An agent-based system for modeling users’ acquisition and retention in startup apps

Amir Sayyed-Alikhani a, Manuel Chica b,c, *, Ali Mohammadi a

a Department of Management, School of Economics, Management & Social Sciences, Shiraz University, Shiraz, Iran
b Andalusian Research Institute DaSCI “Data Science and Computational Intelligence”, University of Granada, 18071 Granada, Spain
c School of Electrical Engineering and Computing, The University of Newcastle, Callaghan, NSW 2308, Australia

ARTICLE INFO

Keywords:
Startups
Customer acquisition
Customer retention
Agent-based modeling
Simulation
Retention modeling

ABSTRACT

Startup companies boost the quality of everyday life in almost all dimensions, and their products and services are of relevance everywhere. One of the most important goals that startups pursue is to increase the number of their users quickly. Users are of two types, new and returning. The present study presents an agent-based model to simultaneously deal with these two types of users by placing them in a preferential attachment network to interact. In this model, new users can be added at each time step according to word-of-mouth (WOM) and marketing activities. To define the retention probability for an agent, a set of real users in the records with the same properties as the agent are looked for to check what they have done in the same situation. To validate and test the model, the agent-based system is first thoroughly verified and then applied to the real data of a startup in the game app industry. After the experiments on a real scenario, the best decisions are made about the users to focus on or incentivize, and the best combination of acquisition and retention policies is adopted. The results show that user retention on the early days of adoption is better than the acquisition of new users. In this regard, acquisition should be focused on when the retention is in an acceptable state. Furthermore, the highest increase in the number of users occurs when there is a good balance between acquisition and retention.

1. Introduction

Startup companies can create innovative software products or services in a limited time and with few resources (Unterkalmsteiner et al., 2016). The main characteristics of startups are rapid evolution, teams of just a few people, and uncertainties about customer needs and the conditions of markets (Klotins et al., 2019).

Startup companies look for rapid growth in the number of their fans, users or customers (Conway & Hemphill, 2019). Their users are of two types including new and returning ones (see Fig. 1). In this case, the main questions are ‘How can startups increase the number of their users?’, ‘What is the effect of different acquisition and retention strategies on the number of users?’, and ‘Which users are more valuable to the startups?’

These questions are significant for almost all businesses, but they would be vital in the case of startups, which have few resources and are designed to grow fast (Graham, 2012). The questions are, indeed, hard to answer by using the metrics and frameworks in the startup literature or by the existing methods in the context of acquisition and retention. Therefore, a general agent-based model is proposed in this study to account for the heterogeneity of users, model acquisition and retention concepts simultaneously and analyze different practical what-if scenarios (Serrano & Iglesias, 2016).

One reason to use an agent-based modeling (ABM) approach is its capability to model the diffusion of products in a natural way, which occurs in the process of acquiring new users (Rand et al., 2018). The agent-based approach has been applied to many different fields of business (Rand & Rust, 2011). For example, it has been used in organizational studies (Cohen et al., 1972; Fioretti, 2013), marketing (Goldenberg et al., 2009; Hu & Lin, 2018; Robles et al., 2020), and supply-chain management (Giannakis & Louis, 2011; Julka et al., 2002). ABM is superior to the traditional equation-based and top-down models because it provides opportunities to model more real market attributes such as adoption threshold, individual responsiveness to information, and heterogeneity in the initial conditions of the population (Rand et al., 2018).

In an ABM approach, the agents are individual autonomous entities with specific properties and behavioral rules. The core concept of ABM is...
that modeling and analyzing complex systems would be possible by defining simple behavioral rules for agents (i.e. rules of behavior with each other and with the environment) and putting these rules together in a model (Rand & Rust, 2011).

In the specific agent-based model of the present study, the agents represent real users and can reflect their heterogeneous nature. A preferential attachment network connects the agents in a simulation process to represent the word-of-mouth (WOM) effects. Thus, the social network helps to establish social relations among real users. The agents of the system can decide to adopt on two main conditions: a) if they have adopted agents in their direct contacts or b) if they are influenced by the marketing activities of the company that offers the product.

