Social Influence and the Diffusion of User-Created Content

Eytan Bakshy
University of Michigan
School of Information
Ann Arbor, MI
ebakshy@umich.edu

Brian Karrer
University of Michigan
Department of Physics
Santa Fe Institute
Santa Fe, NM
karrerb@umich.edu

Lada A. Adamic
University of Michigan
School of Information
Center for the Study of
Complex Systems
Ann Arbor, MI
ladamic@umich.edu

ABSTRACT

Social influence determines to a large extent what we adopt and when we adopt it. This is just as true in the digital domain as it is in real life, and has become of increasing importance due to the deluge of user-created content on the Internet. In this paper, we present an empirical study of user-to-user content transfer occurring in the context of a time-evolving social network in Second Life, a massively multiplayer virtual world.

We identify and model social influence based on the change in adoption rate following the actions of one's friends and find that the social network plays a significant role in the adoption of content. Adoption rates quicken as the number of friends adopting increases and this effect varies with the connectivity of a particular user. We further find that sharing among friends occurs more rapidly than sharing among strangers, but that content that diffuses primarily through social influence tends to have a more limited audience. Finally, we examine the role of individuals, finding that some play a more active role in distributing content than others, but that these influencers are distinct from the early adopters.

Categories and Subject Descriptors

J.4 [**Computer Applications**]: Social and Behavioral Sciences – Sociology; H.2.8 [**Database Applications**]: Data Mining

General Terms

Measurement, Economics, Human Factors

Keywords

social influence, diffusion of innovations, virtual worlds, social networks

EC'09, July 6–10, 2009, Stanford, California, USA.

1. INTRODUCTION

In the digital age, the creation and distribution of digital goods has been democratized. On YouTube, users view millions of videos created by millions of users, on Flickr users upload their own photos and view others', and news are reported on, consumed, and commented on by a distributed network of bloggers and media sources. Perhaps the purest example of a market for user-generated content is that of the virtual world Second Life. The vast majority of the content, in fact pretty much all of the virtual world itself, from buildings to objects to fashion, is created, distributed, and consumed by the users themselves.

The unique property of studying social contagion in Second Life is that one can observe not just adoption in the context of an explicit social network, but also trace direct transfers of user-contributed content owned by users, which we will refer to as *assets*. In Second Life, you can search for interesting places to visit on your own, or a friend or business can give you a landmark – a bookmark that allows you to teleport directly to a location. If upon arriving, you would like your avatar to dance, wave, or make a certain sound, you need to retrieve that *gesture* from your inventory of assets. That gesture may have been given to you by a friend, or you may have purchased it from a store. Such transfer of assets and information presents a unique opportunity to compare diffusion via word-of-mouth to adoption resulting from broadcasts. Depending on the intellectual property rules attached to each object, some assets can be freely copied and shared; one Second Life user can pass on a gesture, hairstyle, or article of clothing to another.

The paper proceeds as follows. After reviewing related work and motivating our approach in Section 1.1, in Section 1.2 we describe the Second Life data set and the characteristics of information diffusion among Second Life users. In Section 1.3 we quantify the properties of asset transfer cascades and their relationship to the social network. We find that assets that are passed from friend to friend tend to produce deeper cascades, but the overall popularity of the asset is lower. We demonstrate that this insight can be used to predict how many additional individuals will adopt an asset over a period of time. Section 2 models the rate of adoption which we find to strongly depend upon the number of adopting friends a user has at any given time. As might be expected, when users have no previously adopting friends, their rate of adoption is related to the popularity of the asset in the population overall. However, once a friend has adopted, the adoption rate increases significantly, especially for less popular, niche assets. In Section 3 we identify two kinds of individuals, influencers who directly influence many of their friends to adopt, and early adopters. We find that early adopters are more likely to adopt without having to first observe their friends, but that they are not necessarily influential in subsequent adoptions. Section 4 concludes and discusses future directions.

1.1 Background & Motivation

The context in which our study occurs, Second Life [23] has been of interest due to the emergence of self-contained economy [22, 4] rich with social conventions [31, 9] that mimic aspects of real-world human and social dynamics. Our study provides a complementary perspective on how individuals influence one another, while contributing to a larger body of work in the measurement of large-scale social phenomena relating to the dynamics of content consumption in online communities.

In the marketing science literature there is a wealth of macro-scale studies of new product diffusion [19]. For example, the Bass model [3] is a differential equation model that predicts adoption based on relative populations of "innovators" that are not influenced by the decisions of others and "imitators" whose adoption depends of the total number of adoptions in the system. Extensions to these models have traditionally not taken into account social structure, nor the individual decision making processes of the adopters. Micro-level studies, such as [6], do model factors that influence the adoption of a product, but have only been studied in the context of small laboratory experiments.

Although the theory of information diffusion in social networks was developed decades ago [25], social contagion has only recently been measurable on a large scale through the digital traces that modern communication leaves behind. Social contagion can be distinguished from viral, unintentional sharing of e.g. human [24, 20] or electronic [21] malaises over networks. One feature of social contagion is that there may be thresholds to infection, with many individuals waiting for several of their friends to adopt before taking the plunge themselves [5]. Unlike disease spread, this diffusion typically has the property that an individual decides whether to accept the contagious object.

The availability of large scale social network data has lead to a number of studies quantifying various aspects of social contagion. Of interest in all these studies is how one might maximize the spread of influence through a social network by selecting a subset of influential individuals to initially infect with an idea or product [11]. Another possibility is to find out early what assets are "hot" by monitoring a subset of individuals that are likely early adopters of popular assets [16]. Although some have modeled adoption simply as a function of observing strangers' actions [26, 30], principally, these studies measured the likelihood that an individual takes an action as a result of their friends' choosing the same action.

Social network information has successfully been used, for example, to predict whether a customer will sign up for a new calling plan once one of their phone contacts does the same [10]. The photos we view and the stories we "Digg" are often the ones we observed our friends consuming [13, 14]. LiveJournal bloggers are more likely to join a group that many of their friends joined, especially if those friends belong to the same clique [2]. Blogs are likely to link to content that other blogs have linked to [27]. The insight that individuals tend to like (or like to have) the same things that their friends like can be used to improve collaborative filtering algorithms [32].

