UC Davis

UC Davis Previously Published Works

Title
Peeking into the Invitation-based Adoption Process of OSN-based Applications

Permalink
https://escholarship.org/uc/item/6m41s7ct

Journal
ACM SIGCOMM COMPUTER COMMUNICATION REVIEW, 44(1)

ISSN
0146-4833

Authors
Rahman, MR
Noel, P-A
Chuah, C-N
et al.

Publication Date
2014

License
CC BY-ND 4.0

Peer reviewed

eScholarship.org Powered by the California Digital Library
University of California

Peeking into the Invitation-based Adoption Process of OSN-based Applications

M. Rezaur Rahman
University of California-Davis
Davis, CA, USA
mrrahman@ucdavis.edu

Pierre-André Noël
University of California-Davis
Davis, CA, USA
pnoel@ucdavis.edu

Chen-Nee Chuah
University of California-Davis
Davis, CA, USA
chuah@ece.ucdavis.edu

Balachander Krishnamurthy
AT&T Labs - Research
Florham Park, NJ, USA
bala@research.att.com

Raissa M. D'Souza
University of California-Davis
Davis, CA, USA
raissa@cse.ucdavis.edu

S. Felix Wu
University of California-Davis
Davis, CA, USA
sfwu@ucdavis.edu

ABSTRACT

Online social network (OSN) based applications often rely on user interactions to propagate information or to recruit more users, producing a sequence of user actions called *adoption process or cascades*. This paper presents the first attempt to quantitatively study the adoption process or cascade of such OSN-based applications by analyzing detailed user activity data from a popular Facebook gifting application. In particular, due to the challenge of monitoring user interactions over all possible channels on OSN platforms, we focus on characterizing the adoption process that relies only on *user-based invitation* (which is applicable to most gifting applications). We characterize the adoptions by tracking the invitations sent by the existing users to their friends through the Facebook gifting application and the events when their friends install the application for the first time. We found that a small number of big cascades carry the adoption of most of the application users. Contrary to common beliefs, we did not observe special influential nodes that are responsible for the viral adoption of the application.

Categories and Subject Descriptors

H.4.3 [**Information Systems**]: Information Systems Applications—*Communications Applications*

General Terms

Measurement

Keywords

Online Social Networks, Social Gifting, Facebook Applications, Information Diffusion, Cascade, Application Adoption

1. INTRODUCTION

Word-of-mouth diffusion of information over online social media has recently been regarded as an important mechanism to increase adoption of a new idea, technology, or product. Such diffusion mechanism has been exploited by organizations and business to grow their number of users (or followers). Targeted advertising is a common strategy with the goal of influencing groups of users (and their friends or followers) in typical social networking or micro-blogging sites like Facebook and Twitter. Therefore, it is of great interest to advertisers to efficiently and effectively identify special individuals, often referred to as the "influentials" [4, 19] or simply "influencers", that can help recruit significant number of other users (e.g., their friends or their followers), leading to large-scale adoptions (or cascades) of new products/information [17, 8]. Although there are prior studies on product adoption process from a marketing and/or mass communication point of view, considerably less is known about user influence and adoption process for OSN-based applications. Indeed, due to the privacy awareness in sharing of user activity data [10], it is challenging to obtain sufficient empirical observations to quantify how a user is recruited or influenced to adopt an application or an idea circulating over online social media such as Facebook. Understanding the adoption process is not only important for analyzing the growth of an OSN-based application, but it is also a significant indicator of its longevity.

In this paper, we provide a first step towards filling this void by performing an in-depth analysis of the adoption process of a popular Facebook gifting application. Facebook, with a user base of over one billion, provides access to a huge collection of social applications that account for \$6+ billion industry as of 2012. We acknowledge that there are numerous channels that can be exploited by Facebook application developers to advertise new features/applications and recruit more users. These include invitations sent from one user to his/her friends from within an application (referred to as *Application Requests* ARs), news feeds, targeted emails, and paid advertisements on users' page. However, it is extremely challenging to collect data on inter-user communications over all the available channels that may influence the adoption process we intend to study. Nevertheless, our previous study [13] reveals that applications in the gifting genre only uses ARs for inter-user communication (e.g., sending "gifts" to one another). Hence, we use gifts and ARs interchangeably in this paper. Since ARs are generated through the application, such inter-user communication can be observed by analyzing user activity data collected at the application server. Understanding the adoption process of gifting applications has its own significance given that gifting is the second most popular genre of Facebook applications. In addition, the findings can also apply to other genre of applications that rely heavily on ARs for influencing users to adopt the application.