After the adoption of a product, the company should also apply retention policies to retain its current users. Our model also presents a novel retention mechanism for the users or agents. This mechanism is based on two main elements to determine the retention probability for each agent. First, the model analyzes the recent patterns of behavior and use for the focal agent. Second, the model compares the focal agent to similar users, who have the same usage pattern, in order to decide if the agent is retained or not. By following this two-step mechanism, the model can approximately determine the final retention probability for that agent.

The model proposed in this study is applied to a real startup that develops game apps. It simulates the real users of the company in the agent-based system. The simulation runs for 55 days, and the model is thoroughly verified and validated in the way suggested by Rand and Rust (2011). The real data for each user come from Google Analytics reports. The validated model is used to answer two specific questions as follows:

1. Which acquisition and retention strategies have the highest impact on the number of users?
2. When does the overall number of users have the highest increase?

This paper is organized in several sections. A review of the literature is performed in Section 2. The agent-based model is then presented in Section 3. Section 4 is given to the verification of the model, its application to a real startup and the way it is validated. Then, some experiments are conducted on the model in Section 5. The main conclusion and some suggestions for future research are detailed in Section 6.

2. Review of literature

2.1. Main metrics to analyze startups

There are more than a hundred metrics recommended for startups (Kemell et al., 2020). Pirate metrics introduced by McClure (2007) is an example. This framework consists of five concepts including acquisition, activation, retention, revenue and referral. Another pack of metrics was introduced by Patel (2016) who mentioned customer acquisition cost, retention, churn, lifetime value, product metabolism, viral coefficient, revenue, activation, and referral. Also, according to Ellis and Brown (2017), a growth-hacking framework can work for acquisition, retention and revenue.

Besides the above-mentioned and all the other metrics, an impressive movement by Ries (2011), known as Lean Startup, stands out in terms of attractiveness and applicability (Bortolini et al., 2018). The Lean Startup emphasizes testing growth and value as two very essential hypotheses. The growth hypothesis focuses on new customers by considering product diffusion among them, but the value hypothesis focuses on retaining customers to evaluate how much they care about the product or the service. In other words, these two hypotheses of the Lean Startup focus on the acquisition and retention metrics (Ries, 2011).

The present study focuses on the two metrics mentioned and emphasized in the Lean Startup for some reasons. Firstly, these two metrics are involved in all the main frameworks of analyzing startups (Croll & Yoskovitz, 2013). Secondly, by considering these two hypotheses simultaneously, startups can analyze the effect of business decisions on the new and returning (or, in sum, all) users (see Fig. 1) and make the right decisions for each category. This is an essential issue because business decisions to increase the number of new and returning users are entirely different and even have different costs (Wertz, 2018). Thirdly, although there are a lot of metrics that startups can take into consideration, it is suggested that they focus on just a few metrics (Croll & Yoskovitz, 2013). Finally, there is extensive research on analyzing businesses according to acquisition and retention (e.g., see Danaher and Rust, 1996, and Erickson, 1993).

In this research, ABM is applied not only to analyze the past and current situations of the startups through the lens of acquisition and retention metrics but also to simulate the effect of different scenarios on these two metrics and the overall number of users. This kind of analysis cannot be done by just tracing the values of acquisition and retention

![Fig. 1. Number of all the users (returning and new) that a startup can have.](image-url)
Retention has a multifaceted nature and consists of the concepts of behavior and attitude as its two main aspects. The behavioral nature of retention involves certain variables such as the number of customers, number of active customers, frequency of buying and recency of buying. The attitudinal nature of retention involves variables like psychological commitment/loyalty, trust and empathy. In other words, the behavioral aspect is related to the behavior, while the attitudinal aspect has to do with hearts and minds (Aspinall et al., 2001).

Of these two aspects of retention, the behavioral view has gained more attention (Aspinall et al., 2001; Huang & Tsui, 2016). This is because the retention rate is influenced more by the behavioral aspect (Hawkins & Vel, 2013). In this aspect, recency and frequency are the two concepts that imply better customer retention and are related to each other. Since frequency fundamentally causes recency (Huang & Tsui, 2016), the present study just applies the concept of frequency to model the retention according to the principle of parsimony (Vandekerkhove et al., 2015). As shown in Table 1, frequency has been applied in most studies through different phrases such as reuse and repurchase.

