

Beyond Friendship Graphs: A Study of User Interactions in Flickr

Masoud Valafar[†], Reza Rejaie[†], Walter Willinger[‡]

[†] University of Oregon

[‡] AT&T Labs-Research

WOSN'09 Barcelona, Spain

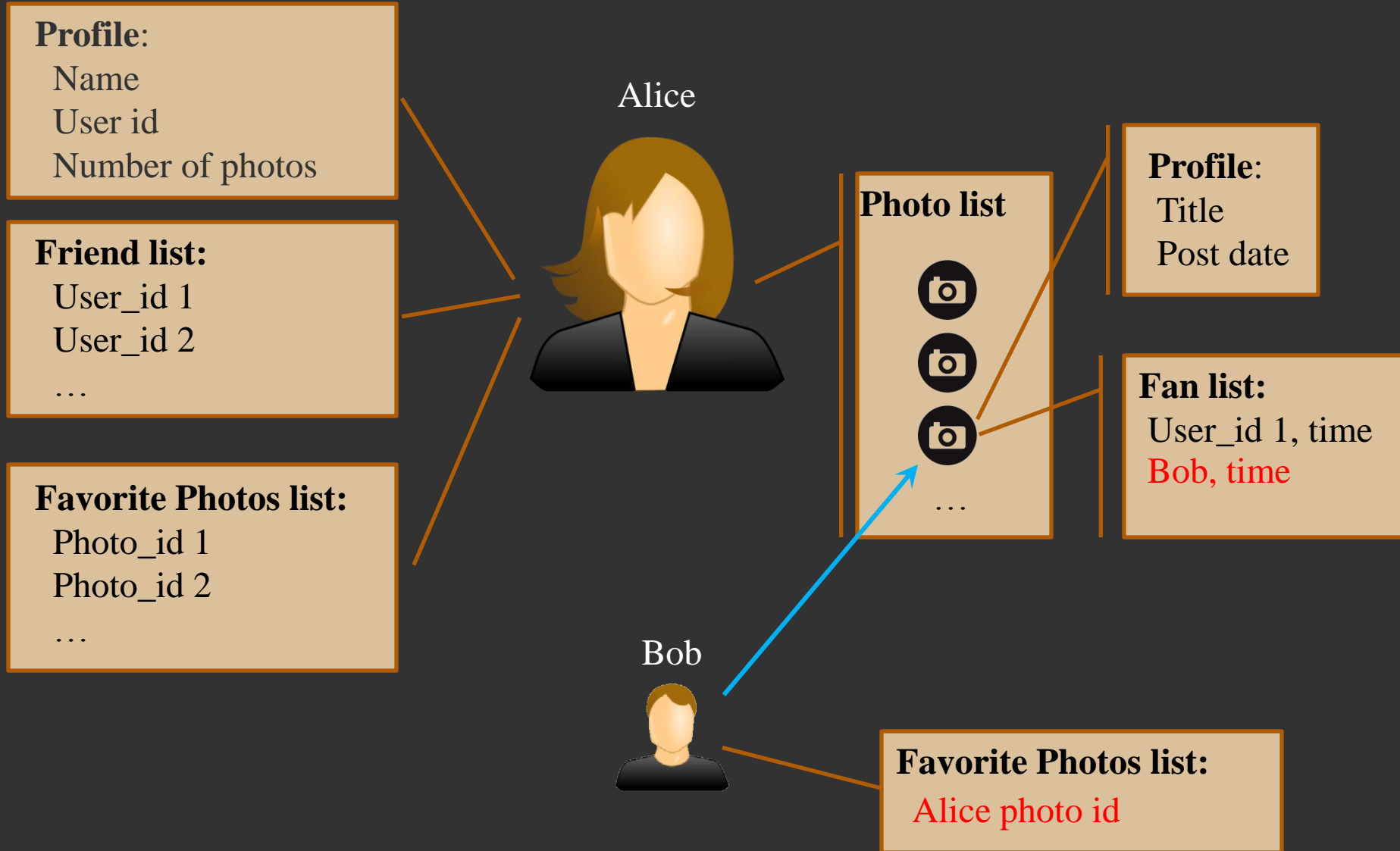
Motivation

- ▣ What does an inferred friendship graph really say about the Online Social Network (OSN) in question?
 - Represents a static, incomplete, inaccurate snapshot of the system
 - Aggregates information over some time period
- ▣ What is the **active** portion of an OSNs inferred friendship graph
 - Requires a notion of “user interaction” and/or of “active user”
 - Inherently dynamic
- ▣ Challenges when moving from inferred friendship to inferred interaction graphs
 - Little (no) incentives for OSNs to make user activity data available
 - Information on user interactions is in general hard to obtain