For our study, we analyze detailed user activity data from a popular Facebook gifting application, *iHeart*, that contains 2 billion entries of user activities (in the form of 'heart' gifts) generated by 190 million users in a span of 64 weeks. In December 2009, iHeart was one of the top three Facebook applications in terms of monthly active users. In a typical Facebook gifting application, when a User A sends a gift to User B (who has not installed the application yet), the Facebook platform will deliver an AR to User B, and User B can either accept it (install the application) or ignore it. Hence, ARs represent a low-cost mechanism to boost the active user base. In our case, we record the time a user installed the iHeart application for the first time. We refer to this event as *adoption* of iHeart, and we track all the ARs received by this user prior to the adoption of iHeart. Note that a user can receive multiple ARs from different friends before he/she adopts the application. However, users who install iHeart to send gifts before receiving any AR are considered the *seed* nodes. In other words, these inter-user communications through ARs form a sequence of activations or adoptions generated by a contagion process, in which nodes cause connected nodes to be activated with some probability. We refer to this sequence of activations or adoptions as a *cascade*. Using the empirical data, we study the characteristics of the cascades to investigate the following questions:

- *What does the adoption process look like?*
- *Which factors may drive large-scale adoptions? How important are those factors to shape the adoption process?*
- *Who are the influential users (if they exist)?*

Unlike a previous study [7] on micro-blogging services, our empirical study reveals that a small fraction of big cascades (top 3%) account for 80% of the user population that adopt the application. We also find that these cascades involve many generations of adopters, unlike the study shown in viral marketing where recommendation chains terminate after just a short number of steps [11]. Our study shows that it is not important to attribute the 'influence' of an adoption to a particular sender when a user can receive multiple ARs from different senders. The users that adopt the application will contribute to the growth of the cascades regardless of the invitation order or count from different senders. We also study hypothetical characteristics of the potential influential users and analyze their impact on the cascades by secluding these nodes from the adoption process in our simulation.

The rest of the paper is organized as follows. Section 2 introduces the background and related work, which is followed by the dataset and the measurement methodology used in Section 3 of this paper. Section 4 shows the measurement-based characteristics of an adoption process. Finally, we conclude the paper in Section 5.

2. RELATED WORK

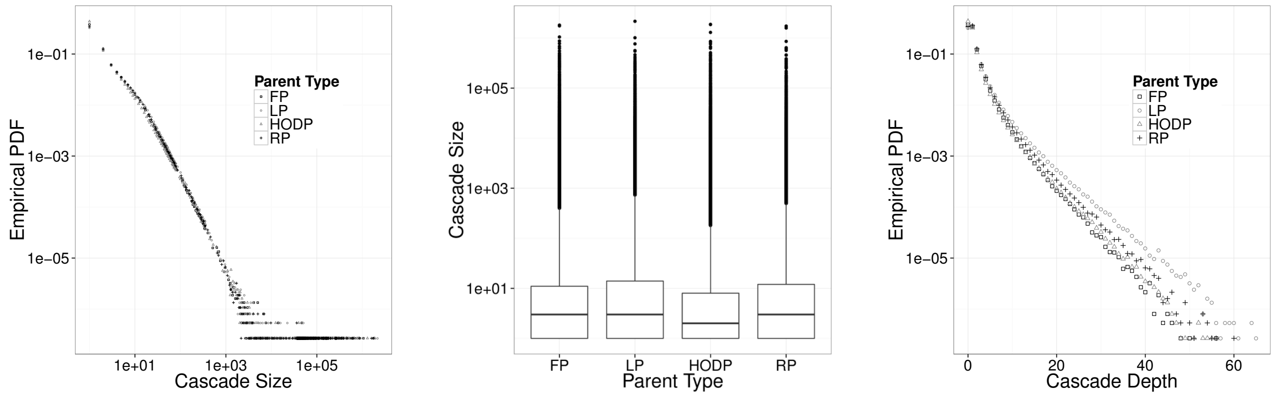
Information diffusion has long been studied by researchers to explore the role of influential users in the adoption of new products [3] or innovations [15]. Many of these studies have tried to understand the cascades or diffusion models and their appearance in the overall population. Recently, Goel et al. [7] studied the cascades of URLs shared in different micro-blogging services as well as the cascades of common interest shared in different communication platforms. Their analysis demonstrated that multi-step diffusions are rare, and the vast majority of the adoptions occur within one hop from a seed node. Similar characteristics are also observed in photo-sharing on Flickr social network by Cha et al. [5], who showed that information spreading on Flickr is limited to the local users around the initiator of the content, i.e., the photo uploader. On the contrary, our empirical study reveals that the biggest cascades (top 3%) that account for the majority of application adoptions (in terms of population coverage) span multiple generations of initiators, i.e., multiple hops away from the original seeders. We further investigate factors that may drive large-scale adoptions observed in our data.