While social network information has been commonly utilized, there are relatively few studies that have included direct transfers between users. A study of person-to-person book and video recommendations found conditions under which such recommendations are successful [15, 17]. A study of online chain letters discovered that as messages diffuse through individuals' email contact networks, they form cascades that are far deeper than one would expect at random [18]. However, information cascades spreading through email were not studied in the context of an explicit social network that would allow one to measure both direct or indirect influence simultaneously.

In contrast to prior work, in this paper we are able to analyze social influence not just indirectly through separate information about the social network and user adoption, but also by accounting for direct transfer of assets between individuals. The direct transfers allow us to more precisely identify influencers who are responsible for a disproportionate fraction of the asset adoptions. Furthermore, we develop a simple model of adoption rates, as opposed to probabilities, that can incorporate information about the evolving social network without needing to make arbitrary decisions about how to subdivide time intervals. This model allows us to clearly illustrate the importance of network effects in the adoption of content.

1.2 Description of Data

The data set in this paper includes time-stamped content ownership data and weekly snapshots of the complete social network over a 130 day period between September 1, 2008 and January 16, 2009, with the exception of the weeks of September 19th and November 14th. We do not have the exact time stamps of when the friendship ties were formed or dissolved, but by using weekly snapshots, we can approximate the coarse evolution of the social graph. At the user level, we have information on when the user first joined Second Life and how many hours they have played. The data also includes the social network of users. The data were provided directly by Linden Lab, the maker of Second Life and no personally identifying information of Second Life's users was shared with the authors.

The social network we observe is made explicit by the users themselves, who add one another as "friends". By default, friends are aware of when their other friends are in Second Life, and if they grant additional permissions, those friends can see where in the virtual world they are located. Some users will even grant each other permission to modify each others' objects. This tends to occur among a small group of users for the purpose of collaboratively creating content. In other cases, users may not grant one another any permissions. Friend permissions do not necessarily need to be reciprocal.

As in many online social networks, the meaning of a friendship tie is somewhat ambiguous and can denote anything from casual acquaintanceship to a close relationship. One user may add another as a friend because they met in Second Life and wished to continue interacting. Or a Second Life friendship may reflect a "first life" relationship that has been carried into the virtual realm. While privacy preferences can vary from user-to-user, we consider the user's "social network" to consist of all friendship linkages that have, at a minimum, reciprocated permissions to see one another's online status. Throughout the paper, we will refer to two users connected in this fashion as "friends" or "neighbors" in the social graph.

Since the subject of the work is on the diffusion of user-created content, we focus on studying content that is freely available, non-trivial to produce, and widely distributed amongst users. Content that can be carried around by a user is called an asset and is stored in the user's inventory. We chose to study *gestures*: transferable animations that allow a user's avatar to carry out programmed physical movements or make sounds. The choice of this type of asset was made because gestures are discrete and simple to trace. In our analysis, we use Linden Lab's definition of an active user: those we have logged in in the 60 days prior to the last observation date (Jan. 2009) and have used Second Life for more than six hours. In addition we focus on the users who have exchanged at least one object with another user between September 2008 and January 2009. We chose gesture assets that had at least sixteen unique owners and were never directly distributed to users by Linden Lab. The former exclusion rule omitted gestures that had not diffused, and the latter excludes gestures the users may have received without opting to. With these restrictions, our sample population contains 100,229 users and 106,499 assets. Because of the long-tail of asset popularity, this represents only a small fraction of the unique 5,327,671 gestures.

Most assets in our data set are owned by a relatively small number of users, and very large assets of size 1,000 or greater make up less than 10% of all assets. This is the familiar long tail, shown in Figure 1, of content popularity; a few gestures are widely adopted by users, while the majority remain of little or niche interest. Interestingly, none of the content has saturated the user population, with the largest assets owned by roughly 10% of the population.

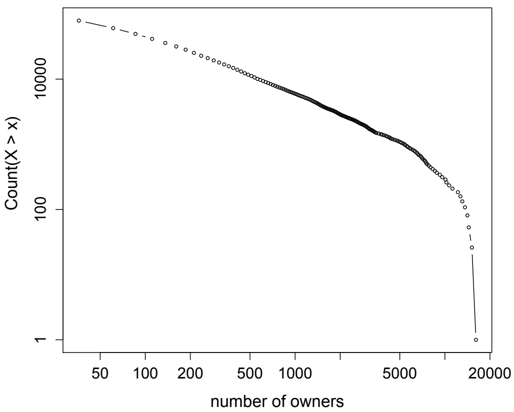

Figure 1: Cumulative distribution of the number of unique owners per asset in our sample population.

The content ownership data comes in the form of asset transfers, that contain the asset, previous owner, next owner, and time-stamp. It indicates that the previous user had given a copy of the asset to the next user. There are a total of 12,585,298 asset transfers over the observation period, 3,409,630 (23%) of which have accurate information about the previous owner. On average, approximately 43% of the observations in each asset have previous owner information. The average is higher than the total percentage because for larger assets there are more observations without previous owner information. Information can be lost, for example, when a user copies or moves assets in their inventory. The extent to which individual assets are missing previous ownership information does not appear to vary systematically with the owner's experience level, their connectedness to others, or how many gestures they own.

The transfers of each asset can be visualized as a cascade forest, with edges drawn between each owner and the previous owner, showing an "infection" path that represents the direct flow of content between users. Where previous owner information is missing, we start a new tree in the forest. Figure 2 shows a cascade forest for one particular gesture. We note a fanning pattern, with some users transferring the gesture to many others.

Of the assets transfers for which we have accurate previous ownership information, 1,754,852 (approx. 48%) of the transfers occurred between friends. This suggests that direct social influence over the social network plays a considerable role in the distribution of content. In addition to direct influence, we find that indirect influence along the social network also plays a large role in adoption. Of those transfers that did not occur between friends, 678,908 (approx. 38%) of the users who had acquired a new asset did so after at least one of their friends had also adopted.

