Gamification, Virality and Retention in Educational Online Platform 
Measurable Indicators and Market Entry Strategy

Ilya V. Osipov 
i2istudy SIA
Krišjāņa Barona iela, 130 k-10, Rīga,
Lv-1012, Latvija

Alex A. Volinsky
Department of Mechanical Engineering, University of South Florida, 4202 E. Fowler Ave., ENB118, Tampa FL 33620, USA

Vadim V. Grishin 
i2istudy SIA
Krišjāņa Barona iela, 130 k-10, Rīga,
Lv-1012, Latvija

Abstract—The paper describes gamification, virality and retention in the freemium educational online platform with 40,000 users as an example. Relationships between virality and retention parameters as measurable metrics are calculated and discussed using real examples. Virality and monetization can be both competing and complementary mechanisms for the system growth. The K-growth factor, which combines both virality and retention, is proposed as the metrics of the overall freemium system performance in terms of the user base growth. This approach can be tested using a small number of users to assess the system potential performance. If the K-growth factor is less than one, the product needs further development. If the K-growth factor is greater than one, the system retains existing and attracts new users, thus a large scale market launch can be successful.

User attraction and retention mechanics are discussed based on the peer-to-peer online language training platform, which utilizes freemium business model. Key system metrics are derived to assess the future commercial potential and making decisions to either fund an advertising campaign, or continue with project technical improvements. The paper can be of interest to venture capitalists as a method to assess freemium projects.

Keywords—Gamification; virality; retention; freemium; K-factor; metrics; open educational resource; e-learning

I. INTRODUCTION

There are numerous products utilizing the freemium model, such as mobile applications, software as a service (SaaS) solutions, shareware software, web applications and others [1]. However, the freemium model is not as simple as it may seem. The authors analyzed the statistics of the users’ behavior in the educational collaborative platform available to everybody as shareware and through the freemium model [2]. The platform is a web site for learning foreign languages with users from all over the World [3]. The main idea of the system is based on the fact that regardless of all the grammar learned in college, students are lacking live interactions with the native speakers to increase their spoken language skills [4]. Finding a native speaker is not an easy task, which typically also requires paying for tutor lessons. It was noticed from the students studying Spanish that the professional teacher is not required for the student to learn basic communication skills. What’s needed is a partner, who’s ready to help using already prepared materials, a Spanish native speaker. Spanish native speakers are eager to learn English in exchange for teaching Spanish. The idea of time banking [5] was used to track how much time each user is learning a foreign language and teaching native language.

The readily available audio-video conferencing technology between the users was combined with the pre-defined lessons, divided into step-by-step cards, understandable by the non-

Fig.1. Typical cycles of the freemium application

Professional teachers, along with the learning/teaching time tracking and an online system of finding and connecting users. This is how the online learning/teaching educational resource, called i2istudy was started [3]. Using this system, which operated from April through August 2014 in the beta mode, and collected over 40,000 users, the authors have conducted several measurements and studies.

Fremium (a combination of the words Free and Premium) business model assumes the maximum product market
distribution, along with the capture and retention of the largest possible number of users. Part of the users, which for various products varies from 3% to 10%, takes advantage of the premium features, allowing the creators not only to pay for the entire system upkeep, including free parts, but also to make a profit [6].

Let’s consider the basic functions of the freemium products aimed at the mass market:

The main (base) function for which the users come (in our case it is learning foreign languages).

User retention, including return users.

User monetization.

User attraction and virality stimulation (existing users attracting new users).

Figure 1 shows a typical cycle of the freemium application. Arrows indicate main user and information flows, including the 4 main cycles:

The in-app cycle is the main application cycle, the core cycle, the basic function for which the user decided to use the application (in this case it’s practicing foreign language skills with native speaker).

The monetization cycle (denoted by the small dollar sign $ in Figure 1). This is an additional cycle, which attracts the most ventrurous people involved in the process, which represents additional features. This cycle is smaller, since it is not available to all participating users (especially in the freemium business model).

The retention cycle is when users leave and subsequently return into the system. To successfully return and retain the users in the system, special means are utilized, from e-mail notifications, social networks and other communication channels reminders of the events, which occurred during the user absence from the system.

Gamification, which is using game mechanics in the non-game context, is actively used in the user retention cycle. This includes motivators, such as game currency (time banking in our case), system content divided into achievement levels, user titles and badges, and peer evaluations (after each lesson both the teacher and the student can rate each other) [3].

