Sizing-up Online Social Networks

Mojtaba Torkjazi, Masoud Valafar, Reza Rejaie, Walter Willinger

Mirage Lab – University of Oregon   http://mirage.cs.uoregon.edu/OSN

1. Introduction

- **Problem:**
  - The size of popular Online Social Networks (OSNs) such as *MySpace* and *Twitter* has been reported to be hundreds of millions of users (and still growing).
  - Little is known about the population of *active users*.

- **Challenge:** Growth of an OSN's overall population over time as well as its population of *active users* cannot be easily determined, mainly because:
  - OSNs are generally silent about declining growth rates or the fraction of users who abandon their accounts.
  - OSNs are typically studied when they are very popular and the fraction of departing or inactive users is likely to be small.

2. Approach

- We describe a simple measurement technique for accurately inferring the growth in user population. It applies to OSNs (such as MySpace and Twitter) that:
  - assign numerical user IDs to users, and
  - their ID allocation strategy can be determined.
- We generate a random user ID in the range of valid numerical user IDs, and then
- By parsing HTML pages or leveraging the API, we collect following information about users:
  - Profile status (deleted, public, or private)
  - Date of the last activity (last login or last tweet)

Table 1: Status of Sampled Profile

OSN	Total	Deleted	Public	Private
MySpace	238,898	101,097 (42.3%)	98,948 (41.4%)	38,853 (16.3%)
Twitter	757,562	86,562 (11.4%)	350,585 (46.3%)	320,396 (42.3%)

3. Growth of User Population

- Each point in Figures 1 and 2 represents a sampled user, where the x-value is the associated user ID and the y-value is her last login.
  - **Account Creation Time:** *The appearance of a clean edge in these figures is interesting. This edge can only be created by users whose last login is shortly after their account creation time. We call such users tourists.*
  - **ID Allocation Strategy:** The closer last login of tourists with greater user IDs to the measurement time suggests a monotonically increasing ID allocation scheme.
- The uniform spread of tourists across the ID space allows us to interpolate the account creation time of any user from her user ID based on the account creation time of her adjacent tourists.