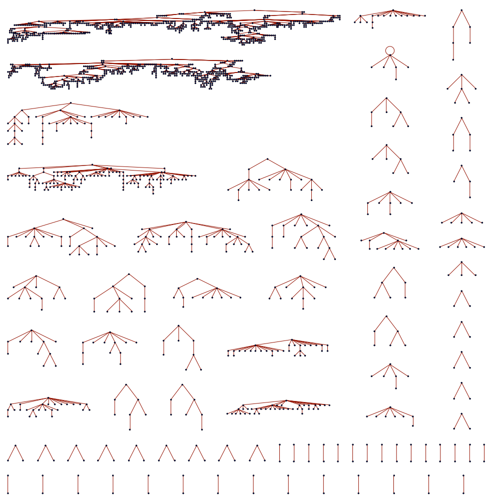

Figure 2: Example of a cascade forest for the Aerosmith(916) gesture. Edges denote transfers of the gesture between users.

1.3 Friend-to-friend vs. one-to-many

Given the above observations on the role of the social graph in the transfer and adoption of content, an important question a viral marketer may wish to answer is how much of a boost one can expect from having customers themselves advertise to one another and distribute the assets [8]. Previous work on book and DVD recommendations found that viral marketing is more effective for niche products as opposed to widely popular ones [15]. We find a similar trend here.

In order to quantify between-user transfers, we look at the following variables for each asset: the total number of adopters for the asset (the asset size or popularity), the percentage of the transfers that were between friends (% direct), and the percentage of transfers that resulted in subsequent transfers by the adopting user (% non-leaf). We find the percentage of non-leaf nodes, which can be thought of as a measure of cascade depth, to be correlated with the percentage of the adoptions that can be accounted for by the social graph ($\rho = 0.42$), indicating that the diffusion along the social network produces deeper cascades for which a higher proportion of users actively participate in the transfer of the asset. But while these cascades tend to be deeper, they are not wider. The average popularity of the asset falls as the proportion of non-leaf nodes and social influence increases. As Figure 3 shows, having more adopters actively transferring assets is actually indicative of the asset not being broadly popular.

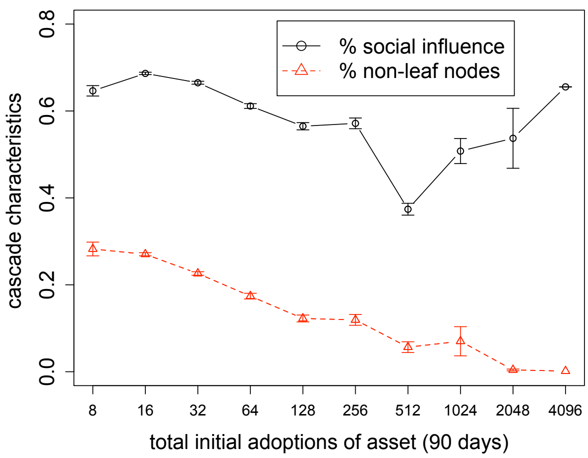

Figure 3: Percentage of non-leaf nodes vs. asset size for assets over the first 90 days of their spread.

One can use the above observation of asset size and the role of social influence to predict the growth in the number of adoptions for a particular asset. We differentiate social influence (having a friend adopt before you do), and direct influence (obtaining an asset from a friend). Not all assets can be obtained from a friend, even if the friend has said asset, because of copy permissions. We therefore separate the assets where no transfers occur between friends (these likely cannot be copied), and ones that do.

We observe the number of adoptions in the first 30 days since the asset is created. We then run a regression to model the number of adoptions in the following 60 days. Besides the initial number of adoptions, we also included the following statistics from the first 30 days: whether the adoption occurred after at least one other friend adopted (% social), the percent of adoptions that are direct transfers along the social network, and the percent of adoptions occurring directly through the social network that resulted in further adoptions. Just two variables yielded the greatest explanatory power: the number of initial adoptions, and the percentage of initial adoptions that can be explained by the social network. We further find that using those same two variables, assets that are transfered from friend-to-friend at least sometimes are more predictable than those that are never passed between friends. A possible reason is that if friends are unable to share assets due to copying restrictions, then the distribution falls on a limited set of individuals, making the sharing of the assets more variable. Although information diffusing through a social network may lead to unpredictable cascades [28], in this case being able to observe such diffusion actually makes the cascade more predictable.

Table 1: Regressing the subsequent number of adoptions on the initial adoptions and percentage that can be explained by social influence. *d* is the restricted set of assets that were observed to have been transfered on the social network.

	all assets	all assets	*d*	*d*
log(initial size)	0.362	0.388	0.508	0.476
% social		-0.808		-0.897
R^2	0.112	0.161	0.164	0.196

As Table 1 shows, unsurprisingly, a higher initial rate of spread translates to a higher number of subsequent adoptions. What is interesting is that the percentage of social adoptions (those that can be easily attributed to friends' adoptions) is *negatively* correlated with the the number of additional users who adopt. This suggests that assets that are diffusing through the social network may be of interest to a smaller subset of individuals. Because of homophily, the tendency of like to associate with like, these individuals are more likely to be friends with one another. So while a niche product may be shared more readily through the social network because the social network reflects niche tastes, the product does not have a wide susceptible audience, and therefore will not be adopted as widely.

While the regression suggests that the overall rate of spread through the social network is slower than through alternate paths, we find individual transfers to be more rapid between friends. Figure 4 shows the distribution of lags between when an individual becomes infected and when they infect either a friend, vs. when they infect a non-friend. First, we note that individuals are most likely to share a gesture within a short time of receiving it, while the context and novelty of the asset are still fresh.

Furthermore, we find that friends will more rapidly share with one another than with strangers: the average time lag between when a user acquires an asset and when they give it to a friend is 53.1 days, compared to the 75.6 it takes them to transfer it to a non-friend. The average time lag between one friend adopting after another (without sharing the asset with one another directly) is 105.2 days, compared to 228.3 days for adopters who are not friends. Although there is a mild cohort effect (with friends being more likely to join Second Life around the same time), it alone would not explain why friends are adopting so closely in time. That there is variation in speed depending on the relationship type is of interest because the speed of a interpersonal link can dramatically effect the fastest route information will take as it spreads, to the point where some slower links play little role at all [12]. It is therefore of interest to model the *rate* of adoption following a friend's adoption, and this is what we undertake in the next section.

