# Hot Today, Gone Tomorrow: On the Migration of MySpace Users\*

Mojtaba Torkjazi, Reza Rejaie
Department of Computer & Information Science
University of Oregon
Eugene, OR 97403
{moji, reza}@cs.uoreogn.edu

Walter Willinger
AT&T Research Labs
180 Park Ave. - Building 103
Florham Park, NJ, 07932
walter@research.att.com

## **ABSTRACT**

While some empirical studies on Online Social Networks (OSNs) have examined the growth of these systems, little is known about the patterns of decline in user population or user activity (in terms of visiting their OSN account) in large OSNs, mainly because capturing the required information is challenging.

In this paper, we examine the evolution of user population and user activity in a popular OSN, namely MySpace. Leveraging more than 360K randomly sampled profiles, we characterize both the pattern of departure and the level of activity among MySpace users. Our main findings can be summarized as follows: (i) A significant fraction of accounts have been deleted and a large fraction of valid accounts have not been visited for more than three months. (ii) One third of public accounts are owned by users who abandon their accounts shortly after creation (i.e., tourists). We leverage this information to estimate the account creation time of other users from their user IDs. (iii) We demonstrate that the growth of allocated user IDs in MySpace was exponential, followed by a sudden and significant slow-down in April 2008 due to an increase in the popularity of Facebook. If such up- and down-turns are symptomatic of OSNs, they raise the obvious question: What are the main forces that enable some systems to compete and strive in the Internet's OSN eco-system, while others decline and ultimately die out?

## **Categories and Subject Descriptors**

C.2.4 [Computer-Communication Networks]: Distributed Systems

## **General Terms**

Measurement

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, to republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

WOSN'09, August 17, 2009, Barcelona, Spain. Copyright 2009 ACM 978-1-60558-445-4/09/08 ...\$10.00.

# **Keywords**

Online Social Networks, User Dynamics and Activities, OSN Eco-system

#### 1. INTRODUCTION

Over the last few years, Online Social Networks (OSNs) such as MySpace [12] and Facebook [5] have attracted hundreds of millions of users and have been responsible for a new wave of popular application over the Internet. The dramatic increase in the popularity of OSNs have prompted network researchers and practitioners to examine the connectivity structure of well-known OSNs [1, 11] and the growth of their user population over time [7, 8, 9].

Given the unwillingness of OSN owners to share information about their systems, measurement is the most promising approach to characterize many of the widely-deployed OSNs and study their growth patterns. A majority of prior empirical studies on OSNs have focused on the characterization of the OSNs' friendship structures inferred from single snapshots taken at a particular point of time. A very small number of prior studies have examined the evolution of popular OSNs using multiple snapshots of the system taken over some periods of time (e.g., a couple of years) [7, 8, 9]. Furthermore, these latter studies have typically focused on the evolution (or growth) of the friendship structure without considering the level of activity by individual users; i.e., the presence of a user implicitly indicated her activity in the system. To our knowledge, most prior studies of the evolution of OSNs have focused first and foremost on the growth of these systems, paying little or no attention to the decline in popularity of OSNs or level of activities among their users.

Capturing a decline in the popularity or user activity of a large OSN is challenging for several reasons. First, to prevent a potential ripple effect on other users, popular OSNs tend to hide or obscure the information about users that have left the system or are inactive. For example, Facebook does not explicitly notify a user when her friends delete their accounts or remove their friendship links. Second, OSNs are often studied when they are very popular and the number of departing or inactive users is negligible. Capturing a decline in the popularity or user activity of a large OSN requires snapshots of the system over a long period of time which is in general expensive. These factors have affected the ability of networking researchers to characterize the down-fall of popular OSNs. Evidence for a decline in the popularity of an OSN or in user activity is typically provided by companies such as alexa.com [2] that monitor the number of accesses to

<sup>\*</sup>The phrase "Hot Today, Gone Tomorrow" is borrowed from [6].

OSN websites. Despite the usefulness of such data, it only provides information about the aggregate pattern of change in user activity or OSN popularity. For example, the estimated number of daily visits to an OSN website does not reveal whether newer users are more active (or more likely to leave the OSN) than older users.