- Figures 3 and 4 demonstrate the growth in total number of users (and its breakdown by deleted, public, and private accounts) in MySpace and Twitter, respectively. These figures reveal two interesting points:
  1. Growth in user population appears to exhibit distinct phases:
     - *Initial:* When the OSN is relatively unknown
     - *Expansion:* When there is a significant hype
     - *Maturity:* When the interest in the OSN is fading
  2. Over time, different OSNs appear to be in different phases of their growth. Focusing on 2009, MySpace has reached its maturity while Twitter is still in its expansion phase.
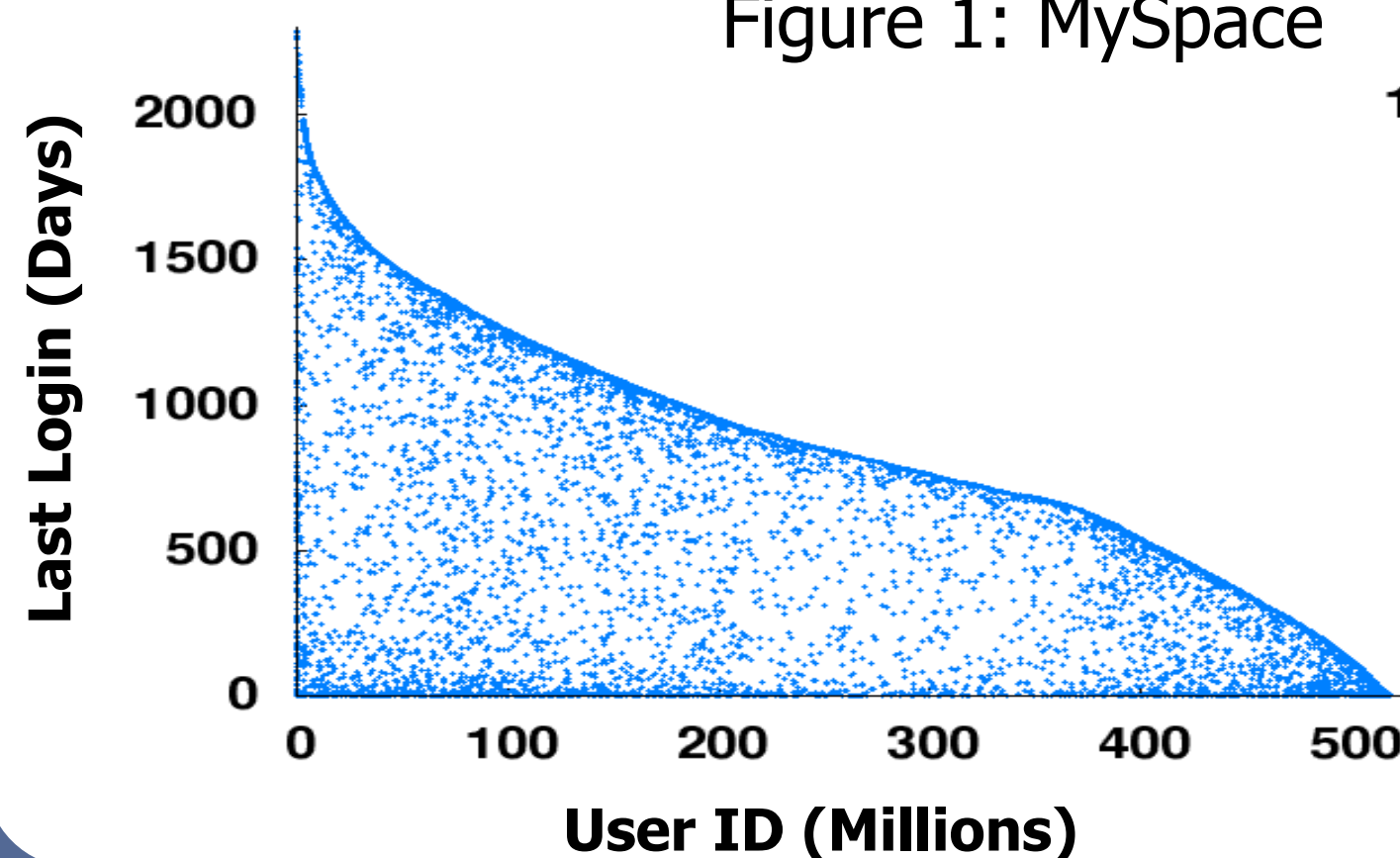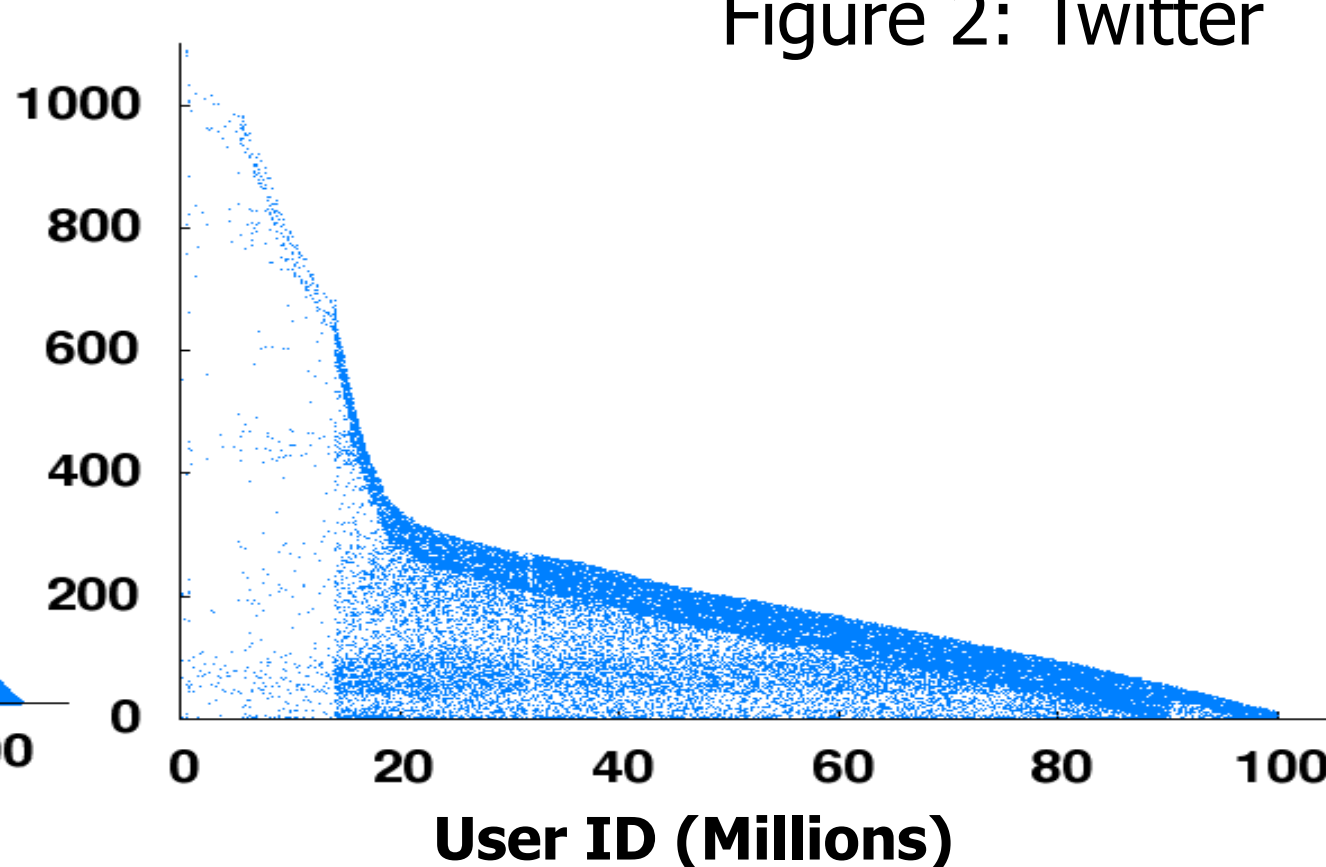