In this study, as retention needs to be modelled at the agent level, the modelling is conducted on the basis of one of the latest definitions of retention proposed by Ross (2018), namely “a measure of the likelihood that a customer will return.” Therefore, this definition is applied to model retention through the frequency concept.

### 2.4. Acquisition versus retention

Focusing on acquisition or retention is a challenge for marketers (Furman et al., 2019). In this section, the previous studies are analyzed concerning this challenge in different aspects. At the end of the section, the main shortcomings of those studies are summarized, and the corresponding solutions are presented.

By applying a decision calculus approach to maximize customer equity, Blattberg and Deighton (1996) determined the optimal spending on acquisition and retention. Berger and Bechwart (2001) considered this subject in different market conditions and tried to allocate budgets to acquisition and retention. In another study, Pfeifer (2005) analyzed the subject in a situation where the cost of acquisition was five times bigger than that of retention. As the results showed, the optimal allocations to acquisition and retention were dependent on costs. Calciu (2008) compared the research works of Pfeifer (2005) and Blattberg and Deighton (1996) and proposed a generic formulation for both optimization methods. He also assumed that a lost customer would not go forever and (s)he might come back.

Reinartz et al. (2005) developed a conceptual framework and a model to balance acquisition and retention resources to maximize the long-term profitability of firms. Their model answered the questions of how much to spend across communication channels and how to spend. Investing in acquisition and retention in a competitive context was investigated by Musalem and Joshi (2009). They also provided suggestions about the type of customers that firms should concentrate on in both acquisition and retention activities. On the basis of the fact that acquisition and retention are intertwined, Dong et al. (2011) developed an incentive mechanism to simultaneously consider acquisition and retention. They analyzed the negative effect of acquisition on retention (i.e. the spoiling effect) on the firm performance when acquisition occurs through direct selling and delegation. They found that this negative effect noticeably bears upon acquisition and retention and the firm profit.

Ovchinnikov et al. (2014) presented a dynamic programming model for the optimal allocation of acquisition and retention under capacity limitations. They considered personalized retention offers for the existing customers and non-personalized acquisition offers for the general public. They also suggested the retaining of high-value customers instead of the acquisition or retention of low-value ones. Their research criticized the way firms approach the CLV values when they do not have an ample capacity to acquire new customers. King et al. (2016) introduce a dynamic programming model to allocate the resources of a firm optimally. They found that, when the customer base of a firm grows, the firm should change its focus from acquisition to retention.

In order to maximize the profit of firms, Rhouma and Zaccour (2018) proposed a dynamic programming model to optimize customer acquisition and retention expenditures for a subscription service. They applied the model in the telecommunications sector. Their empirical...
results showed that firms should invest more in retention than in acquisition. Chang et al. (2020) extended the work by Rhouma and Zaccour (2018) through adding competition. They applied a differential game model to optimally allocate acquisition and retention expenditures in a duopolistic market. In such a market, both firms invest in both acquisition and retention to expand their market shares. Their results showed that, for each firm, optimal investments in acquisition and retention depend on the effectiveness of acquisition and retention rather than on the number of customers or market shares.

After this overview of the studies in the field, the specific aspects of the existing pieces of research are explored in Table 2. Although different methods, such as the use of analytical and statistical models, have been applied in the field so far, none of them has considered the social interactions among entities. Bringing such interactions into focus is the important novelty of the present study because interactions between users or customers affect both acquisition (Goldenberg et al., 2009; Rand & Rust, 2011) and retention (Nitzan & Libai, 2011). The heterogeneity of customers or users is another important issue in marketing (Allenby & Rossi, 1998; Rand et al., 2018). As shown in Table 2, just a few of the analyzed studies consider the heterogeneity issue. Moreover, verification and validation are usually not performed in most of such studies, while it is one of the cornerstones of the present work.