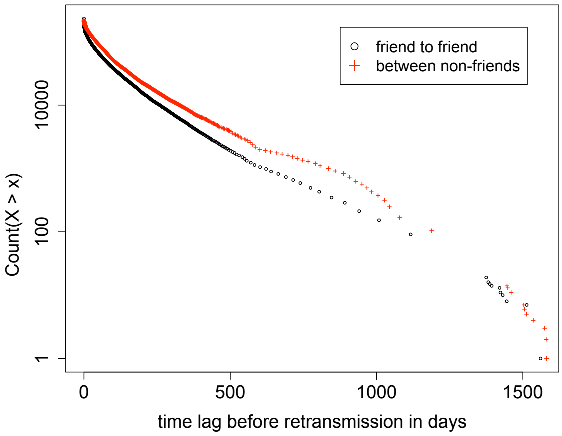

Figure 4: Delays between a users' adoption and retransmission times, for assets with 100-200 adopters.

2. MODELING ADOPTION

As a Second Life user observes others adopting particular assets, she may not only be more likely to adopt the asset herself, but the rate at which she does so may quicken as she observes more and more of her friends adopting. In order to characterize social network effects on user adoption in Second Life, we utilize a simple model of users' adoption rates. We show how with slightly different assumptions the same model can be applied to adoption rates both at the asset and at the user levels. Our results are compared to a Cox proportional hazards model with time-varying covariates that incorporates other possible influences such as the total number of adopters in the user population. We show that the estimates produced by the Cox model are consistent with the results of our simple model.

2.1 Formulation

One way in which this neighbor influence has been measured before is by computing the probability of adoption as a function of the number of neighbors who have already adopted in some time interval [2]. To be more precise, one counts the number of individuals who have not adopted that have k neighbors who have adopted at the beginning of the time interval and then compute the fraction of these individuals who have adopted at the end of the time interval.

An improved and related approach, used by [1], considers the probability of adoption within many identical discrete time intervals, rather than just one. Our approach presents a further refinement by utilizing a continuous time model of adoption where we have stochastic rates of adoption rather than probabilities of adoption. We consider rates of adoption from two perspectives: at the level of adopting a particular asset and at the level of the user. In the former case, we assume that the rates of adoption are characteristic of a particular asset, are fixed in time, and the same for all users. These assumptions are analogous to the assumptions used in [1] and [2]. At the level of the individual user, we assume that a particular user's rates are fixed in time and equivalent for all assets, but that they differ from user to user.

We first explain the model formulation from the perspective of a particular asset computed over the entire population of users. A user enters into state k at the moment that their k_{th} friend adopts the particular asset. The model assumes that once an individual is in state k, the time until they adopt, T_k, is exponentially distributed, i.e. they draw an exponentially distributed random variable T_k with mean $1/\lambda_k$ where λ_k will be referred to as the adoption rate for state k. If an avatar's state changes before they reach their adoption time, they discard that time and draw a new time from the next exponential distribution corresponding to their new state. There are three ways in which a user can exit state k. If one of their existing neighbor adopts or they become friends with someone who has already adopted (adding an edge in the social network), they advance to state $k+1$. If they end a friendship with an adopter (deleting an edge in the social network), they return to state $k-1$.

We use maximum-likelihood to estimate λ_k from the available data for each asset. To do this, we have to compute the probability of observing the data given the model. Let t_k^i be the total amount of time the ith user spent in the k state and θ_i be one if the user adopted by the end of our observation period or zero if the avatar did not adopt. For the users that did adopt an asset, let a_i be the state from which that avatar adopted. Then the probability (density) of the data given the model is

$$\prod_i \lambda_{a_i}^{\theta_i} exp(-\sum_k \lambda_k t_k^i). \tag{1}$$

We can further simplify the probability of the data given the model by defining A_k to be the number of individuals who adopted from state k and $M_k = \sum_i t_k^i$ to be the total amount of time spent in state k over all individuals. Then

$$\prod_k \lambda_k^{A_k} exp(-\lambda_k M_k). \tag{2}$$

Maximizing with respect to the model parameters yields

$$\lambda_k = A_k/M_k, \tag{3}$$

as the maximum-likelihood estimate of the rates, assuming a uniform prior over the model.

We make a further distinction based on the population of measurements used to calculate the characteristic rates in our model. For a particular asset, it's unclear whether the entire population of users should be included in the calculation. The reason for not including all users is that some individuals may never want to acquire the asset regardless of the number of their neighbors that adopt. Including all users for each asset is what has been done previously, which carries the assumption that all individuals considered will adopt if one waits a sufficiently long time. However, a user may never want to adopt, no matter how long they have been exposed to it. For example, Aerosmith gestures may be a taste that a particular user will never acquire. Rather, individuals are selective in their adoptions, and will resist both advertising and social influence if an asset does not match their tastes or interests. Therefore, our alternative approach is to estimate the rates only using measurements from the observed user population that has adopted the asset. We can be sure that this population wants the asset, but of course, there may be other individuals who want the asset but have just not acquired it yet.

Since there are advantages and disadvantages in including the non-adopting population in our measurements, we report our results for both specifications, referring to the respective calculations as utilizing the entire population and the adopting population of users. We note that our population of all users is still restricted to users who have adopted at least one asset during the time period, which means that all users were susceptible to adopting in general. To specify to the adopting population only, we follow the above derivation only including users that were observed to adopt the asset. This adjustment again leads to Eq. 3, where now M_k is the total amount of time spent in state k over individuals that adopted the asset.

As we mentioned above, one can model many users adopting the same asset, or one can model a particular user as they adopt different assets. Calculating adoption rates for a particular user over the entire population of assets is also simple. We again use maximum-likelihood to estimate λ_k for each individual using every asset. Let t_k^i be the total amount of time a user spent in the k state for the ith asset, θ_i be one or zero if the avatar adopted or did not adopt the ith asset respectively, and a_i be the state from which that user adopted the ith asset. Then the probability (density) of the data for that individual given the model is again Eq. 1. Defining A_k to be the number of assets adopted from state k and $M_k = \sum_i t_k^i$ to be the total amount of time the individual spent in state k over all assets, and then maximizing with respect to model parameters leads to Eq. 3. As in the analysis for particular assets, we also can decide to only include assets that the user was observed to acquire. This specification results in M_k being the amount of time that an individual has spent in state k over all assets that they were observed to adopt. Again, we report our results for both cases for each user, which we refer to as either utilizing the entire population and the adopted population of assets.