Figure 1: MySpace
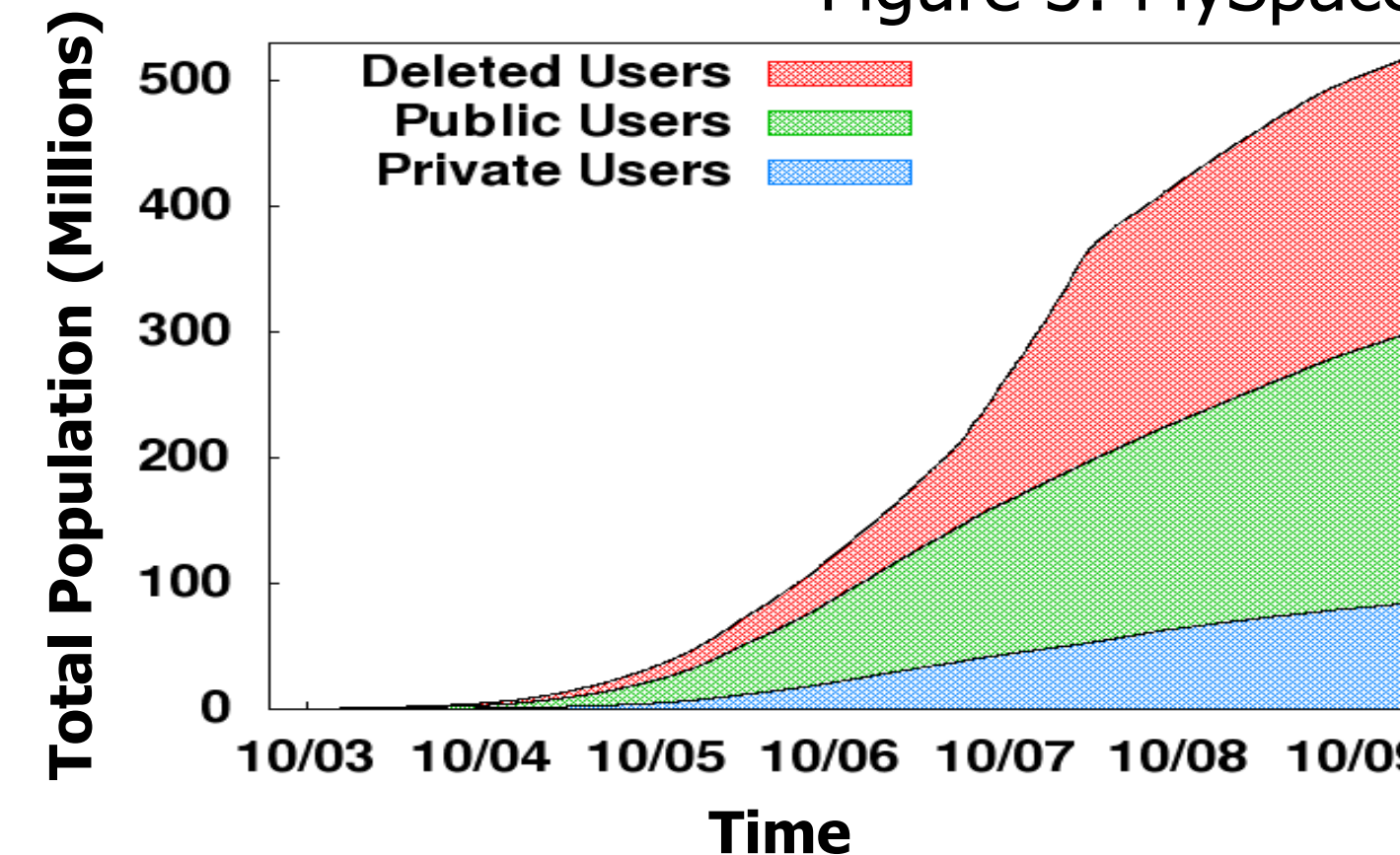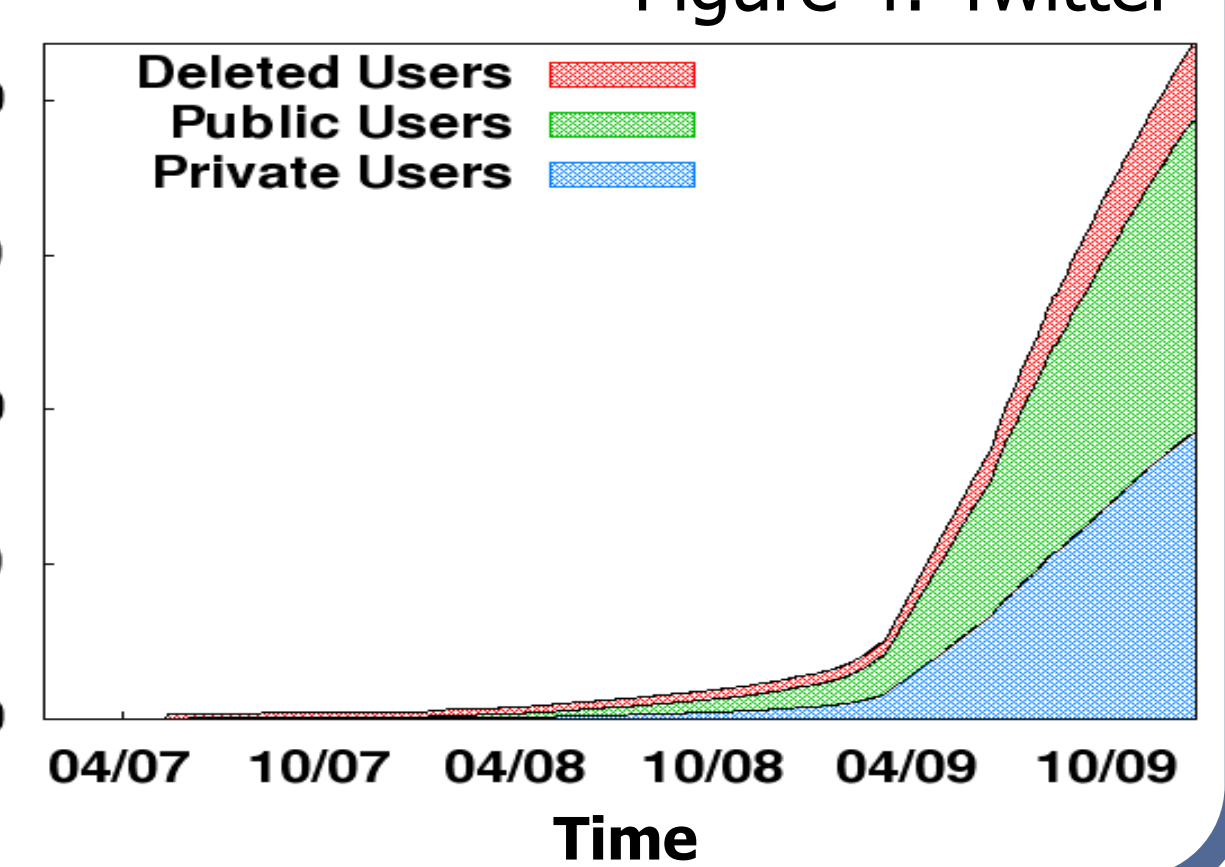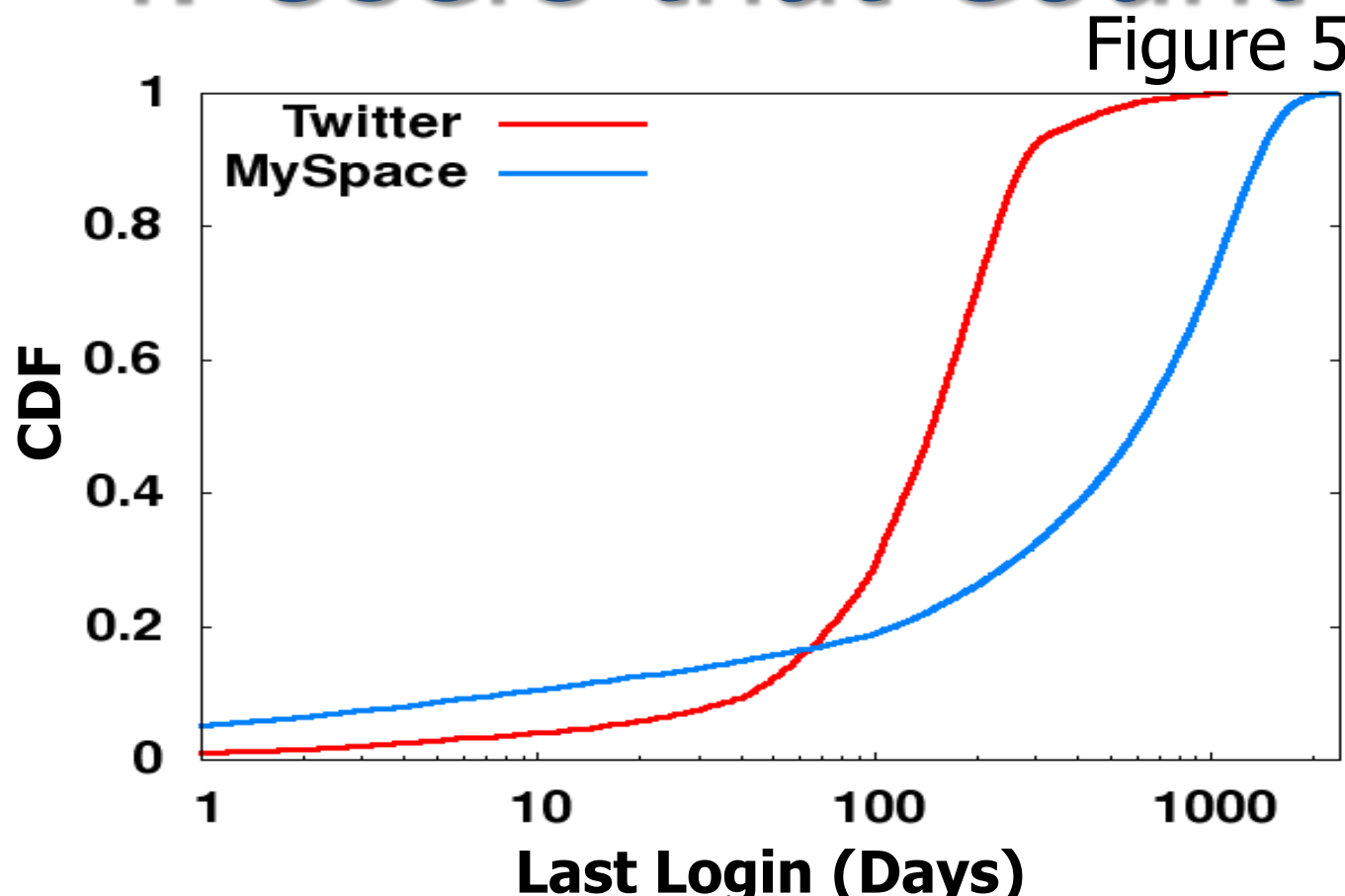

Figure 2: Twitter

Figure 3: MySpace

Figure 4: Twitter

4. Users that Count

Figure 5

- Figure 5 shows the CDF of the duration of time between a user's last login and the time of our measurement.
- Only 7.9% of MySpace users and 3.5% of Twitter users have logged in within the last 10 days of the measurement date (*active users*).

- Table 2 summarizes our main results in estimating valid and active population of MySpace and Twitter.

Table 2

OSN	Total	Existing	Active
MySpace	518.0M	298.9M (57.7%)	41.1M (7.9%)
Twitter	86.4M	76.6M (88.6%)	3.0M (3.5%)

5. Conclusions and Open Problems

- We found that *active* population of OSNs is an order of magnitude smaller than the reported (total) population.
- Exploring the smaller portion of *active* OSN users, could result in substantial efficiencies in future OSN measurement.
- We discovered that OSNs' growth significantly varies within different phases and across various systems.

- Determining the underlying social, economic, or technological factors that affect the growth of an OSN population enables us to meaningfully study the structure and evolution of such systems.
- Given the likely migration of users among OSNs, it is interesting to investigate the impact of the emergence of a new OSN on the growth of current popular ones.