Numerous studies have focused on identifying influential users and analyzing their roles in information diffusion. Unfortunately, attributing influence to specific users is challenging because the network over which word-of-mouth influence spreads is generally unobservable. Liben-Nowell and Kleinberg [12] attempted to trace information flow on a global scale using Internet chain-letter data. They concluded that diffusion occurs along trees that are very deep and narrow, and the diffusing process can be very complex, involving large number of steps. Similarly, diffusion chains in Facebook news feed are shown to be extremely long and not usually limited to a single chain-reaction event [12]. Moreover, the analysis of word-of-mouth exchange of URLs shared on Twitter [14] reveals that the propagation trees are two orders of magnitude wider than their depth. On the other hand, Watts has shown a simple model [16] that explains how the connectivity of the underlying network drives the cascade growth. Continuing with this argument, Watts and Dodds [18] explained that it is not the influential users but the critical mass of the easily influenced individuals who are responsible for the massive cascade growth.

Although identifying influential users is usually an intractable problem due to the lack of empirical data, a growing number of researches are focusing on maximizing the influence of selected seeds (for initiating the information cascades). For example, Kempe et al. [8] found near optimal algorithms for selecting a maximally influential set of seeds for several classes of models. Also, it is found that users in the dense core of the network maximize the diffusion of disease when initiated by those users [9]. On the other hand, Bakshy et al. [1] showed that quantifying potential influential users on Twitter is costly and word-of-mouth diffusion needs to target a large number of those individuals. A clearer distinction among the different individuals in terms of level of influence in user created contents is made in a later study by Bakshy et al.[2] in the context of a time-evolving social network. This study concluded that some individuals play a more active role in the content transfer than others, but did not assess their overall impact on the final adoption size. In contrast, we perform several "what-if" scenarios by specifically removing active individuals (e.g., seeds of large cascades) and analyze how the lack of their actions/influence affect the overall adoption process. Our results show that the adoption process is a complex system based on influence of many users rather than a chosen few.

3. DATASET AND MEASUREMENT METHODOLOGY

In this paper, we use detailed user activity data collected from a popular Facebook gifting application launched in June 2009, *iHeart*, which was operated by Manakki, LLC at the time of data collection. The dataset we studied was anonymized and donated by Manakki, LLC, which covers the major lifespan of the application (64 weeks since its initial launch until it declines in popularity) and contains over 2 billion gift exchanges among 190 million users. Method-



(a) Size distribution (b) Size box plot (c) Depth distribution

Figure 1: Comparing characteristics of cascades based on different parent types

ology to collect this data is presented in our earlier study by Nazir et al. [13]. However, by opposition to [13] that studies overall user interaction graph, here we focus specifically on analyzing the adoption process of the application. Towards this end, we track when a user installs and/or uses the application for the first time, and all the invitations or ARs received by the user prior to adopting the application. It is important to note that a not-yet recruited user can receive ARs from (and hence be influenced by) multiple friends before installing the application. This is a common feature across popular OSN-based gifting applications. This also differentiates our work from prior studies of epidemic or virus spread where a user is typically infected by the first contact with another infected individual. After a user adopts the application, he/she will in turn send out invitations to his/her friends, thereby recruiting more users and starting a cascade of adoptions. In the following, we describe how to extract information about such cascades from the empirical data and how to attribute influence to a parent node from a candidate set of senders.

3.1 Adoption Process as a Cascade

The inter-user communications through ARs in iHeart form multiple chain events, in which nodes cause connected nodes to be activated (i.e., adopting the application) with some probability. To study the adoption process, first we identify the seed nodes, defined as users who start using iHeart without receiving an AR from any other users. Other users (non-seed nodes) tend to adopt the application after receiving one or more ARs from their friend(s). Users are considered *activated* when they send the first gift to another user through iHeart; the time of this action is referred to as the *activation time*. We view such activations as a result of the *influence* of the original sender(s), who are considered *potential parents* of the user nodes. In general, users may send gifts to friends who already adopted the application or to users yet to be activated. In the former case, the invitations do not contribute to the adoption process, and hence such activities are dropped from our analysis. If a node sends multiple gifts to the same user, we consider the first gift-sending time as the *influence time* of that node.