We take into account all adoptions that occur before Sept 1, 2008, when we first started receiving weekly social network snapshots. After this point, the adoption times the t_k^i are used to estimate the rate parameters, since our network data moving forward in time is more accurate.

2.2 Analysis

We first report on the differences in adoption rates as a function of the number of adopting neighbors for small and large assets separately. Asset size denotes simply the total number of adopting users for the asset. We also consider the trends across all assets, and "new" assets that appeared after Sept. 1, 2008. Examining new assets helps us avoid confounds such as large assets being in the later stages of their adoption curve. Figure 5 shows that adoption rates increase with the number of previously adopting neighbors a user has, whether one considers all users or just the adopters, and whether one includes all assets or just newer ones. When one considers all users, the rate increase is initially convex, suggesting that having two, rather than just one adopting friend increases the likelihood that a user will adopt at all. This is in agreement with previous analyses [15, 2, 1], which found that the probability initially increases steeply with k but then shows diminishing returns as k increases further.

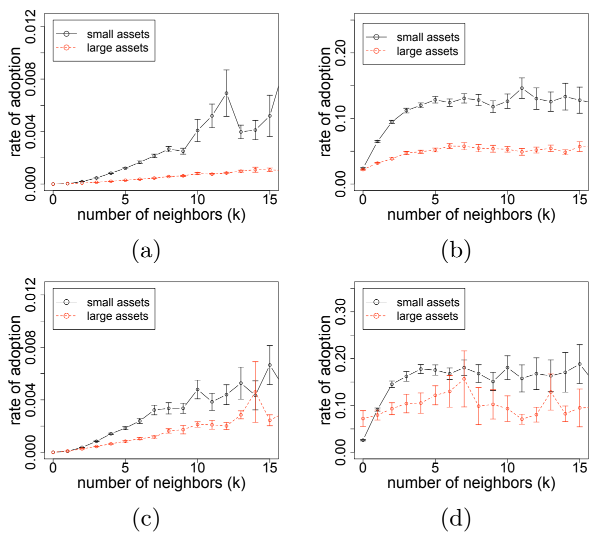

Figure 5: The average rate of adoption of assets as a function of adopting neighbors, k. The black curve corresponds to assets that are owned by 50-500 users, and the red curve corresponds to assets owned by 500 or more users. (a) entire population, all assets (b) adopting population, all assets (c) entire population, new assets (d) adopting population, new assets. The rates are in units of inverse days.

Once we consider the population of just the adopting users, the rates do not show as steep of an initial gain as they did for all users. This is because now the rates do not reflect a binary outcome of whether or not the user adopts at all, but rather how much more quickly a susceptible user adopts following the adoption of multiple neighbors. For smaller assets that have between 50 and 500 adopters, the rate doubles between having no adopting neighbor to having one, with the increase more pronounced for new assets. It then increases roughly another 60% when a second neighbor adopts.

What is most striking, however, is that this rate of adoption as a function of the number of neighbors increases more rapidly for smaller assets. These plots confirm our intuition from Section 1.3 concerning the relationship between relative popularity and channels of influence. The increase in rate appears most strong for more niche items, whereas neighborhood effects appear to play less of the role for more popular assets. This suggests that what is driving the adoption of more popular assets must lie at least partly outside of the social network. For large assets, those with ≥ 500 adopters, λ_0 is 4.73 times higher than for smaller assets. For newer assets, this ratio is 7.43. Because collectively users spend much more time in the $k = 0$ state (having no adopting neighbors) than in the $k > 0$ states, a small difference in λ_0 can lead to significant differences in asset size. For example, across assets with between 50 and 500 adopters, the total length of time spent by all users in the $k = 0$ state is a factor of 190 times greater than the total length of time spent with at least one adopting neighbor. We obtained qualitatively similar results when we varied the cutoff between small and large assets.

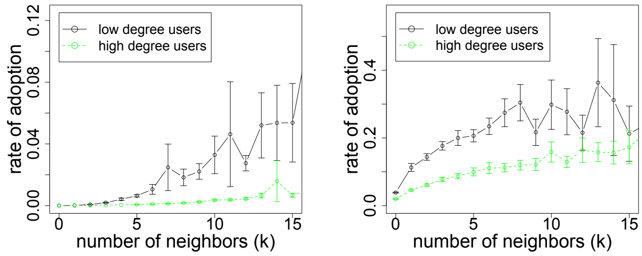

Figure 6: The average rate of adoption for users as function of adopting neighbors, k. The black curve corresponds to users of low degree that have 15-100 friends, and the green curve corresponds to users with 100-1000 friends. Left: entire population of assets. Right: adopted population of assets. The rates are in units of inverse days.

Table 2: Cox proportional hazards model with time-varying covariates. All estimates have p < 0.001.

parameter	estimate	error
mean degree	-0.00134	0.00009
log(assets owned)	0.03391	0.00601
cohort	0.62933	0.01176
log(usage)	-0.18349	0.00470
adopting neighbors	0.32795	0.00902
log(adopting users)	-0.04634	0.00754

We next turn to an analysis of user-characteristic adoption rates. We average the data over all individuals, where we divide the data into high and low degree, as shown in Figure 6. Interestingly, the users with high degree tend to adopt at comparably lower rates than their lower degree counterparts. This suggests that individuals may accumulate many friends, especially in online contexts, but consequently any individual friend holds less influence.

We can further understand the above results by turning to a more general approach to adoption rates using a Cox proportional hazards model with time-varying covariates [7]. This model allows us to include additional fixed covariates such as average degree of the user observed over the time period, the number of assets owned, the user's cohort (from 0 to 5 years, 5 being the most recent), and usage (in days). We also utilized the number of adopting neighbors and number of adopting users for each asset as time-varying covariates. Results from the regression are shown in Table 2.