In order to meet the above-mentioned concerns, this research seeks to deal with the interactions among agents through a preferential attachment network and by applying ABM. This approach of modeling is one of the best techniques to study the heterogeneity of individuals. Verification and validation tasks are also conducted to ensure the correctness of the model and its results.

3. Description of the acquisition and retention agent-based model

3.1. General structure

The model proposed in this study presents a finite set of agents including prospects, adopted users and returning or active users of a startup. These agents form the market size of the startup. The agents occupy the nodes of a social network. The edges in this network represent the connections and interactions among the agents that can lead to the spread of the WOM. The agents have two behaviors including a) adoption, which can occur just one time and is influenced by WOM and marketing activities, and b) retention or returning to use the app again, which can occur at each time step after adoption. Fig. 2 shows the general structure of the model with the acquisition and retention components of it.

The parameters that form the inputs of the model are WOM, marketing activity, retention, and the network topology. The outputs are the agents’ properties, the number of new and returning users, and the sum of the users including them all.

At the initial step of the model, both new and adopted agents are created along with their properties and the network structure. To define the population and the properties of the previously adopted agents at the first step of the simulation, the users who have already adopted the app and have a chance to return must be taken into consideration. One way to do this is to consider the average lifetime of the users. For example, consider an app that is used just for 60 time-steps on average. This means, in each time step, on average, only the users who adopted this app from the previous time step up to 60 time-steps before that time have a chance to use it again, not all of the previously adopted users. Therefore, the previously adopted users who have the chance to return should be defined separately for each startup according to its settings.

The model should run at the most for the duration between two developments of the app. Otherwise, the results will be affected by several development processes. This may lead to the misunderstanding of the relations between the development processes and the results.

3.2. Adoption mechanism

This part of the study addresses previous agent-based diffusion models that used the concepts of the Bass model, as in Goldenberg et al. (2001) and Stephen & Lehmann (2016). The product or service adoption, thus, occurs through marketing activities and WOM. The adoption probability of a user agent \( i \) or \( P_i \) is calculated as follows (Goldenberg et al., 2001):

\[
P_i = 1 - (1 - p)(1 - q)^i
\]

where \( p \) is the effect of marketing activities, \( q \) is the effect of WOM, and \( i \) is the number of active users connected to agent \( i \).

In conventional diffusion models, any adopted customer can spread WOM after adoption (Bass, 1969; Goldenberg et al., 2001; Watts & Dodds, 2007). In the model presented here, however, given that each previously adopted user in each time step has a chance to return to use the app or not (i.e., be active or not), it is just the active users who are allowed to transfer the WOM to the non-adopted users.

3.3. The retention mechanism

This section of the paper explains how the model calculates the retention probability at the agent level by using the frequency concept.
(Ross, 2018). In this regard, two properties are defined to indicate the frequency of use for every agent 

\( i \) as follows:

1. Total days \((T_i)\): the number of the time steps from the time step when the first installation of the app occurs.

2. Active days \((k_i^+; \tau \in [0, T_i])\): the number of the time steps the agent uses the app for its first \( \tau \) time steps.

The goal of the model in this part is to calculate the probability for the agent \( i \) to use the app again at the next time step (i.e., when its total days property is \( T_i + 1 \)). To achieve this goal, an analysis is conducted of the behaviors of other similar real users of the app from the population of users \( Z \). Also, what they did under the same conditions (i.e., when having the same properties as agent \( i \)) is assessed. These calculations, indeed, serve to clarify who, among all the real users, used the app with the same frequency as agent \( i \) and how many of them used it again at \( T_i + 1 \). This ratio helps to define the retention probability for agent \( i \) at \( T_i + 1 \).