As defined earlier, a cascade is a chain of adoption actions initiated by a seed node (through ARs). We represent a cascade as a tree (rooted at a seed node), where the children nodes are activated after receiving ARs from the parent nodes. These children nodes will in turn recruit next generation of nodes to adopt the iHeart application (hence growing the *depth* of the cascade tree). The *leaves* are the silent users who received one or more gifts but never sent further ARs to any users. In other words, any users except the leaves remain active for a time period and participate in the growth of the adoption of the application. The duration of the time between and including the first and last activity of any node is referred to as the *active lifespan* of the users. Active lifespan also represents the persistence of a user and governs how long they actively contribute to the growth of the application adoption process. In the next section, we will investigate how to attribute the influence on application adoption to one *parent* out of potential parents.

3.2 Attribution of Parent Nodes

Based on the definitions above, our analysis of the iHeart data reveals that there are 3.7 million seeds and 146.8 million leaves. Slightly less than 50% of the users are activated after receiving one invitation (AR); in this case, the parent nodes are clearly identified. The remaining users receive gifts from two or more friends before activation and hence the potential parents of the user is a set of senders from which the user received at least one gift before its activation. In most cases, the potential parent sets contain between two to 30 nodes, except for a few outliers that may receive up to 1000 invitations before activations. To simplify our analysis, we attempt to attribute the influence of adoption to only one parent node. For this purpose, we need to decide which of the senders is the effective/influential parent responsible for the activation of a particular user. We consider the following four parent types (a variation of the types used in the study by Bakshy et al [1]) that are defined based on AR reception order or senders' activity counts (note that these can be clearly determined from the empirical data):

- **First parent (FP):** User v is considered user u's first parent if u receives its first AR from v before u's activation, i.e. the first AR sender in the chronologically ordered potential parent set of u.
- **Last parent (LP):** User v is considered user u's last parent if v is the latest parent from which u received an AR before u's activation. In other words, LP is the last AR sender in the chronologically ordered potential parent set of u.
- **Highest out degree parent (HODP):** v is considered u's highest out degree parent if v has the highest out

degree among the users in potential parent set of u.

- **Random parent (RP):** v is considered u's random parent if v is chosen randomly from the potential parent set of u. This random choice is similar to distribute the influence among all the potential parents.

Based on the above heuristics, we collect the size and the depth of the cascades and their number of occurrences. Fig. 1(a) shows the distribution of cascade sizes for iHeart using different parent types. We notice the differences in the size distributions for FP, LP, HODP, and RP are insignificant. Moreover, Fig. 1(b) shows that FP, LP, and RP have very similar performance on cascade sizes, while HODP results into more positively skewed distribution of the cascade sizes than the remaining three types. Similarly, the distribution of cascade depths in Fig. 1(c) does not have much noticeable difference across different parent definitions except that LP leads to a slightly heavier tail. Based on these observations, we found that the different mechanisms to attribute influence to a single parent node does not seem to have any significant impact on the characteristics of the overall adoption process. Therefore, for the sake of brevity, we only show the results from analyzing cascade trees using FP heuristic to determine the single parent of active users.

4. EMPIRICAL RESULTS

In this section, we analyze the characteristics of 3.7 million adoption cascades (each associated with a distinct seed) observed in the iHeart data. In addition, we investigate factors that contribute to large cascades, and examine if there exist any distinctive features that mark important seed nodes or influential parent nodes. We only present results based on FP heuristic, but similar trends are also observed when considering other heuristics.

4.1 Characterizing Cascades of Application Adoption

According to our definition in Section 3, a cascade is initiated by a seed node. The smallest cascade size is one, which could happen when all the nodes that the seed send ARs to are already active (children of other parent nodes). Since the popularity of iHeart has declined towards the end of our data collection, we consider that to be the ending time of all cascades. Any cascade, big or small, causes a fraction of the overall user population (190 millions) to adopt the application and become active, which we refer as the *coverage* of the population. Fig. 1(a) shows that most of the cascade sizes we observed are tiny and account for only a small fraction of the overall population. In fact, 37.7% of the cascades are of size 1, 72.9% have sizes less than 10, and 95.5% have sizes less than 100. From Fig. 1(c), we observe that only a small fraction of cascades have tree depth greater than 2, but there are cascades that span more than 60 generations. The average cascade size and depth are 50.87 and 1.37, respectively. Fig. 2(a) shows the cumulative coverage of user population contributed by cascades ordered from the largest to the smallest (on x-axis). Note that the top 3% (about 10^5) out of the 3.7 million cascades account for the activation of the 80.0% of the overall population. We will focus on these large cascades and examine their growth patterns in the following section.