As in the previous model, we find that the number of adopting neighbors has a significant and positive effect. We also see that high average degree does indeed have a negative effect on the adoption rate. By itself, the overall popularity of an asset does increase the rate of adoption, as suggested in Figure 5(d). In combination with the other factors, however, overall popularity has a weakly negative effect in the rate of adoption. Finally, we see that users that have signed up recently tend to adopt friend's content more rapidly, and that this effect decreases with experience. The results indicate substantial heterogeneity in user behavior, which we further investigate in the next section where we look for influential users and early adopters.

3. INFLUENCERS AND EARLY ADOPTERS

Thus far we have observed social influence from the point of view of the adopter – finding that the rate of adoption increases as one observes more and more friends adopting. This suggests that each friend holds some influence, and that having more adopters among one's friends increases the "hazard" that one will catch the bug and adopt as well. But one may also pose the question of whether all adopters are equally contagious to their friends. More specifically, using data on user-to-user asset transfers among friends, we can examine whether a few individuals are responsible for distributing assets.

3.1 Concentration of influence

First, we look at the distributions of transfers per individual, shown in Figure 7. The distributions are heavy tailed, indicating that a majority of individuals play a negligible to small role in distributing assets, while a handful of users disproportionately contribute to the dissemination of content. Some of the heavy-tailedness may be explained by primary content providers (i.e. store owners) whose role includes marketing assets to individuals. While approximately 52% of the transfers occur between non-neighboring users, many transfers occur at similar scales between users that are affiliated with one another, or more strongly, have at least three other friends in common.

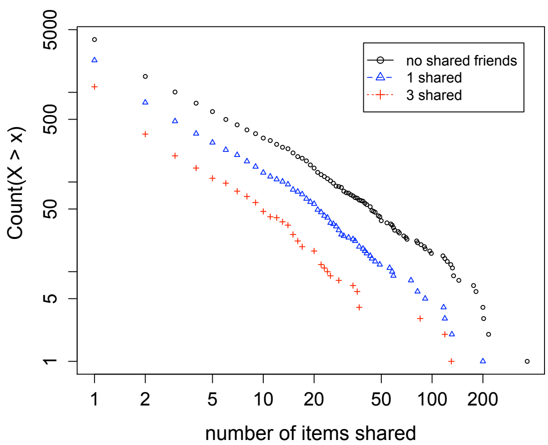

Figure 7: Distributions of the number of assets shared by users with other users with whom they share a specified number of friends in common.

We can also measure the entropy of users who are responsible for transfers and compare it against a null model where each subsequent adopter receives the asset from a randomly chosen previous adopter. The entropy is simply computed using the proportion of transfers that can be attributed to each user in the cascade who shared at least one asset. The null model has two parameters, the total number of owners of the asset n, and the proportion p of missing edges in the cascade. At each time step, the null model adds a new owner, who with probability p starts a new tree, and with probability $(1-p)$ picks one of the previous owners uniformly at random as its parent node. The null model was computed for each asset using the corresponding (n, p).

We find that the distribution of entropies from the data, measured in bits, has a mean of 2.72, which is significantly lower than that of the null model (3.48), (t = 97.08, p = 0). This indicates that the actual distribution of assets is more concentrated than one would expect if every previous adopter participated with equal probability.

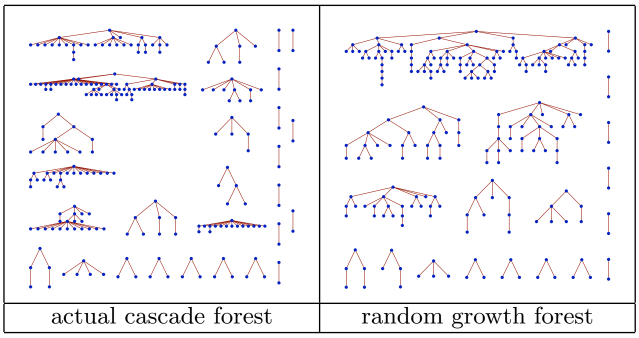

Figure 8: Comparison between actual growth of cascades and a null model where each previous adopter is equally likely to be sharing assets.

An obvious distinction between the null model and the actual cascades is the tendency of the observed cascades to be concentrated on the social graph, with many users adopting after their friends do. As we mentioned before, 48% of the direct transfers occur on the social graph. A null model that takes just any previous adopter as the source of an asset would pick a friend 6.6% percent of the time. This is calculated by computing the fraction of previous adopters who are friends for each transfer with accurate previous owner information and dividing by the total number of such transfers. Unsurprisingly, direct sharing is more a feature of friendship ties than simply a desire to share with others.

3.2 Strength of influence

The number of times a user transfers assets is an unambiguous influence measure. However, it doesn't capture how successful a user would be in a competition where one's friends could obtain assets from others. We propose a simple measure, γ, that compares the number of times a user A infected one of its friends B, against the expected number given the odds that B was not infected by one of their other adopter friends. For example, if B had 2 other friends besides A who had previously adopted, and B obtained the asset through a friend, then the probability that A was the infector is 1/3. This adds 1/3 to the expected number of transfers for A.

We measure γ =*(transfers - expected)/expected* for all users who had at least 20 instances where one of their friends acquired an asset through a social tie after they did. Figure 9 shows what if odds were even that the adopter receives the asset from any one of their friends, the user's γ scores would be narrowly distributed around 0 – they would be doing no worse or better than odds. In contrast, the distribution of observed γ scores is highly skewed – approximately 74%, fall below 0 and while the remaining 26% are more influential than odds. The actual gammas have a mean of -0.286.

A further question one might have is whether a user can be influential in distributing many assets or just a few. The overall correlation between the number of transfers a user made and the number of assets they were sharing was highly positive ($\rho = 0.63$), but still displayed a wide range of behaviors. One user influenced 73 transfers to friends involving just 2 different assets while another made 104 transfers involving 16 different assets. In yet another case, a user distributed 46 assets in 47 transfers. This implies that some users share only a few select items while others share less discriminately.