In this way, \( T_i \) and \( k_i^+ \) are also defined as the total days and the active days properties for the real users respectively. In the population of \( Z \), users with two characteristics are looked for, those who have the same active days property as agent \( i \) when \( \tau \) is equal to \( T_i \) (i.e., \( k_i^+; \tau = k_i^+; T_i \)) and those whose total days properties or \( T_i \) are equal to or greater than \( T_i + 1 \). This subset of real users is denoted with \( S \). For using the app, the behavior of the members of \( S \) is similar to that of agent \( i \) at their first \( T_i \) time steps. In addition, their actions (i.e., using or not using the app again) at the next time step (i.e., when their total days property is \( T_i + 1 \)) are known. The number of these users is denoted with \( \sum_{u=1}^{s} f_u^{i+1} \) where:

\[
f_u^{i+1} = \begin{cases} 
1, & \text{if } k_u^+=k_i^+;T_i \text{ and } T_u \geq T_i + 1, \\
0, & \text{otherwise}.
\end{cases}
\]

Next, in the previous subset of the real users, namely \( S \), those who use the app at the next time step or at \( T_i + 1 \) are identified. While these users have the same frequency of use as agent \( i \) at their first \( T_i \) time steps (i.e., \( k_i^+; \tau_i = k_i^+; T_i \)), they also use it again or return when their time step is equal to \( T_i + 1 \). The number of these users is represented with \( \sum_{u=1}^{s} e_u^{i+1} \) where:

\[
e_u^{i+1} = \begin{cases} 
1, & \text{if } k_u^+-T_i = k_i^+; T_i \text{ and } k_u^+;T_i+1 = k_i^+; T_i + 1, \\
0, & \text{otherwise}.
\end{cases}
\]
preferential attachment network is chosen to be used (Barabasi & Albert, 1999). The choice is made because a high number of real-world social networks match this synthetic network (Moya et al., 2017) and it has already been applied in this context (e.g., see Rand & Rust, 2011; Stummer et al., 2015).

In this type of network, some agents or users have a disproportionately large number of connections to other agents, but some others have a small number of connections (Newman, 2001). To generate this network based on the Barabasi-Albert model, certain issues must be considered. These issues include a) an initial arbitrary network with \( v_0 \) nodes each of which must have at least one connection to the others and b) addition of a new node with \( v \) links (\( v \leq v_0 \)) in each step of the network evolution. The destination node of each new link is chosen by the following probabilities:

\[
w(x) = \frac{\text{deg}(x)}{\sum_y \text{deg}(y)}
\]  

In this equation, \( w(x) \) is the selection probability for node \( x \) which is equal to its degree (\( \text{deg}(x) \)) divided by the total degrees of all the nodes (\( \sum_y \text{deg}(y) \)) (Sayama, 2015).

![Fig. 3. An illustration of the agent i and the retention probability calculation steps.](image)

Data for 55 days from Google Analytics reports

Transforming the data to apply at the agent level

Market size: 3 millions
Scale-free network generated by the BA algorithm

![Fig. 4. Some details about the input data, market size and the research processes.](image)
4. Verification and validation of the model

4.1. Experimental dataset from a real startup case

The model has been applied to a real startup (some details are presented in Fig. 4). The startup develops game apps one of which is analyzed in this study. It is an online game in the action genre where users fight other users. The game is free to play, but each user can improve his or her power by buying extra in-game currencies named coins and diamonds. This game has a leader board that gives in-game currencies to top players.

The model simulates the users’ data over 55 days from the 23rd of December 2017 until the 15th of February 2018. Each day is represented by a time step. At each step, the numbers of the returning and the new users are calculated. More details will be given in the validation section (Section 4.3).

4.2. Model verification

Verification determines the correspondence between the implemented model and the conceptual model (Wilensky & Rand, 2015). In this part of the study, the steps expressed in Rand and Rust (2011) are followed. In the documentation step, the description of the model design in terms of general structure, adoption mechanism, retention probability and social network topology serves as the conceptual model documentation, and the code is also documented. In the programmatic testing step, however, unit testing is done by analyzing different parts of the model separately with different input values or functions. If the individual parts of the model work properly, they are combined and the whole model is verified again. Code walkthroughs are done by the presentation of each part of the model to an ABM expert and the application of his feedbacks. In the present study, debug walkthroughs were done for the model by checking the corresponding results in each step. Although it is difficult because of the complexity of most agent-based models (North & Macal, 2007), it was done in this study by putting together different parts of the model step by step and checking the overall functionality at each step. In testing the cases and during the scenario steps, corner cases were examined by setting \( p = 1.0, 0.1, r = 0.1 \), \( p = 0.1, r = 1.0 \), \( p = 0.5, 0.5, 0.0 \), and \( p = 0.5, 0.5, 0.5 \). In all these cases, the agents behaved as expected. Moreover, several sampled cases were selected with different \( p \), \( q \), and \( r \) values as well as different properties for previously adopted agents. The model was then implemented, and it functioned properly. Specific scenarios were used to test the model behavior in the retention part. To do this, the equation \( q = 0 \) was postulated, which means “no new user.” The behavior of the previously adopted users was traced as well. According to the real data, there was a declining behavior expected in the number of the returning users, which was because the users’ lifetime is not infinite. In this scenario, the model responded properly, and its response was in accordance with the customers’ lifetime. As another experiment, relative value testing was conducted, while one out of \( p, q, r \) was changed and the remaining items were kept constant.