Characterizing the growth of large cascades: As described in Section 3.1, each cascade can be represented as a tree rooted at the seed node. Nodes activated by the seed node are represented as children nodes at depth one in the tree, and are considered the *first generation* of adopters influenced by the seed node. The tree expands both in depth and width as each of these children nodes in turn recruit other nodes, leading to the 2nd, 3rd, and in general, n-th generation of adopters. We define *shell size* as the population size of each generation, i.e., number of activated users at a particular tree depth. To study the growth of the application adoption process, we analyze how the shell size typically changes at each tree depth for large cascades. Fig. 2(b) shows the shell sizes associated at each tree depth of the top 10 cascades (ordered by size). At the beginning of a cascade, the shell size generally increases with each tree depth until it reaches the peak (can be between 5 to 20 hops away from the seed), and then it starts to decrease as the later generations become less and less effective in recruiting more nodes. This seems to indicate that the large number of early adopters has notable influence in expanding the cascades.

Fig. 2(b) suggests that the application adoption does not grow at a same pace for every generation/depth. Therefore, we plot the differential shell size growth (i.e., difference in shell size from depth n to $(n+1)$ in Fig. 2(c). We observe that there are four distinct phases of cascade growth process:

- *Accelerating positive growth:* The shell size keeps increasing with each depth before reaching a local maxima. This corresponds to increasing growth with average of more than one child per node. Obviously, the longer this phase lasts, the more it contributes to the overall adoption size.
- *Decelerating positive growth:* After reaching the local maxima, the growth (while still more than one child per node) starts decreasing. This leads to a slowdown of the overall adoption growth towards a more stable state.
- *Accelerating negative growth:* In this phase, the shell size shrinks at a faster and faster pace, with average growth of less than one child per user.
- *Decelerating negative growth:* The shell size is still shrinking but at a slower and slower pace. This also refers to an average growth of less than one child per user but the growth is becoming less negative.

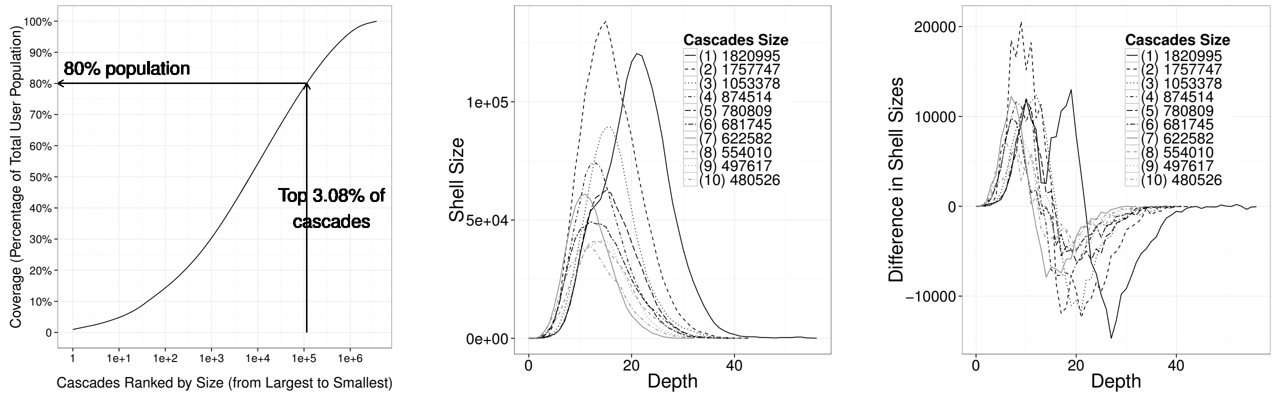
Most cascades we observed follow these four phases sequentially. However, a few of the big cascades exhibits two accelerating positive growth phases, e.g., cascade 1 represented by solid line in Fig. 2(c). This suggests that there exist other factors (e.g., potential existence of influential nodes) that can help further amplify the adoption cascades. We investigate the potential roles of important/influential seed and parents nodes in the next section.

4.2 Factors Contributing to Large Cascades

Identifying important or influential seed nodes is crucial for a variety of applications (e.g., marketing, political propaganda, etc.). In this section, we examine if there are any distinct features associated with seeds that succeed in generating the top 3% cascades, referred to as the *large-scale* cascades. We also investigate the impact of a few user characteristics that are typically observed in OSN applications.

