *Harvard Business Review, 78*(3), 179–188.

Kaufman, L., & Rousseeuw, P. J. (1990). *Finding groups in data: An introduction to cluster analysis.* Hoboken, NJ, USA: John Wiley & Sons.

Libai, B., Biyalogorsky, E., & Gerstner, E. (2003). Setting referral fees in affiliate marketing. *Journal of Service Research, 5*(4), 303–315.

Mardsen, P. (2010). *Social commerce: Monetizing social media.* Hamburg, Germany: Syzygy Deutschland Gmbh.

Monaghan, S., Lavelle, J., & Gunnigle, P. (2017). Mapping networks: Exploring the utility of social network analysis in management research and practice. *Journal of Business Research, 76*, 136–144.

Naylor, R. W., Lamberton, C. P., & West, P. M. (2012). Beyond the "like" button: The impact of mere virtual presence on brand evaluations and purchase intentions in social media settings. *Journal of Marketing, 76*(6), 105–120.

Norušis, M. J. (2011). *SPSS 19.0 statistical procedures companion.* Prentice Hall.

Norušis, M. J. (2014). *SPSS 13.0 statistical procedures companion.* Prentice Hall.

Pavlou, P. A., Liang, H., & Xue, Y. (2007). Understanding and mitigating uncertainty in online environments: A principal-agent perspective. *MIS Quarterly, 31*(1), 105–136.

Qiu, Y., & Rao, R. C. (2017). Increasing retailer loyalty through use of cash back rebate sites (May 9, 2017). Retrieved from SSRNhttps://ssrn.com/abstract=2965870.

Rapp, A., Trainor, K. J., & Agnihotri, R. (2010). Performance implications of customer-linking capabilities: Examining the complementary role of customer orientation and CRM technology. *Journal of Business Research, 63*(11), 1229–1236.

Rousseeuw, P. J. (1987). Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics, 20*, 53–65.

Ryan, D. (2016). *Understanding digital marketing: Marketing strategies for engaging the digital generation.* Kogan Page Publishers.

Stephen, A. T., & Galak, J. (2012). The effects of traditional and social earned media on sales: A study of a microlending marketplace. *Journal of Marketing Research, 49*(5), 624–639.

Stephen, A. T., & Toubia, O. (2009). Deriving value from social commerce networks. *Journal of Marketing Research, 47*(2), 215–228.

Swamynathan, G., Wilson, C., Boe, B., Almeroth, K., & Zhao, B. Y. (2008). Do social networks improve e-commerce? A study on social marketplaces. *Proceedings of the first workshop on online social networks* (pp. 1–6). New York, NY, USA: ACM. (Paper presented at SIGCOMM'08 ACM SIGCOMM 2008 Conference Seattle, WA, USA, August 17–22, 2008) https://doi.org/10.1145/1397735.1397737.

Vana, P., Lambrecht, A., & Bertini, M. (2017). Cashback is cash forward: Delaying a discount to encourage further spending (June 30, 2017). Retrieved from: https://ssrn.com/abstract=2588611 or https://doi.org/10.2139/ssrn.2588611.

Voorhees, C. M., Fombelle, P. W., Gregoire, Y., Bone, S., Gustafsson, A., Sousa, R., & Walkowiak, T. (2017). Service encounters, experiences and the customer journey: Defining the field and a call to expand our lens. *Journal of Business Research, 79*, 269–280.

Weber, L. (2009). *Marketing to the social web: How digital customer communities build your business.* Hoboken, NJ, USA: John Wiley & Sons.

Zhang, M., Hu, M., Guo, L., & Liu, W. (2017). Understanding relationships among customer experience, engagement, and word-of-mouth intention on online brand communities: The perspective of service ecosystem. *Internet Research, 27*(4), 839–857.

Zhang, H., Lu, Y., Gupta, S., & Zhao, L. (2014). What motivates customers to participate in social commerce? The impact of technological environments and virtual customer experiences. *Information Management, 51*(8), 1017–1030.