The viral cycle consists of the existing users inviting new users from the external environment (e-mail, social networks, blogs, forums, personal websites, applications, and other communication channels), including the new users accommodation.

Besides, the diagram in Figure 1 also shows different user flows into the web application, including organic “word of mouth” users, bookmarks, search engines, motivated and purchased users, along with the invited users [7]. The downward arrow shows users lost directly from the web front page (landing page), as well as from any other of the mentioned cycles. It should be noted that the application cycles: the core, viral, retention and monetization cycles are antagonistic, as they are competing for the user attention, which is always lacking. The system developers must understand which cycles have priority.

When creating freemium products, there are two main business approaches:

Purchasing and other paid user attraction (traffic). Part of the traffic can be monetized by selling additional premium services and attracting new users by spending the money earned. The key factor in this approach is money, thus a successful monetization model is required to involve a significant percentage of users in the paid mechanics to maintain the balance. The positive balance must exist between the revenues from the existing customers minus the cost of attracting new users. Moreover, the cost of attracting new users can be substantial, and there is a risk not to recover this high cost from monetization. Pluses of this approach include fast money earning, and that the K virality factor (K-factor) [8] can be less than one (discussed below).

Involving existing users into the product promotion through virality. It is necessary to ensure that the virality coefficient (the K-factor) is significant, which for a number of products is difficult and even unattainable. The volume of users with this approach is growing exponentially until it reaches saturation [9]. The product can contain features from both approaches, with the emphasis on monetization and the emphasis on expanding the user base. However, given the limited users attention, one of the approaches must be dominant.

II. VIRAL USER BASE EXPANSION

Let’s consider the second freemium approach with an emphasis on virality. David Skok, successful venture capitalist, wrote about the freemium virality-emphasized products [10]:

“in a typical business the single biggest expense is sales and marketing and recognize that offering a free product/service is an extremely smart way to acquire customers at a low cost that can then be monetized in a different way.” “Another powerful effect of using the free strategy is that it usually results in a far larger customer base using the free products, who become proponents for your company. This expanded footprint or market share can have a huge effect on the price that acquirers or investors are willing to pay for your company, as they recognize that even though these customers have yet to be monetized, they represent a great potential for future monetization. Twitter and Facebook
are two perfect examples of this.” “Another way of looking at the importance of footprint or market share is to recognize the importance of market leadership. In the tech industry, market leadership is usually self-reinforcing unless the company does stupid things to annoy its customers. Even if you have gained market leadership by giving away a product/service for free, the financial markets and acquirers realize that market leadership is worth a significant premium over niche players that may have more revenue.”

However, the strategy of viral user attraction cannot be utilized forever. Seufert, the author of the book Freemium Economics [1] presents a graph, similar to the one shown in Figure 2. All efforts invested in the virality mechanics will not bear fruit when the market niche is already saturated, and all potential users either already use the product, or know about it, but prefer not to use it. Obviously, this is the best time to refocus the product and change the user’s attention to monetization, which was discussed as the strategy number one.

III. VIRALITY REALIZATION METHODS (INVITATIONS)

The viral marketing requires several components: the sender, the message and the medium for dissemination, including recipients, along with the context in which the message is received. There are two ways for the user to invite new users:

TABLE I. THE NUMBER OF PUBLISHED INVITATION LINKS (OPEN INVITATIONS) AND HOW MANY USERS POST THESE LINKS ON THE WEEKLY BASIS

<table>
<thead>
<tr>
<th>Week</th>
<th>Publishes total</th>
<th>Published users total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

TABLE II. THE NUMBER OF THE NEWLY JOINED USERS INVITED THROUGH THE OPEN LINKS, INCLUDING ALL TYPES OF INVITATIONS

<table>
<thead>
<tr>
<th>Week</th>
<th>Invited by public link</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Fig.2. Saturation point in the total number of users with time. Adapted from Seufert, 2014 [1]
Open invitations - is the viral mechanism, where the user places invitations in social networks, blogs and personal web pages, etc. to address an undefined set of individuals.

Direct personal invitations initiated by the existing users to the potential new users using different means of communication by e-mail, personal communication, social networks, SMS, etc.