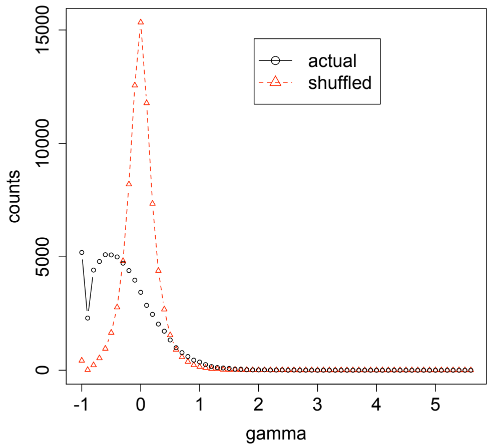

Figure 9: Users' influence using the γ measure, for actual and randomized transfers.

Who are these influencers and what are their characteristics? Interestingly, even though users with a higher number of friends tend to have been around longer ($\rho = 0.13$), have more assets ($\rho = 0.16$), and have made more transfers in total ($\rho = 0.14$), a user's γ score is negatively correlated with their number of friends ($\rho = -0.17$[1]). This is likely because maintaining strong ties with many individuals is more difficult, hence influencing any single one is less likely. We observe, for example, that the number of assets shared by two friends is mildly correlated with the strength of their tie ($\rho = 0.10$), as given by the number of friends the two have in common.

A higher γ is slightly negatively correlated with the number of assets ($\rho = -0.05$), but highly positively correlated with the number of transfers to friends per asset owned ($\rho = 0.35$). This means that influencers don't necessarily have more assets than others, but the ones that they do have, they like to share with their friends. Overall, we find that users who are sharing a higher number of assets and making more transfers tend to be sharing less popular ones $\rho = -0.15$, again suggesting, as in Section 1.2 that assets shared tend to be niche products.

We also examine whether those who are directly responsible for their friends' adoptions tend to acquire assets earlier. While users who have more transfers per asset tend to be "earlier" in their adoption ($\rho = -0.06$), both in terms of absolute rank (they were the r^{th} person to adopt) and relative rank (they were among the first p% of users to adopt), a user's γ score and relative adoption rank are uncorrelated.

[1]numbers of friends and assets were log-transformed before their correlation was measured

Altogether, combining the age, number of friends, number of assets, average adoption rank, and average number of transfers in a linear regression model yields an R^2 of 0.17 for a user's γ score.

3.3 Early adopters

This still leaves the question of whether the very earliest adopters might be different as a group from other users. We select users who have 20 or more gestures and are on average among the first 5% of adopters for all assets they own. This corresponds to being the 15th adopter on average across the assets one owns. For the analysis below we obtained qualitatively similar results when we selected an early adopter group of approximately the same size, but slightly different criteria: adopting 40 or more gestures, and being among the first 10% of users to acquire them.

We compare the early adopter group against the group of 50,000 users who have also acquired 20 or more gestures, but are on average in the latter half of adopters for those gestures. We can immediately rule out some factors relating to whether a user becomes an early adopter. The early adopters were on average born just 68 days earlier, meaning that joining Second Life earlier yields only a slightly higher advantage in being one of the first adopters of an asset. Early adopters have actually had a bit less playtime than the later adopters (40 hours), and have an average of 8 fewer friends (for an average of 61 and median of 33). Clearly the early adopters are neither especially early, active, nor gregarious.

The very earliest adopters distinguish themselves in other ways. For the assets that they eventually adopt, the rate of adoption before any of their friends adopt, $\lambda_{k=0}$, is twice as high as that of the laggard group ($t = 4.2, p < 0.0001$), as is their rate of adoption under initial social influence, $\lambda_{k=1}$, though this difference was not as significant ($t = 2.3, p < 0.05$). This indicates that they are more susceptible to adopting assets early (when none or one of their friends have adopted), although on average they own 20 fewer assets than late adopting users (t = 10.3, p = 0). Perhaps, being trendsetters, they resist acquiring assets that have become too common.

Finally, we examine the direct influence that these early adopters wield, and find that their γ scores, though closer to odds (-0.08) than that of the later adopters (-0.22) are not particularly impressive. The number of transfers they make is not significantly higher than the laggard group, even though the assets they adopt eventually grow to be more popular than those owned by laggard group ($t = 5.5, p < 10^{-7}$). Previously simulated models of social influence over social networks have established a negative link between being an early adopter (easily succumbing to a new trend) and therefore been less influential [29]. This is not the case for the most extreme early adopters in Second Life. But the overall trend for all users is a very slight but statistically significant negative correlation between the probability that one adopts before one's friends do, and both γ ($\rho = -0.015, p < 0.001$) and number of transfers the user makes ($\rho = -0.02, p < 10^{-7}$).

In summary, we identified some users as influential, and others as early adopters. They don't appear to be one and the same, with the early adopters being more easily susceptible early on, but not being more likely to share their finds. We were able to identify some characteristics of both early adopters and influencers, however, these characteristics alone cannot be used reliably to identify such users. The size of a users' social network is just one of the variables that was of little help in identifying influencers, although the social network itself is responsible for many of the transfers.

4. CONCLUSION AND FUTURE WORK

In this paper we examined the interplay of social networks and social influence in the adoption of online content. Roughly 48% of transfers occur along the social graph, the remainder occurring between users who are not friends. We find that assets whose transfers typically occur through the social graph tend to have deeper transfer cascades measured as a higher proportion of non-leaf nodes, but tend to grow more slowly. This suggests that social networks are an important medium for diffusion of *niche* information in Second Life.

We applied models of social contagion that capture the rate at which users adopt following the adoption by one of more of their friends. We find that the rate of adoption increases as more of one's friends adopt, and that this is more significant for smaller, niche assets. We also find that someone who has many friends is less likely to be influenced by any particular one. A user with many ties would have difficulty maintaining all of them, increasing the probability that many of the ties are weak and therefore hold less influence. Indeed, we found a slight correlation between the strength of a tie and number of assets that are transferred between two friends.