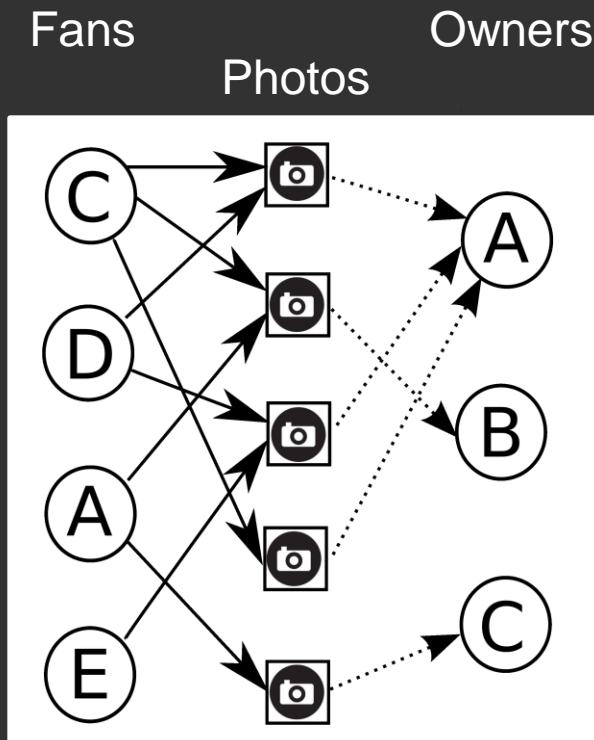
This Study

- ▣ Main focus is on characterizing user interactions in Flickr
 - (Indirect) fan-owner interactions through photos shared among users
 - Based on representative snapshots of fan-owner interactions
- ▣ More specifically, we focus on
 - Extent of user interactions
 - Locality (and reciprocation) of interaction
 - Relationship between user interaction & user friendship
 - Temporal patterns of interactions
- ▣ Related studies
 - Chun et al.'08
 - Viswanath et al.'09 – WOSN'09

User Interactions in Flickr



Indirect Fan-Owner Interactions



- ▣ Users interactions/relations are indirect
 - Through photos
- ▣ Users as owners
 - Photo list (photos they post)
 - “**Favored photos**” (photos they post with at least 1 fan)
- ▣ Users as fans
 - Photos they declare as their “**favorites**”
 - Favorite photo list

Data Collection in Flickr

- ▣ Flickr-specific issues
 - Provides well-documented API
 - Imposes a rate limit for querying the server of 10 queries/second
 - Has well-known user ID format (e.g., 12345678@No2)
- ▣ Data collection method 1 (crawling owned photo lists)
 - ▣ Query server for IDs of all photos owned by a user
 - ▣ Separate query to server for each photo to obtain IDs of all its fans plus associated timing info
 - ▣ Obtain fan-owner interactions from the owner side
- ▣ Data collection method 2 (crawling favorite photo lists)
 - Query server for IDs of all favorite photos of a user along with the IDs of their associated owners with no timing info
 - Obtain fan-owner interactions from the fan side

Data Collection: Method 1

- ▣ Dataset I (Interactions of random users)
 - Leveraged known user ID format
 - Identified about 122K **random** users
 - Extracted **user-specific** information
 - ▣ Profile, friend list
 - ▣ Favorite photo list
 - ▣ Photo list, photo profiles (timing info)
 - ▣ Photo fan lists (timing info)
- ▣ Number of queries needed is on the order of number of photos (slow and inefficient)
- ▣ Dataset I provides a (relatively small) **representative** sample of detailed fan-owner interactions in Flickr (with timing info)

Data Collection: Method 2

- ▣ Dataset II (Interactions of users in main component of friendship graph)
 - Used 122K sampled users as seeds
 - Crawled their friendship graph via their friend lists
 - Identified **main component** (MC) of the friendship graph
 - Collect list of favorite photos and their owners for all MC users and any new user we encounter as an owner of a favorite photo
 - Miss negligible fraction of interactions with singleton users/fans or unreachable fans within MC
- ▣ Number of queries needed is on the order of number of users (efficient and fast)
- ▣ Dataset II provides a large snapshot of indirect fan-owner interactions within MC without any timing info

Dataset 1 vs. Dataset 2

	# photos	#favored	#favorite	#users	#fans	#owners
Singletons	835,970	3,734	24,078	101,210	2,638	1,230
MC users	2,646,139	142,391	532,333	21,127	4,053	5,075