4.2.1 Features of seeds in large cascades

Since the cascade is a chain of user activations initiated by the seed node, the depth of the cascade tree represents the *reach* of influence of a seed node, i.e., how many generations of adopters are recruited. The first question we explore is



(a) Coverage of top cascades (b) Growth of cascades shells (c) Differential shell size growth

Figure 2: Characterizing cascades of application adoption

whether seed nodes with further reach (or larger tree depth) generate larger cascade sizes. The other two features we examine are the active lifespan and the out-degree of the seed nodes, with the rationale that seeds that stay active longer or send invitations to large number of friends may recruit more of the users, hence leading to large cascades. Fig. 3 shows the scatter plot of the cascade sizes with respect to the three features of their seeds: reach (or maximum tree depth), active lifespan, and out-degree. We also compute the Pearson correlation coefficients between cascades sizes and three seed features for all and large-scale (top 3%) cascades. Results are summarized in Table 1. We conclude from the figure that cascade size is neither correlated with the active lifespan nor the out-degree of the seeds (correlation coefficient of 0.048 and 0.009, respectively). Although there is generally a positive trend where larger cascade sizes are associated with larger tree depths, a wide range of cascade sizes are found for the seeds at a certain reach (tree depth). The presence of this noise results in very low correlation coefficient (0.3). In summary, our empirical analysis shows that the features commonly believed to be the traits of effective seed nodes fail to show any positive correlation with large-scale application adoption.

Table 1: Pearson correlation coefficients between cascade sizes and various features of seeds

Cascades	Reach	Active Lifespan	Out Degree
All	0.2008214	0.03875115	0.006206429
Large-scale	0.3102366	0.04755791	0.009425999

4.2.2 Who are the influential nodes?

In this section, we seek to understand whether there exist important or influential individuals that are responsible for large cascades. Section 4.1 discussed that not all cascades are effective in propagating the application adoption (97% of the cascades have population size of less than 150). However, it is unclear whether the seed nodes of large-scale cascades are indeed the influential users that drive the cascade growth. In addition, the longer a user stays active in the application, the more likely he/she will recruit more users since they send out more invitations over a longer period of time. For this reason, they can potentially be "influential" in the growth of the adoption cascades. We analyze the active lifespans of parent nodes observed in our cascades and find that more than half of the parents remain active in the adoption process for more than 80 days. We also find that 95% of the overall active parents come back at least once after their first activity. Nevertheless, a huge number of parents have short lifetime (50% parents have active lifespans of less than 3 months), and these massive short-term users may have significant impact on the overall growth of an adoption (e.g., based on the popularity of the nodes). In fact, our previous study [6] showed that gifting applications tend to experience a higher churn of users, i.e., users remain active only for a short period of time. Hence, one cannot rule out the possibility that short-term users may also contribute to the growth of a popular application.

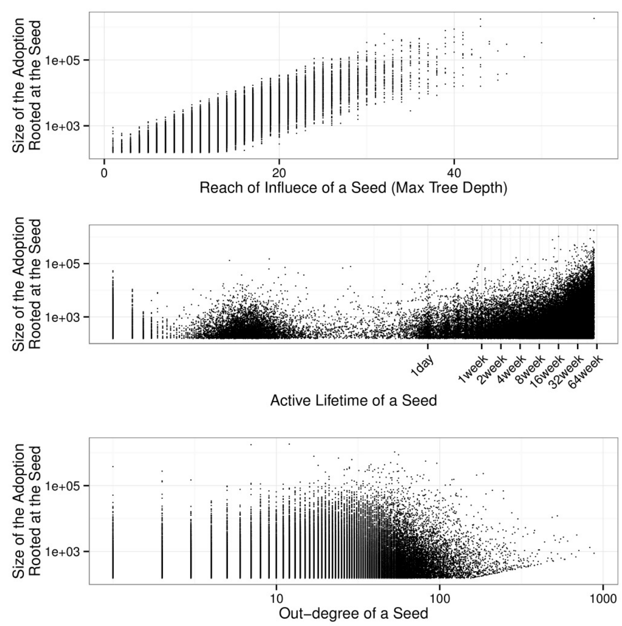
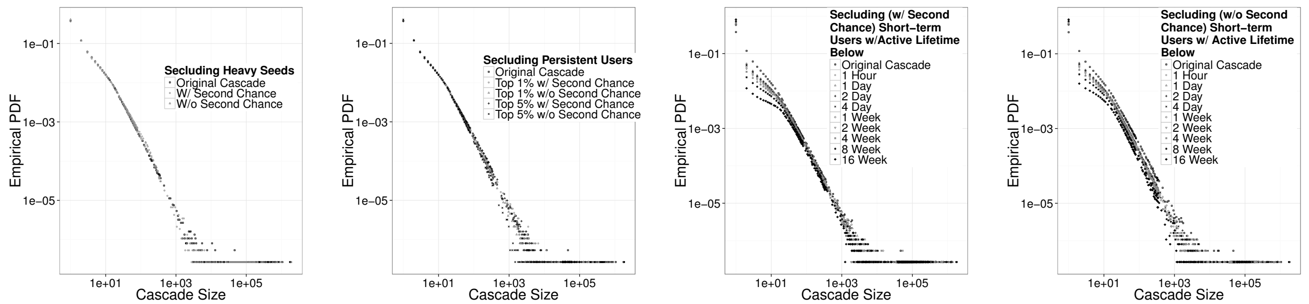