4.3. Calibration of the model and the validation details

Validation is done to check whether the implemented model corresponds to reality or not (Wilensky & Rand, 2015). In this section of the study, the steps expressed in Rand and Rust (2011) are followed again. So, in the micro-face validation, product adoption occurs by the effect of marketing activities and WOM while each adopted agent returns to use the product again according to his or her history of use. These points seem valid “on the face”. Also, each agent has a social position in the preferential attachment network with information about his or her neighbors, not the whole population.

At the macro-face level, reference is made to the retention part of the model as the well-accepted agent-based model in Goldenberg et al. (2001) has been used for the adoption part. The aggregated patterns for the retention part of the model seem logical because, in each time step, the number of the returning agents must be equal (in the best case that rarely happens) or lower than that of all the adopted agents. This behavior is supported by many cohort analyses reports in the startup context.

Calibration determines the correspondence between the simulation outputs and the real data. The training part relates to the real data of this study and some others such as Ross (2018). In this section, the model is calibrated manually (Rand & Rust, 2011) to approximately define the range of the parameters. After this manual calibration, the exact values of \( p, q \), and \( r \) parameters. First, the numbers of the new and the returning agents in each time steps have to be similar to those in the training part of the real data. Second, the model has to generate agents who have the same features (i.e. “total days” and “active days”) as real users.

In this way, the data of the study are divided into training and test sets. Indeed, 85% of the data is for training, and 15% belongs to testing. This division is based on the hold-out approach proposed by Chica and Rand (2017). In the next step, the model is calibrated manually (Rand & Rust, 2011) to approximately define the range of the parameters. After this manual calibration, the exact values of \( p, q \), and \( r \) and the approximate maximum and minimum values of \( o_{df} \) are found. In order to fine-tune the \( r \) value, BehaviorSearch (Stonedahl & Wilensky, 2010) is used. It interfaces with NetLogo (Uri Wilensky, 1999) and can serve to find the parameters of NetLogo simulations. By the use of BehaviorSearch, the standard generational Genetic Algorithm (Holland John, 1975) is put to practice. The algorithm has a mutation rate of 0.03, population size of 50, crossover rate of 0.7, population model of a generational type, and tournament size of 3. The mean squared error (MSE) function is also used to calculate and minimize the deviation of the model outputs from the real data in the training part.

As Fig. 5 shows, the agents generated in the model have almost the same characteristics as real users after the calibration of the parameters. In order to do the empirical output validation, the model is first validated according to stylized facts. The adoption part of the model shows s-shaped curves in accordance with all technology adoption curves (Rand & Rust, 2011), and the retention part displays a descension in the number of the returning users according to the empirical evidence in the real data of this study and some others such as Ross (2018).

In the real-world data validation and after all the parameters are fine-tuned, the model is executed for 30 times using NetLogo’s BehaviorSpace (Wilensky, 2003). In addition, the overall number of the all users and the returning ones is obtained in each time step. Next, the data are imported to R (R Development Core Team, 2013) and averaged within each parameter set across all the 30 runs. Fig. 6 presents the results of the simulation versus the real data. The training part related to the calibration and the test part show that the simulation outputs are similar to reality (\( \text{RMSE} = 1.523 \) and \( \text{MAPE} = 6.080 \)).
5. Analysis of the results

In this section, the validated and calibrated model is used to analyze a real scenario and perform sensitivity evaluation. Since acquisition and retention policies are different in terms of implementation and consequences, the first issue to deal with is the benefits of user acquisition and user retention in comparison. In this regard, the effects of adding new users and returning previously adopted returning users on the overall number of the users over time are compared (Section 5.1). Next, a sensitivity analysis is done on the acquisition and retention parameters (Section 5.2).