In this paper, we examine the evolution of user population and user activity in one of the largest OSNs, namely MySpace. MySpace has several features that collectively enable us to collect representative samples of user accounts, distinguish deleted (or invalid) from valid user accounts, and quantify the level of activity among existing users using their last login information. We downloaded more than 360K profiles of randomly selected MySpace users and determined whether these profiles are invalid (or deleted), private, or public.

Using our data set, we characterize the pattern of departure and the level of activity among MySpace users. Our main findings can be summarized as follows. First, a substantial fraction (41%) of MySpace IDs are associated with user accounts that have been deleted. A relatively large fraction of these deleted accounts belong to newer users, i.e., older MySpace users are more loyal. Second, more than 75% of users with public accounts have not visited MySpace for more than 100 days. This number grows to 85% for users with public accounts that have not visited MySpace for more than 10 days. There is a higher level of activity on private accounts than on public accounts. Users with a very low level of activity appear to be surrounded by users with the same level of low activity, i.e., users appear to abandon MySpace in groups. Overall, out of 445 million allocated users IDs, only 85 and 55 million of them have been active during the last 100 and last 10 days, respectively. Third, the relation between user ID and last login of a user exhibits a very clean "edge" that represents those users who have left the system shortly after creating their account (i.e., tourists). This observation allows us to estimate the account creation time of individual users based on their user ID. This in turn enables us to identify the rate of growth in the allocation of user ID (i.e., arrival rate of new users). Our results show that MySpace experienced an exponential growth in its user population for a four-year period ending in early 2008. However, since April 2008, MySpace has seen a significant and sudden slow-down in its growth. Fourth, we corroborate our findings with additional evidence and conclude that this recent drop in the popularity of MySpace is directly related to the growing popularity of Facebook. The observed evolution pattern of MySpace suggests that many of the existing OSNs can be expected to go through a similar life cycle that is in part determined by a collection of social and technological factors and reflects an OSN's ability to compete against newer OSNs.

The rest of the paper is organized as follows. In Section 2, we describe some of the features of MySpace that we exploit for our study and present our measurement methodology for data collection. We characterize the patterns of departures and activities among MySpace users in Section 3. In Section 4, we discuss the underlying causes for the decline in the popularity of MySpace and in the activity among MySpace users. Finally, we speculate why other successful OSNs are likely to experience MySpace-like up- and down-turns during their life time.

## 2. MEASUREMENT METHODOLOGY

MySpace is one of the most popular OSNs with a few hundreds of millions of reported users. We first discuss a number of features that are unique to MySpace and that we rely on for our approach. To this end, we exploit the simple fact that since the MySpace server generates an error message for a large and thus unassigned ID, at any point in time during an experiment, we can easily determine the maximum ID that has already been assigned up to that time.

# 2.1 Monotonic Assignment of Numeric IDs

We gathered couple of strong evidences that MySpace assigns numeric user IDs in a sequential fashion. First, we repeatedly run an experiment whereby we first identified the currently smallest unallocated ID and checked that all the larger IDs are unallocated as well. After waiting for a short period, we found that that smallest unallocated ID and other IDs after it were now all allocated. Second, we also examined all the IDs within a particular range of IDs and checked the gaps between consecutive deleted IDs. Examining the resulting distributions of these gaps between all consecutive deleted accounts for different ranges of values (as shown in Figure 1(a)) provides no indications of any patterns for the allocation of IDs. Although not conclusive, these tests strongly support the claim that MySpace allocates user IDs in a sequential fashion.

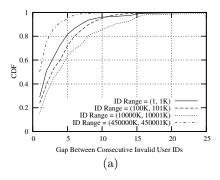
An implication of the apparently sequential ID assignment strategy in MySpace is that there is a direct relationship between the ID of a user and that user's age in the system. While we can determine the hourly and daily rate of arrival for new users, it is difficult to estimate this rate for users that arrived in the past. Thus, estimating the age of a user in the system directly from its user ID is in general non-trivial. Lastly, the ability to determine the maximum ID allocated up to any given time also enables us to identify the range of valid user IDs at a particular point of time and generate random IDs to select random (and thus representative) samples of MySpace users.