Figure 3: Cascade size vs. seed's features

Based on the discussions above, we design a series of "what-if" experiments to gauge the impact of the following three *potentially influential* group of users on the overall growth of the application adoption process:

- **Heavy seeds:** Seed nodes that are the root of large-scale cascades.
- **Persistent users:** Users whose active lifespan is above a set threshold.
- **Mass of short term users:** Active users who has short lifespans (below a chosen threshold).

In our experiments, we seclude each class of users (e.g.,



(a) Impact of secluding the heavy seeds

(b) Impact of secluding the persistent users

(c) Impact of secluding the mass of short-term users (w/ second chance)

(d) Impact of secluding the mass of short-term users (w/o second chance)

Figure 4: Impact on cascade size distribution due to seclusion of "important" nodes

heavy seeds) from their subsequent children nodes as a way to simulate their influence on the adoption process. Assume that H represents one of the three potentially influential group of users, and a node $h \in H$ is a candidate to be secluded. Since the children of h in the cascade tree may have received multiple ARs from other nodes, we consider two scenarios. In the first scenario, we assume there is a *second chance* for the children nodes to be activated even after h is secluded. In other words, if a node n has an alternative parent node p where $p \notin H$ and p is not a descendant of v ($\forall v \in H$), then n and its subsequent actions (e.g., activating other children nodes) are preserved. Note that since the individual cascade tree is rooted at a distinct seed node, if the alternate p is a descendant of a different seed node, then secluding h may lead to shrinking of one cascade, and expansion of another. In the second scenario, we remove all the sub-trees rooted at n without considering second chance.

Secluding the heavy seeds: First, we seek the effect of removing the heavy seeds that are the roots of large (top 3%) cascades. Secluding the heavy seeds without considering the second chance (i.e., without alternative parent nodes) essentially discards all the adoption trees rooted at them, and the seeds become the only member of their respective cascades. However, when we consider second chance, we do not see significant difference between original cascade size distribution and the distribution after secluding the heavy seeds in Fig. 4(a). This implies that the heaviness of the seeds hardly have any effect on the distribution of the adoption size.

Secluding the persistent users: We classify users with active lifespan above a set threshold as *persistent*. We rank the users based on their active lifespan and select two thresholds that correspond to identifying the top 1% and top 5% of the long-lived users. Fig. 4(b) shows the effect of secluding these highly persistent users from the adoption for the two scenarios: with and without second chance. For the former case, the distributions of cascade sizes before and after secluding the persistent users remain the same. Without second chance, removing persistent users does not affect small-scale cascades, but does reduce the size of large-scale cascades significantly (up to 1-2 orders of magnitude).

Secluding the mass of short-term users: Users with active lifespans shorter than a threshold τ are called short-term users. We consider different thresholds τ (varying from one hour to 16 weeks) and study the impact of secluding short-term users by removing all users with active lifespans less than τ. Fig. 4(c) shows the result when we consider second chance, and Fig. 4(d) shows the result without considering second chance. In both cases, we compare the simulated distributions of cascade sizes with the original distribution obtained using FP (as seen in Fig. 1(a)). When τ increases, a larger fraction of short-term users are removed, and one could expect a complete change in the distribution. However, in Fig. 4(c), we do not notice any significantly differences. Instead, secluding a massive number of short-term users mostly affects the small-scale cascades. Large-scale cascades are found to be more resilient against the seclusion of numerous short-term users (the slopes of the distributions remains the same for big cascades). Similarly, even when second chance is not considered, the slopes of the distributions for large-scale cascades do not change, as shown in Fig. 4(d). However, due to the impact of the loss of numerous short-term users (by not considering the second chance), the distributions of small-scale cascades shift further down.

In summary, none of these three potentially influential groups have any significant impact on the overall adoption cascade sizes. We attribute this to the fact that a user may receive invitations from a diverse group of senders (and be influenced by any one of them) in a typical OSN-based gifting application. Results shown in Fig. 2, 3, and 4 are based on the cascades using FP heuristic. Similar trends are observed considering the heuristics using LP, HODP, and RP. This implies that attributing influence to a specific parent node has little to no effect on the overall adoption process.