5.1. The most valuable user to target

In this section, the effects of retaining different users on the overall number of the users are comparatively investigated over time. To do so, the model is run three times with three different settings added to the base model. In the first setting, three new agents (i.e., agents with total days’ properties equal to 0) are introduced in the population of the adopted agents. In the second setting, three agents in the population with $1 < \text{Total days} < 15$ and the highest “active days” properties are returned. This second setting, which considers those users who were adopters in the early days of using the app, are essential users to retain in the freemium gaming industry (Milosevic et al., 2017)). Finally, in the third setting, three agents with $26 < \text{Total days} < 35$ and the highest “active days” properties are returned. The reason for choosing agents with the highest “active days” properties is that they have more chance to return in comparison with the rest of the agents. These three settings or scenarios are set in each time step between the 5th and 15th. This fact is shown by the two vertical lines in Fig. 7. In this scenario, the model is run for 30 times for each case using NetLogo’s BehaviorSpace.

The results of the three above-mentioned changes are presented in Fig. 7. By comparing the gray and blue lines in the Fig. 7, we see that, as the lifetime of the added returned agents increases, the positive effect on the overall number of the agents decreases. A comparison of the blue and yellow lines provides evidence that retaining newly adopted agents (yellow line) is worth more than acquiring new agents (blue line). This finding is valuable since user retention costs less than user acquisition (Milosevic et al., 2017; Wertz, 2018).

5.2. Effects under different acquisition and retention scenarios

In this sub-section, the calibrated model is applied to analyze the effect of changes in the acquisition and retention parameters on the overall number of the users after 55 steps. The main goal is to understand the settings which lead to the best increase in that overall number. Fig. 8 shows three plots with changes in the parameters and their results. Each square in each plot is the average of 30 Monte Carlo simulations and represents the overall number of the users after 55 time-steps.

The results of the three above-mentioned changes are presented in Fig. 7. By comparing the gray and blue lines in the Fig. 7, we see that, as the lifetime of the added returned agents increases, the positive effect on the overall number of the agents decreases. A comparison of the blue and yellow lines provides evidence that retaining newly adopted agents (yellow line) is worth more than acquiring new agents (blue line). This finding is valuable since user retention costs less than user acquisition (Milosevic et al., 2017; Wertz, 2018).

In this sub-section, the calibrated model is applied to analyze the effect of changes in the acquisition and retention parameters on the overall number of the users after 55 steps. The main goal is to understand the settings which lead to the best increase in that overall number. Fig. 8 shows three plots with changes in the parameters and their results. Each square in each plot is the average of 30 Monte Carlo simulations and represents the overall number of the users after 55 time-steps.

The X-axis in all the three plots represents the changes under the effect of WOM (or $q$). The minimum value in this axis represents the current condition in the startup. This value means that the adoption probability for a non-adopted user in each contact with active users is 0.009. Regarding this axis, an analysis is performed of the effect of the increase in the probability from 0.009 to 0.7.

The Y-axis in all the three plots represents the changes under the effect of marketing activities (or $p$). Again, the minimum value in this axis represents the current condition in the startup, which is the output
of the calibration process. This value means that the adoption probability for a non-adopted user as a result of marketing activities is $7.0 \times 10^{-5}$. Regarding this axis, an analysis is performed of the effect of the increase in the probability from $7.0 \times 10^{-5}$ to $0.01$.

The difference among the three plots is in the changes occurring in retention. The plot in the left has no increase in the retention. In the next two plots, however, the retention of the users increases by 15% and 30% respectively.