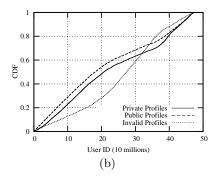
# 2.2 Providing Explicit Profile Status

A unique feature of MySpace is the availability of explicit information on whether a profile is public or private. MySpace also provides information about whether a given profile is still valid or has been removed. More specifically, when one requests access to a profile that has been deleted, the system responds with the message "Invalid Friend ID. This user has either canceled the membership, or the account has been deleted." A profile is invalid because the user has either left MySpace or her account has been deleted by the system administrators. MySpace deletes profiles for violations of its "Terms of Service" and has done so more aggressively in the recent past due to news about hosting illegal content (e.g., nude photos, racist content, videos or pictures with sexual content or excessive violence, and gang related content). MySpace is one of the few (and quite possibly the only popular) OSNs that openly reports information about deleted profiles. In fact, many other popular OSNs such as Facebook do not even inform a user when their friends remove their friendship links or leave the system for good.

## 2.3 Availability of Last Login

Almost all the valid public and private MySpace profiles contain the "date of last login" for the corresponding user.





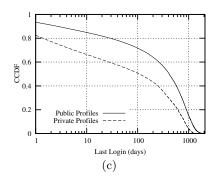


Figure 1: (a) User ID allocation for different ranges of user ID, (b) Distribution of user ID, (c) Distribution of last login.

Our examinations revealed that the last login status of individual users is indeed updated by the OSN server after each user login. While the granularity of last login information is relatively coarse (*i.e.*, date as opposed to date and time), it still provides very valuable information to assess the level of activity among users.

# 2.4 Global Visibility

MySpace allows access to all profiles in the system with no authentication. For a given user ID (e.g.,user\_id), the corresponding profile can be easily accessed using a URL of the form http://www.myspace.com/user\_id. Downloading the file at such a URL will provide the information that is necessary to determine whether the account is deleted or valid; and for valid accounts, whether it is private or public. While for private profiles, we have only access to the date of the user's last login, in the case of public profiles, the entire user information in the profile is available to us.

# 2.5 Data Collection

The largest MySpace ID at the time of our measurement was 455,881,700. Our data collection started on Feb. 26th 2009, 12:47 am PDT, and lasted for 11.5 hours. Using 50 parallel samplers, we generated 362,283 random user IDs and downloaded the profiles of the corresponding MySpace users. This number of samples corresponds to a sampling rate of approximately 0.01%. Using HTML parsing, we post-processed the downloaded profiles to extract the following information:

- Profile Status: This could be private, public or invalid,
- Last Login Date: only for private and public profiles,
- List of Friends: only for public profiles.

For a random subset of sampled public profiles, we also downloaded the profile of all the listed friends to examine any possible correlation between the behavior of sampled users and their friends. We were unable to parse the last login date for 0.96% of public and 0.08% of private profiles. Closer examinations revealed that these profiles either do not provide the last login date or provide erroneous information (e.g., last login date for a user was 1/1/0001). We removed these small percentage of profiles from our data set.

# 3. CHARACTERIZING PROFILES

## 3.1 User Departure

We are first interested in the questions "Do MySpace users actually leave the system?" and "Are newer users more likely to leave than older ones?" In part, the answer to the first question is given in Table 1 that summarizes the break-down of our sampled profiles into invalid (or deleted), public, and private profiles. Interestingly, more than 41% of the selected random user IDs, while within the valid range, were invalid. This suggests that a substantial fraction of MySpace users have left the OSN, either voluntarily or because their accounts have been deleted by MySpace. About 17% of MySpace profiles are valid private profiles and the remaining ones (close to 42%) are valid public profiles. Given that we know the largest valid user ID at the time of our measurement, these statistics imply that the population of valid MySpace users at the time of our experiments was around 268 million.

Total		Public	Private
362K	149K (41.2%)	150K (41.5%)	63K (17.3%)

Table 1: Sampled MySpace user profiles broken down by number of invalid (deleted), public, and private profiles on 2/26/2009

Concerning the second question, Figure 1(b) shows the distribution of each group of profiles across the entire range of user IDs. This figure indicates that for a large initial portion of the ID space (i.e., users who joined the system early on), the number of public profiles is between 5-10% higher than the number of private profiles. Towards the end of the ID space (i.e., more recent users), this difference reaches to zero and at the same time, the percentage of invalid profiles exceeds the percentage of private and public profiles. Figure 1(b) also shows that the number of private and public profiles in the first half of the ID space is respectively 5% and 10% higher than in the second half of the ID space. To compensate for this, the percentage of invalid profiles in the second half of the ID space is some 15% higher than in the first half. These observations suggest that relatively speaking, MySpace users who joined the system earlier have been more loyal and are more likely to keep their accounts than users who joined more recently.