5. CONCLUSION

In this paper, we analyze user activity data from a popular Facebook gifting application and characterize the invitation-based (through ARs) adoption process. We consider a chain of adoptions initiated by a seed node as a cascade. We find that most cascades are tiny (less than 156 users), and the top 3% of the cascades account for 80% of the population. Contrary to popular belief, our empirical study of large-scale cascades shows that they are not influenced by seed nodes with large out-degree or long active lifespan. The fact that a user can receive multiple invitations from different senders makes the adoption process resilient to the removal of the most active and persistent parent nodes. Although we studied the adoption process of a popular Facebook gifting application, we plan to evaluate the results shown here with a future study on less or more attractive gifting applications.

6. ACKNOWLEDGMENTS

This work is supported in part by National Science Foundation CNS-1302691 grant.

7. REFERENCES

[1] E. Bakshy, J. M. Hofman, W. A. Mason, and D. J. Watts. Everyone's an influencer: quantifying influence on twitter. In *Proceedings of the fourth ACM international conference on Web search and data mining*, WSDM '11, pages 65–74, New York, NY, USA, 2011. ACM.

[2] E. Bakshy, B. Karrer, and L. A. Adamic. Social influence and the diffusion of user-created content. In *Proceedings of the 10th ACM conference on Electronic commerce*, EC '09, pages 325–334, New York, NY, USA, 2009. ACM.

[3] F. M. Bass. A New Product Growth for Model Consumer Durables. *Management Science*, 50(12 supplement):1825–1832, Dec. 2004.

[4] J. Berry and E. Keller. The influentials: One american in ten tells the other nine how to vote, where to eat, and what to buy. Free Press, January 2003.

[5] M. Cha, A. Mislove, and K. P. Gummadi. A measurement-driven analysis of information propagation in the flickr social network. In *Proceedings of the 18th international conference on World wide web*, WWW '09, pages 721–730, New York, NY, USA, 2009. ACM.

[6] B. Estrada, A. Nazir, L. X. Liu, C.-N. Chuah, and B. Krishnamurthy. Evolution of user activity with time on third-party facebook applications. Technical Report ECE-CE-2011-5, Department of Electrical and Computer Engineering, University of California, Davis, 2011. http://www.ece.ucdavis.edu/cerl/techreports/2011-5/.

[7] S. Goel, D. J. Watts, and D. G. Goldstein. The structure of online diffusion networks. In B. Faltings, K. Leyton-Brown, and P. Ipeirotis, editors, *ACM Conference on Electronic Commerce*, pages 623–638. ACM, 2012.

[8] D. Kempe, J. Kleinberg, and E. Tardos. Maximizing the spread of influence through a social network. In *KDD '03: Proceedings of the ninth ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 137–146. ACM Press, 2003.

[9] M. Kitsak, L. K. Gallos, S. Havlin, F. Liljerosand, L. Muchnik, H. E. Stanley, and H. A. Makse. Identification of influential spreaders in complex networks. *Nature Physics*, 2010.

[10] B. Krishnamurthy and C. E. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the 2nd ACM workshop on Online social networks*, WOSN '09, pages 7–12, New York, NY, USA, 2009. ACM.

[11] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. *ACM Trans. Web*, 1(1), May 2007.

[12] D. Liben-Nowell and J. Kleinberg. Tracing information flow on a global scale using Internet chain-letter data. *Proceedings of the National Academy of Sciences*, 105(12):4633–4638, Mar. 2008.

[13] A. Nazir, A. Waagen, V. S. Vijayaraghavan, C.-N. Chuah, R. M. D'Souza, and B. Krishnamurthy. Beyond friendship: modeling user activity graphs on social network-based gifting applications. In *Proceedings of the 2012 ACM conference on Internet measurement conference*, IMC '12, pages 467–480, New York, NY, USA, 2012. ACM.

[14] T. Rodrigues, F. Benevenuto, M. Cha, K. Gummadi, and V. Almeida. On word-of-mouth based discovery of the web. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference*, IMC '11, pages 381–396, New York, NY, USA, 2011. ACM.

[15] T. W. Valente and R. L. Davis. Accelerating the Diffusion of Innovations Using Opinion Leaders. *The ANNALS of the American Academy of Political and Social Science*, 566(1):55–67, Nov. 1999.

[16] D. J. Watts. A simple model of global cascades on random networks. *Proceedings of the National Academy of Sciences*, 99(9):5766–5771, Apr. 2002.

[17] D. J. Watts and P. S. Dodds. Influentials, networks, and public opinion formation. *Journal of Consumer Research*, 34:441–458, 2007.

[18] D. J. Watts and P. S. Dodds. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441–458, Dec. 2007.

[19] G. Weimann. The infuentials: People who influence people. State University of New York Press, 2003.