According to the plot on the left, without an increase in the retention, acquisition activities (i.e., $p$ and $q$) would not increase the overall number of the users significantly. With an increase in the retention in the next two plots, however, the increase in the number can even be 12 times more significant. Therefore, it is recommended to assign the first priority to retention (Olsen, 2015).

According to the plot in the center and the one on the right (where the retention is increased), if the startup does not try to acquire new users, the overall number of the users will not grow significantly. So, when the retention is good enough, it is time for acquisition. If not, no considerable growth will occur.

The important notion in all the three plots is that a combination of improvements in parameters, rather than the improvement in one parameter, can lead to the best increase in the overall number of the users.

Fig. 8 can serve as a compass for this startup because it shows when and under what conditions the expected growth will occur. It can also guide the startup in different situations. For example, if the company does not have enough money to run advertisement campaigns, it can compensate it with non-organic WOM. This is like what Dropbox does; it gives extra gigabytes to those users who invite new users).

6. Conclusion and suggestions for future studies

In this research, an agent-based model of user acquisition and user retention is proposed to determine the overall number of the users. For the acquisition part, the research benefits from the rich literature on the modeling of product adoption and diffusion. In the retention part, the agent-based model is developed according to the user’s pattern of use and the data of the past users. To test the functionality of the model, verification was done first, and then the model was validated with the data of a real startup in the game category. Later, a scenario was run, and a sensitivity analysis was done on the acquisition and retention parameters to gain new insights.
Two major results were achieved in this research. First, besides trying to gain new users, it would be a great idea to focus on activating previously adopted users, which costs less. The return of newly adopted users affects the overall business better than gaining new users does. Second, paying attention just to acquisition or retention activities will not lead to a massive growth in the number of users; if both acquisition and retention improve proportionally, the growth can be surprisingly high.

An important contribution of this study is the introduction of an agent-based retention model that excellently fits conceptual models and real data. The main contribution, however, is the simultaneous modeling of acquisition and retention by the agent-based modeling approach. The importance of this modeling is that both retention and acquisition are essential for businesses in general and for the analysis of startups in specific (for example, see Chen (2016), Croll & Yoskovitz (2013), and Ellis & Brown (2017)).

This study can be expanded in other directions. Social networks influence diffusion (Muller & Peres, 2019), and, thus, the effects of applying different network topologies (e.g., small-world and stochastic networks) can be compared and investigated. Although it is hard to gather the data on the demographic characteristics of users (e.g., age, gender, location, etc.), these heterogeneities have effects on acquisition and retention (Rhouma & Zaccour, 2018) and can be analyzed through the extension of the model presented in this study. Churn prevention (Milosevic et al., 2017) is another issue that can be investigated by the analysis of the past behavioral patterns of churning users. In this regard, the effects of giving different motivations to the users who are susceptible to churn (i.e., who behave like churners) would be an interesting topic for further research. Finally, focusing on the purchase history of users, as suggested by Sela et al. (2018), is another significant issue to explore. In this case, the explorer can a) extract the characteristics of payers in the past data and use them in a model to define the obtained revenues and b) analyze the effect of retaining the premium users (e.g., by giving in-game currencies or push notifications) on the revenue.

CRediT authorship contribution statement

Amir Sayyed-Alikhani: Conceptualization, Methodology, Software, Validation, Formal analysis, Investigation, Resources, Data curation, Writing - original draft, Visualization. Manuel Chica: Conceptualization, Methodology, Validation, Formal analysis, Writing - review & editing, Supervision, Resources. Ali Mohammadi: Supervision, Project administration.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

The authors would like to thank Arman Balali-Moghaddam for providing access to real data and the Google Analytics support team for some helpful suggestions. We would also like to thank Mohammad Jalal Karimi for his valuable comments. This work is supported by the Spanish Ministry of Science, Andalusian Government, and ERDF under grants EXASOCC (PGC2018-101216-B-I00), SIMARK (P18-TP-4475), and the Ramon y Cajal project (RVC-2016-19800). Barabasi, A.-L., & Albert, R. (1999). Emergence of scaling in random networks. Science (New York, N.Y.), 286(5439), 509–512. https://doi.org/10.1126/science.286.5439.509
null