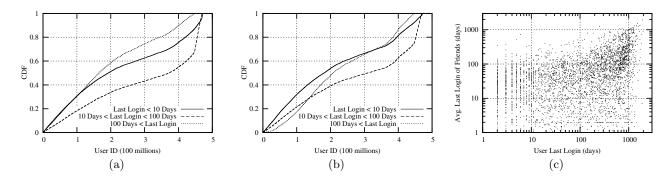


Figure 2: User activity. (a) User ID distribution for public users with different range of activity, (b) User ID distribution for private users with different range of activity, (c) Relation between activity of a user and her friends.

# 3.2 User Activities

Next we are interested in the questions "How active are users who have a valid public or private profile?" and "Do inactive users tend to leave MySpace individually or in groups?" To shed light on these questions, we measure the level of activity of users with a valid public or private profile in terms of the date of their last login: A user whose last login was within the last 10 days is considered to be more active than a user whose last login dates back 100 days. Figure 1(c) depicts the CCDF of the duration of time between a user's last login and our measurement date for all users with valid public or private profiles who have been in the system for more than 100 days. The figure shows that only about 30% of the users with a valid public profile have logged in within the last 100 days, and this number drops to about 15% for those users who have logged in within the last 10 days. The figure also shows that in relative terms, users with valid private profiles are more active - roughly 50% and 25% of them logged into the system during the last 100 and 10 days, respectively. If we conservatively assume the active users are those who have logged into the system during the last 100 days, we obtain an estimated population of active MySpace users (both public and private) at the time of our experiment of slightly more than 96 million users. This estimate drops to around 55 million users if we only count those users who logged in during the last 10 days. These observations suggest that a large fraction of users with a valid profile are not actively using MySpace and might very well have abandoned the system.

To explore potential correlations between the age of a user in the system and the user's level of activity (measured in terms of last login date), Figure 2(a) depicts the distribution of user IDs with valid public profiles for three groups of users whose last login date was within the following windows (relative to our measurement date): (i) within the last 10 days (i.e., very active users), (ii) between the past 10 to 100 days (i.e., moderately active users), (iii) more than 100 days ago (i.e., inactive users). Figure 2(b) depicts the same information for users with valid private profiles. Figure 2(a) demonstrates that active users are roughly evenly distributed across the ID space, with only a slightly larger fraction of them (about 8%) within the first half of the ID space. While the percentage of inactive users in the first half of the ID space is around 17% larger than in the second half, for the moderately active users the opposite is true;

their percentage in the second half is around 13% larger than in the first half of ID space. The distributions of users IDs with valid private profiles with different level of activity across the ID space follows a similar trend, but are generally more even as shown in 2(b). While the percentage of very active and inactive users in the first half of the ID space is slightly larger than for public users, the percentage of moderately active users is larger in the second half when compared to users with valid public profiles.

To answer the question about whether inactive users leave MySpace individually or in groups, we focus on the inactive users with valid public or private profiles and check for possible group formations by inactive users in their underlying friendship graph. To this end, recall that for each randomly selected user with a valid public profile, we have also collected the profiles of all the listed friends. Using this information, we examine the correlation between the level of activity of selected users and their immediate neighbors. Figure 2(c) is a scatter plot that shows for each randomly selected user the number of inactive days (i.e., time since the user's last login) on the x-axis and the average number of inactive days among that user's neighbors on the y-axis. The figure indicates no strong correlation between the activity of a user and the (average) activity of its neighbors, except for the very inactive users who have not logged in for more than a few hundred days. In particular, the top-right part of the figure suggests that inactive users tend to have inactive friends, implying that inactive users are more likely to abandon MySpace in groups than one by one.