Typically it is hard to account for all open invitations. The authors used simplified statistics by calculating how many people were invited by this method, and how many people were able to use this method to initiate invitations. The system calculates how many open invitations were made by each user (via built-in system instruments), and how many new users joined as a result (including open invitations initiated by the user and not generated by the system). Table I lists the number of published invitation links (open invitations) and how many users posted these links on the weekly basis.

Statistics reflects only built-in invitation publication mechanisms. Table II lists the number of the newly joined users invited through the open links, including all types of invitation. Personal direct invitations allow calculating all parameters and quantifying all steps of the viral cycle. The system accounts for how many users make personal invitations, how many invitations are generated per each user, how many invitations reach the addressee, how many recipients come to the service, and how many register and get involved in the learning/teaching process. Table III lists the number of individual invitations and the number of users that sent such invitations, while Table IV lists the number of users who joined as a result of direct personal invitations.

IV. THE VIRAL CYCLE AND THE K-FACTOR

Let’s define the metrics parameters. Here, the term user means registered and authorized user of the service.

dU stands for daily users;
dNU are daily new users;
dAU are daily active users (users who spent more than 5 minutes in the system);
U is the total number of all users;
IU is the number of invited users;
Di is the total number of invitations per day;
AiPSU is the average number of invitations per spreading user (AiPDSU is the same per day);
AiPU is the average number of invitations per user;
DIU is the number of daily invited users;

IPi is the ratio of people who accepted an invitation to the number of invitations sent (conversion percentage).

Conceptually the K-factor is the average number of additional users introduced to the product by each user [1]. For practical purposes we calculated the K-factor as the ratio between the users attracted through viral methods in a certain time period, to all active users in this time period [11]. Theoretically, we should have used the previous time period, setting it equal to the duration of the viral cycle. However, the length of the viral cycle is difficult to establish, since the reaction to the invitation timeline is very short in our system, as in a typical case, sending invitations and accepting them gets completed in one day, or less. For calculating the K-factor, only new users (dNU), or all users in a certain time period (dU) can be used, along with the active users in a certain time period (dAU). The authors used active users (not considering uninterested users, who spent very little time in the system) as the base, as it gives the most accurate results in our estimates. For practical purposes some sources used only new users (dNU) as a base, comparing all users attracted virally with all types of new users. The authors believe this is not quite correct, since all active users contribute to virality and not only new users, thus:

\[
Local K_{factor} = \frac{dU}{dAU}
\]

(1)

The term K-factor comes from epidemiology, “in which a virus having a K-factor of 1 is in a “steady” state of neither growth nor decline, while a K-factor greater than 1 indicates exponential growth and a K-factor less than 1 indicates exponential decline.” [12]. The K-factor, which is also called viral coefficient in the literature, can be calculated as the number of invitations sent by each user multiplied by the conversion percentage of the new users [13]. For example, if the average number of invitations per user is 5, and 20% of the invitees register in the system, i.e. become new users, then the K-factor = 5*0.2 = 1. Time-independent K-factor, averaged over the whole time of the system operation, the authors call the global K-factor, which is calculated in the following way. The conversion percentage, IPi is the number of invited people, IU, divided by the number of invitations, i:

\[
IP_i = \frac{IU}{i}
\]

(2)
AiPU is the average number of invitations per user, calculated as the average number of invitations per user, i, divided by the total number of users, U:

\[
AiPU = \frac{i}{U}
\]  
(3)

Then the global K-factor is calculated as the product of the average number of invitations per user, AiPU (equation 3) and the conversion percentage, IPi (equation 2):

\[
Global \ K_{factor} = AiPU \cdot IP_i
\]  
(4)

The K-factor dynamics reflects the users’ mood swings, and how they react to the introduction, activation or deactivation of one or another viral mechanics, involving them in the activities of inviting new users, and whether these mechanics are well accepted. Thus, for practical purposes the authors utilized the local K-factor, calculated daily. The authors call it the daily K-factor, dK-factor, which is based on the daily active audience, dAU. This daily K-factor is

\[
Daily \ K_{factor}, dK = dAU \cdot IP_i
\]  
(5)

Figure 3 shows the i2istudy foreign language educational platform K-factor dynamics.