We further find that some individuals play a more active role in the transfer of assets than others. A random cascade model, where any node is equally likely to cause another adoption, yields a higher entropy than the empirically observed cascades. But the variability in influence cannot be attributed to the social network alone: when we measure the direct influence an individual has on a particular friend, this influence is negatively correlated with the number of friends. Finally, the early adopters, while being more susceptible to adopting content without waiting for many of their friends to so, do not wield greater influence over others.

In future work we would like to examine the effect of fees on the transfer of assets. Individuals may behave rather differently when assets are costly to acquire. They may either seek to keep up with the Jonses or be less likely to succumb to peer influence because of the associated cost. Another interesting dimension for exploration is that of copyright. Copyright may inhibit the spread of assets, favoring the spread of those where users are free to share and modify the content. The effect of users' ability to modify content created by others, and more generally collaborate in this virtual space, would be a fascinating subject of study.

5. ACKNOWLEDGMENTS

We thank Alex Dailey, Jimmy Li, and Everett Harper of Linden Lab for providing the data used in this study and for valuable discussions. We would also like to thank Theodore Iwashyna for his help with the Cox model. This work was supported by NSF IIS-0746646 and NSF IGERT-0654014.

6. REFERENCES

[1] A. Anagnostopoulos, R. Kumar, and M. Mahdian. Influence and correlation in social networks. In *KDD*

'08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining, pages 7–15, New York, NY, USA, 2008. ACM.

[2] L. Backstrom, D. Huttenlocher, J. Kleinberg, and X. Lan. Group formation in large social networks: membership, growth, and evolution. In *KDD '06: Proceedings of the 12th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 44–54, New York, NY, USA, 2006. ACM.

[3] F. M. Bass. A new product growth for model consumer durables. *Management Science*, 15(5):215–227, 1969.

[4] E. Castronova. A Test of the Law of Demand in a Virtual World: Exploring the Petri Dish Approach to Social Science. *SSRN eLibrary*, 2008.

[5] D. Centola and M. Macy. Complex Contagions and the Weakness of Long Ties 1. *American Journal of Sociology*, 113(3):702–734, 2007.

[6] R. Chatterjee and J. Eliashberg. The innovation diffusion process in a heterogeneous population: A micromodeling approach. *Management Science*, 36(9):1057–1079, 1990.

[7] D. R. Cox and D. Oakes. *Analysis of survival data*. Chapman & Hall, London, 1984.

[8] P. Domingos and M. Richardson. Mining the network value of customers. In *KDD '01: Proceedings of the seventh ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 57–66, New York, NY, USA, 2001. ACM.

[9] D. Friedman, A. Steed, and M. Slater. Spatial Social Behavior in Second Life. *Lecture Notes in Computer Science*, 4722:252, 2007.

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

[11] 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, New York, NY, USA, 2003. ACM.

[12] G. Kossinets, J. Kleinberg, and D. Watts. The structure of information pathways in a social communication network. In *KDD '08: Proceeding of the 14th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 435–443, New York, NY, USA, 2008. ACM.

[13] K. Lerman. Social information processing in news aggregation. *IEEE Internet Computing*, 11(6):16–28, 2007.

[14] K. Lerman and L. A. Jones. Social browsing on flickr. In *ICWSM*, 2007.

[15] J. Leskovec, L. A. Adamic, and B. A. Huberman. The dynamics of viral marketing. In *EC '06: Proceedings of the 7th ACM conference on Electronic commerce*, pages 228–237, New York, NY, USA, 2006. ACM.

[16] J. Leskovec, A. Krause, C. Guestrin, C. Faloutsos, J. VanBriesen, and N. Glance. Cost-effective outbreak detection in networks. In *KDD '07: Proceedings of the 13th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 420–429, New York, NY, USA, 2007. ACM.

[17] J. Leskovec, A. Singh, and J. Kleinberg. Patterns of influence in a recommendation network. In *Pacific-Asia Conference on Knowledge Discovery and Data Mining (PAKDD)*. Springer, 2006.

[18] 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, 2008.

[19] V. Mahajan, E. Muller, and F. M. Bass. New product diffusion models in marketing: A review and directions for research. *Journal of Marketing*, 54(1):1–26, 1990.

[20] M. Newman. Spread of epidemic disease on networks. *Physical Review E*, 66(1):16128, 2002.

[21] M. Newman, S. Forrest, and J. Balthrop. Email networks and the spread of computer viruses. *Physical Review E*, 66(3):35101, 2002.

[22] C. Ondrejka. Aviators, moguls, fashionistas and barons: Economics and ownership in second life. Available at SSRN: http://ssrn.com/abstract=614663.

[23] C. Ondrejka. A piece of place: Modeling the digital on the real in second life. *Social Science Research Network Working Paper Series*, June 2004.

[24] R. Pastor-Satorras and A. Vespignani. Epidemic Spreading in Scale-Free Networks. *Physical Review Letters*, 86(14):3200–3203, 2001.

[25] E. M. Rogers. *Diffusion of Innovations*. Free Press, New York, fourth edition, 1995.

[26] M. J. Salganik, P. S. Dodds, and D. J. Watts. Experimental study of inequality and unpredictability in an artificial cultural market. *Science*, 311(5762):854–856, 2006.

[27] X. Song, Y. Chi, K. Hino, and B. Tseng. Information flow modeling based on diffusion rate for prediction and ranking. In *Proceedings of the 16th international conference on World Wide Web*, pages 191–200. ACM Press New York, NY, USA, 2007.

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

[29] D. Watts and P. Dodds. Influentials, Networks, and Public Opinion Formation. *Journal of Consumer Research*, 34(4):441, 2007.

[30] F. Wu and B. Huberman. Novelty and collective attention. *Proceedings of the National Academy of Sciences*, 104(45):17599, 2007.

[31] N. Yee, J. Bailenson, M. Urbanek, F. Chang, and D. Merget. The Unbearable Likeness of Being Digital: The Persistence of Nonverbal Social Norms in Online Virtual Environments. *CyberPsychology & Behavior*, 10(1):115–121, 2007.

[32] R. Zheng, F. Provost, and A. Ghose. Social Network Collaborative Filtering. Working paper CeDER-8-08. Center for Digital Economy Research, Stern School of Business, New York University., 2007.