## 3.3 User Arrival

The last question we address here is "What does the user ID indicate about the user's account creation time?" To study this issue, we check whether there is any relationship between a user's ID and the days since that user's last login. Figures 3(a) and 3(b) show scatter plots of user ID (on x-axis) and the days since the user's last login (on y-axis) for our set of randomly selected public and private users, respectively. While the monotonic decrease of these graphs is intuitive and fully expected, the appearance of a clean "edge" in these graphs is surprising. Such a clean edge can only be explained by users whose last login occurred immediately or very shortly after the creation of their accounts. We refer to these users as tourists for obvious reasons. This discovery of MySpace tourists is sup-

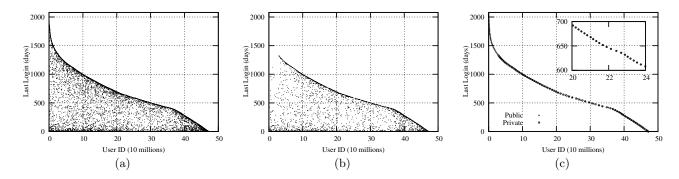


Figure 3: (a) Last login of public users, (b) Last login of private users, (c) Last login of tourists.

ported by the fact that the region below this clean edge is more or less uniformly covered by our samples, and not a single data point exists above the edge. More specifically, since the last login of a tourist occurs shortly after the creation of her account, other users with higher user IDs who have created their accounts after this tourist cannot have an earlier last login, even if they have logged in only once right after creating their accounts. We note that about 32% of the public and 18% of the private users are tourists. Moreover, both public and private tourists are roughly uniformly scattered across the entire ID space. Figure 3(c) shows only the private and public users that are tourists, and reveals that the edges associated with tourists in both groups are perfectly aligned.

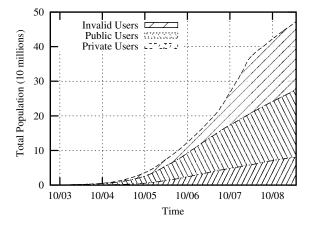


Figure 4: Growth of MySpace

Given the relatively uniform spread of tourists across the entire ID space, we consider their last login time as a good estimate of their account creation times, and accurately estimate the creation time of all of our sampled accounts based on their associated user ID. Leveraging this information, the top line in Figure 4 shows the value of each user ID (*i.e.*, total number of created MySpace profiles or an upper bound for the MySpace user population) as a function of profile creation date. We further divide the total number of user profiles into invalid (top region), public (middle region) and private (bottom region) to demonstrate the growth of the total number of profiles in each category over time.

Figure 4 illustrates several interesting features. First, the slope of the bottom line shows a linear rate of increase in the total number of private users. The slope of the second line represents the combined rate of increase in the number of private and public users and also shows a roughly linear behavior over time. Second, the slope of the top line in Figure 4 represents the rate of growth in the population of MySpace profiles over time. We observe that MySpace experienced a slow growth at the beginning, from the time it was launched until about February 2005. During that period, on average, 34K new users joined the system each day. What follows is a period of exponential growth, from February 2005 til April of 2008, during which the average profile creation rate was about 315K new profiles per day. Around April of 2008, the clearly visible "knee" in Figure 4 indicates a significant and sudden slow down in the growth of the MySpace user population. In fact, we observe a return to a linear growth, where the average daily rate of newly created profiles is around 265K profiles per day. Interestingly, Figure 4 also reveals that a disproportionally large fraction of profiles created during this last phase are invalid (or deleted), i.e., relatively many new users delete their profiles within a year from when the profile was created. In summary, Figure 4 suggests that MySpace is experiencing a life cycle reminiscent of the logistic curve nature of technology adoption as modeled in diffusion of innovations theory [13]. In the next section, we discuss factors that might explain the observed life cycle for MySpace.

## 4. THE LIFE CYCLE OF MYSPACE

## 4.1 Informal Evidence

Examining various public reports suggests that the observed slow-down in the growth rate of MySpace population is directly related to the emergence of another OSN, namely Facebook. In [15], the research firm comScore [4] reported that the number of unique monthly visitors for Facebook has surpassed that of its rival MySpace. Additional evidence is given in Figure 5 that shows the percentage of daily accesses reported by Alexa.com [2] for three major OSNs: Facebook, MySpace and Orkut. According to this metric, Facebook surpassed MySpace around April 2008 and continues to grow at a rather steady pace while the popularity of the other two OSNs is clearly decreasing. While these trends have been reported in the popular press and are in general well-known, they are typically based on qualitative data such

as estimated number of daily visits to an OSN website. In contrast, our study not only confirms these trends, but does so by directly examining MySpace profiles. Moreover, our quantitative approach enables us to examine these trends or other observed features of MySpace in ways that are not feasible with more qualitative data such as the one provided by Alexa.com.