It is important to understand that if the K-factor is less than unity (e.g. 50%), in the absence of retention (when the loyalty of existing users is zero), the system growth attenuates. In the best scenario, such virality mechanism partially compensates the users’ loss as a result of the normal loyalty retention cycle decrease. This K-factor increases the effectiveness of paid user attraction. For example, purchased 100 paying users get involved in the viral mechanics and invite additional 50 users, which reduced the average price of each user and saved the budget. If the K-factor is greater than 1 (say it is 200%), it leads to the geometric progression growth of the user base. For example, purchased 100 users attract 200 new people, and if the K-factor remains the same, the new users will attract 400 new people, and so on. Virality works as long as the entire mass of potential users will not reach saturation in their social matrix and a given market, as schematically illustrated in Figure 2.

V. VIRALITY, RETENTION AND MONETIZATION RELATIONSHIP

In his book Freemium Economics, Eric Benjamin Seufert [1] in the Virality and Retention section, on page 175 wrote: “Virality and retention exist on opposite sides of the acquisition threshold: virality describes how users are introduced to a product, and retention describes how long users remain with a product. But in essence, both sets of metrics measure the same general sense of delight users feel for a product, manifested in different ways. To that end, virality and retention generally exhibit a positively correlated relationship: products that users are inclined to return to over a long period of time are also likely to be products that users invite others to join.”

In our opinion, virality and retention are characteristics amenable to manipulation by the creators of the product. Even a weak product can successfully maintain good retention and virality performance if appropriate mechanics and effects (impact, gamification) are well integrated into the product and successfully motivate users to these actions. This situation resembles a grocery store, where buyers are manipulated by the layout, marketing, branding, packaging and a discount system, and buy groceries that are not the best and healthy as a result [14].

Certainly all three parameters: virality, retention and monetization are related. Users with high product loyalty get increasingly involved in the mechanics of virality and monetization [15, 16]. Despite competition for the user’s attention, these mechanisms may spur one another, and all sorts of techniques, such as gamification, which is usually considered in the literature as part of retention, can serve monetization and virality. Oddly enough, monetization, can also spur virality and retention. For example, premium paid services can be alternatively earned by participation in the viral and gamification programs. The dollar price of these premium options demonstrates their value to the users. For example, when the user knows the cost of acquiring new premium options for real money, it may be easier to motivate
the user to earn these premium options by performing certain tasks and actions, such as inviting friends.

It is important that the experience of using the main basic functions of the product cause admiration, then the virality and retention mechanics come into play. Virality and retention should be balanced. For example, with perfect virality and poor retention, the growth of the user base, caused by the successful virality, will compensate for the loss of the same base due to disloyal users. The opposite situation of poor virality with excellent retention leads to the product and its user base stagnation, and eventual defeat by the competitors.

Coefficient of the product audience growth, K-growth, can be expressed as a sum of the coefficients of the viral K-factor and the retention factor, K-retention:

$$K_{growth} = K_{factor} + K_{retention}$$

(6)

Equation 6 is the main formula of the freemium product growth, based on the viral spreading. It is clear that this formula does not take into account alternative methods of attracting users, such as paid users and organic users, who came through search engines, word of mouth, or due to the brand popularity. K-retention is always less than one over a long period of time, since no products can retain its audience 100% at all times.

$$Local \ K_{retention} = \frac{dU - dNU}{dU_{-1}}$$

(7)

Where dU is the daily audience for a given day; dU-1 is the previous day audience and dNU are the new users for this time period. It’s convenient to use only active audience for calculations, by taking into account only the new users that have become active, but not all registered users. Similar situation is with the new invited users, among which only active users are accounted for:

$$Local \ K_{retention} = \frac{dAU - dNU}{dAU_{-1}}$$

(8)

For example, if the viral K-factor is 20 %, and K-retention factor is 90% (i.e. 9 out of 10 people are coming the next day), the growth coefficient will be 0.2 + 0.9 = 1.1 and the system will grow on its own by 10% of its daily (or other accounting period) audience. Coefficient of the system self-growth can be represented as:

| TABLE V. DATA AND PARAMETERS CALCULATIONS FOR THE I2STUDY PROJECT |
|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|---------------------|
| week | 05.05 | 12.05 | 19.05 | 26.05 | 02.06 | 09.06 | 16.06 | 23.06 | 30.06 | 07.07 | 14.07 | 21.07 | 28.07 | 04.08 | 11.08 | 18.08 |
| All active users (xAU) | 297 | 801 | 867 | 979 | 1080 | 1213 | 1827 | 2126 | 1763 | 2624 | 2572 | 1924 | 1716 | 1810 | 1576 | 947 |
| New active users (xNU) | 239 | 575 | 572 | 643 | 695 | 776 | 1358 | 1464 | 1140 | 1881 | 1643 | 1129 | 998 | 1066 | 824 | 358 |
| Invited active users (xIU) | 5 | 9 | 10 | 32 | 54 | 81 | 85 | 41 | 73 | 68 | 52 | 65 | 77 | 52 | 42 |
| K-Factor = xIU / xAU, percent | 2 | 1 | 1 | 3 | 5 | 7 | 5 | 2 | 4 | 3 | 2 | 4 | 4 | 3 | 4 |
| K-Retention = (xAU - xNU) / xAU, percent | 76 | 37 | 39 | 39 | 40 | 39 | 36 | 29 | 42 | 35 | 31 | 37 | 43 | 42 | 37 |
| K-Growth = (xAU - xNU + xIU) / xAU, 1, percent | 79 | 38 | 42 | 45 | 48 | 46 | 38 | 33 | 46 | 37 | 33 | 41 | 48 | 44 | 40 |
\[
K_{\text{growth}} = \frac{dAU_1 - dNU + dIU}{dAU_0}
\]

(9)

This is the ratio of the audience from the next period of time without accounting for the new users, but accounting for the users invited by the viral techniques, divided by the audience from the previous time period. If \(dNU\) and \(dIU\) are equal, then all new users get involved through viral methods exclusively (there is no paid and organic traffic), and in this case:

\[
K_{\text{growth}} = \frac{dAU}{dAU_{-1}}
\]

(10)

Table V shows the data and the corresponding parameters calculations for the i2istudy project.

As a side note, it is necessary to take into account that the very properties of the product may be a barrier to its viral spread. For example, users absolutely don’t want to advertise to their friends that they participate in dating services [17], which negates any virality efforts. On the contrary, the users promote their morning runs and other physical exercises, even without strong viral mechanics and ingenious motivations [18, 19]. An obstacle to the viral spread may be excessive annoyance of the viral mechanics, which can be negatively perceived by the existing users, and even considered as spam by the invitation recipients [20]. In addition, it’s a common mistake to promote business-to-business (b2b) services using virality methods, which usually gives poor results, with the exception of some individual cases.

If the \(K\)-growth factor is less than one, then the product cannot grow and loses users with all the consequences for the product and its team. However, if the team can achieve the \(K\)-growth factor greater than one, the product grows exponentially. It is the ultimate goal for the product team to achieve non-paid user base growth. This is necessary to achieve the project’s capitalization exceeding the investment in the purchase of the user base. Note that the positive \(K\)-growth factor can compensate for other shortcomings, such as the quality of the product itself. For the overall project development strategy, investment in virality and retention is a viable alternative to investments in advertising and public relations. This is often a cost-effective solution, since compared with the cost of development (programmers’ salaries, etc.), marketing and associated staff costs can be quite high. The freemium product development diagram in terms of the \(K\)-growth factor is shown in Figure 4.

VI. CONCLUSIONS

When building a freemium product, it is wise to take it to the market and to work out the viral and retention mechanics on small volumes of paid audience, since these mechanics can be easily evaluated statistically and analyzed. Having a positive \(K\)-growth factor, venture capital funds can be attracted, and the project can be brought to a large market (Figure 4).

As for the specific product, which was used to conduct these studies, it is clear that the value of the \(K\)-growth factor varied around 40% (last line in Table V), which was not a satisfactory. As a result, the product will be reworked and improved. First of all, studies involving test users identified a misunderstanding problem, where users stopped using the product before figuring out its functions and features.

Thus, the first task is to develop and improve intuitive user interface. Second is to encourage and motivate existing user invite their friends to grow the overall user base. Further studies, based on the principles and metrics outlined in this paper, will show how successful these improvements have become.

The authors believe that the described method can be applied for a wide variety of freemium products, providing objective means of assessing maturity of information services and applications, which have not yet seen large market launch. Venture capital is often used for scaling up the business and increasing the user base. Thus, venture capitalists need the means and methodologies of assessing investment risks in different products.

ACKNOWLEDGMENT

The authors would like to thank the i2istudy.com team for their dedicated efforts: Anna Prasikova, Ilya Poletaev, Andrei Poltanov, Elena Bogdanova, Vildan Garifulin, Mihail Shagiev and Franziska Rinke.
REFERENCES

www.ijacsa.thesai.org

18 | P a g e