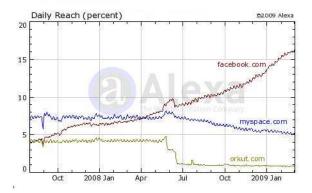


Figure 5: Comparison of daily access to three popular OSNs by Alexa, captured on March 05, 2009

# 4.2 Why Do Users Abandon Popular OSNs?

The departure of users after OSNs become very popular appears to be a common phenomenon and not specific to a particular OSN. However, in the specific context of OSNs, this behavior has been attributed to two major and somewhat related reasons [6, 10]. First, as OSNs become huge, it becomes increasingly more challenging to effectively link its like-minded users together. Furthermore, the more popular an OSN is, the more likely it is that it attracts attackers and ad spammers, which in turn increases the frequency of privacy violations and unwanted messages. The result is more alienated users who are more likely to abandon their current OSN for a competing, more user-friendly OSN. In this sense, popular OSNs appear to become the victim of their own success. For example, it will be interesting to observe the future evolution of a currently popular OSN such as Facebook. Second, existing OSNs appear to be very vulnerable to the arrival of new "fashions" among users (e.g., messaging with Twitter). Even in the absence of new fashions, it is in general difficult to keep the user interested in one and the same OSN for a long time without introducing new features and applications. In the absence of constant innovation by the OSN owner, the initial excitement of users fades away and they gradually abandon the system, typically without any prior warning.

To date, the Internet has already witnessed the rise and fall of a few OSNs, such as Friendster [3] and Yahoo! 360 [14]. This suggests that OSNs, similar to so many other innovative services and products, may also have a natural life cycle. However, our knowledge of the details of their life cycles remains very limited. Moreover, our understanding of the socio-technological factors, that enable some OSNs to compete and thrive in the OSN marketplace while others struggle and ultimately disappear, remains limited as well. While networking researchers can be expected to develop new OSN architectures that can scale to support hundreds

of millions of users and provide adequate performance, they are less likely to predict the type of applications and new services that an OSN needs to offer in order to retain its existing users while attracting new users. Both problem areas can benefit tremendously from measurement studies of existing OSNs. Toward this end, this paper demonstrates that there are unique opportunities for collecting and analyzing OSN-specific measurements that can provide a great insight about how current users of OSNs use the system and migrate between them.

## 5. REFERENCES

- Y.-Y. Ahn, S. Han, H. Kwak, S. Moon, and H. Jeong. Analysis of Topological Characteristics of Huge Online Social Networking Services. In WWW, May 2007.
- [2] Alexa. http://www.alexa.com.
- [3] M. Chafkin. How to Kill a Great Idea! *Inc. The Daily Resource for Entrepreneurs*, June 2007.
- [4] comScore. http://www.comscore.com.
- [5] Facebook. http://www.facebook.com.
- [6] Knowledge@Wharton. MySpace, FaceBOOK and Other Social Networking Sites: Hot Today, Gone Tomorrow? Wharton Journal, May 2006.
- [7] R. Kumar, J. Novak, and A. Tomkins. Structure and the Evolution of Online Social Networks. In KDD, Aug. 2006.
- [8] J. Leskovec, L. Backstrom, R. Kumar, and A. Tomkins. Microscopic Evolution of Social Networks. In KDD, Aug. 2008.
- [9] J. Leskovec, J. Kleinberg, and C. Faloutsos. Graphs over Time: Densification Laws, Shrinking Diameters and Possible Explanations. In KDD, Aug. 2005.
- [10] J. Levy. MySpace or ByeSpace: MySpace Growth Tailing Off As Site Population Reaches Critical Mass. Wall Street Journal, Oct. 2006.
- [11] A. Mislove, M. Marcon, K. P. Gummadi, P. Druschel, and B. Bhattacharjee. Measurement and Analysis of Online Social Networks. In *Internet Measurement Conference (IMC)*, Oct. 2007.
- [12] MySpace. http://www.myspace.com.
- [13] E. M. Rogers. Diffusion of Innovations. Free Press, 5th edition, 2003.
- [14] E. Schonfeld. Yang Decides to Shut Down Yahoo 360—Nobody Notices. *TechCrunch*, Oct. 2007.
- [15] E. Schonfeld. FaceBOOK Is Not Only The World Largest Social Network, It Is Also The Fastest Growing. TechCrunch, Aug. 2008.