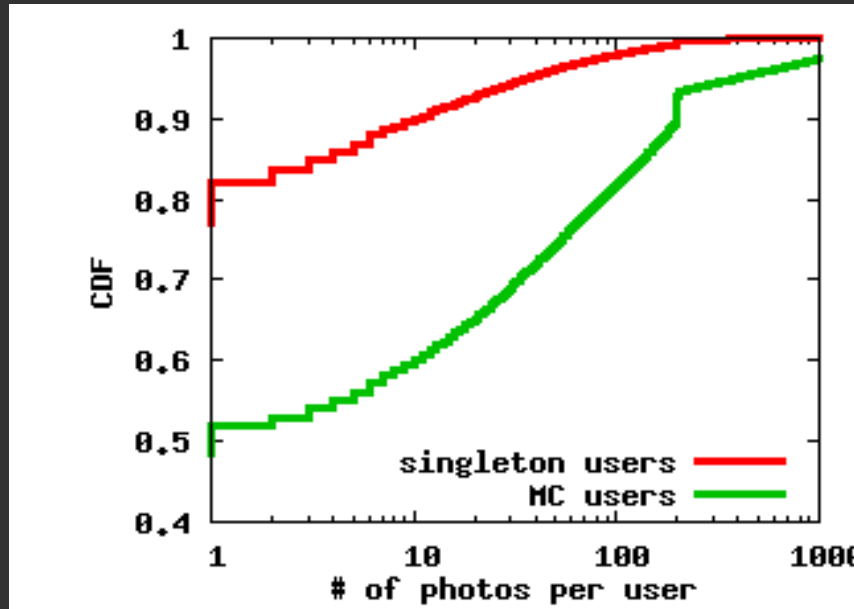
- ▣ Dataset I: small, yet detailed
 - Most of the randomly selected users are inactive singletons
 - MC users are more active than singleton users
- ▣ Dataset II: large, but less detailed
- ▣ Estimate of total user population in Flickr
 - Dataset I: 1 out of 6 of our randomly selected users are in MC
 - Dataset II: Est. total Flickr population = $6 * 4.14M = 25M$ (as of mid-08)

	# favorite photos	# users	# fans	# owners
Interaction in MC	31,495,869	4,140,007	821,851	1,044,055

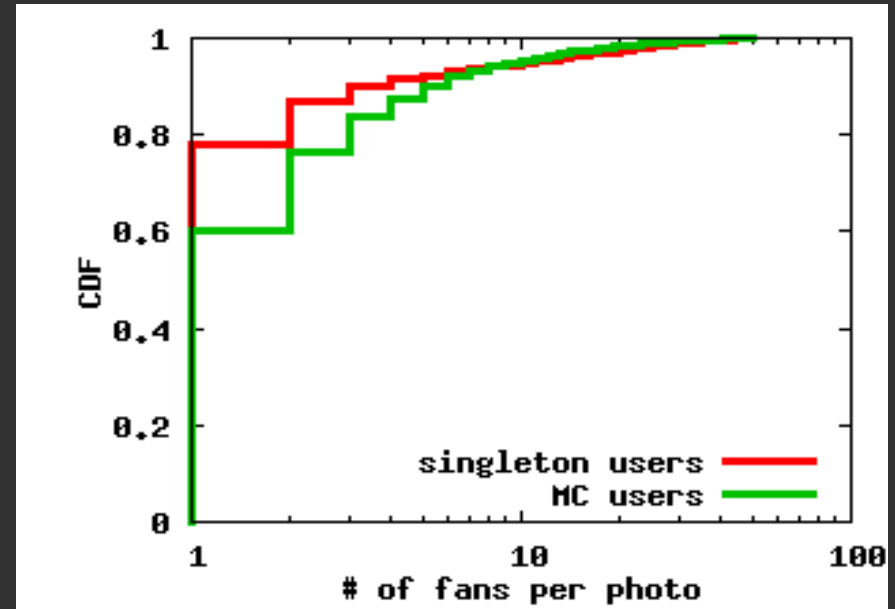
Highlights of our Flickr Study

- ▣ Extent of overall fan-owner interactions
 - More than 95% of fan-owner interactions occur among users in the MC of the Flickr friendship graph
- ▣ Extent of fan-owner interactions in MC
 - The most active users in Flickr form a core in the interaction graph and are responsible for the vast majority of fan-owner interactions
- ▣ Temporal properties of fan-owner interactions
 - There exists no strong correlation between age and popularity of a photo
 - The majority of fans of a photo arrives during the first week after the photo is posted
- ▣ Note: The results are typically based on Dataset I and are validated (where possible) using Dataset II

Extent of Fan-Owner Interactions (I)



- Posted photos
 - Only about 20% of singleton users post 1 or more photos
 - About 50% of MC users post 1 or more photos

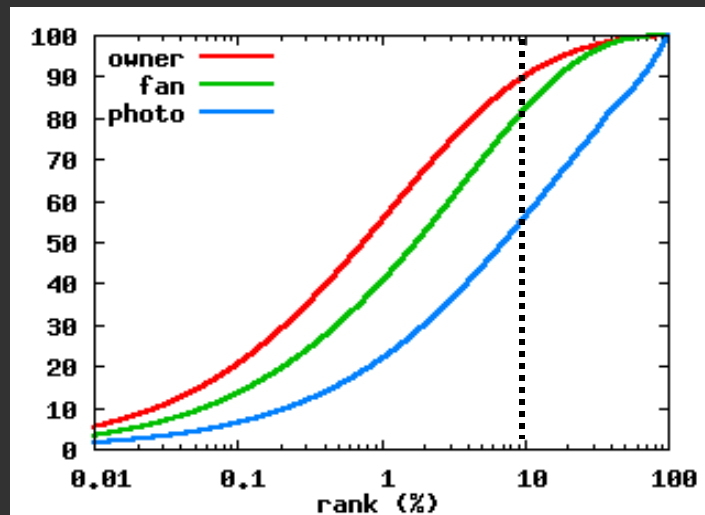
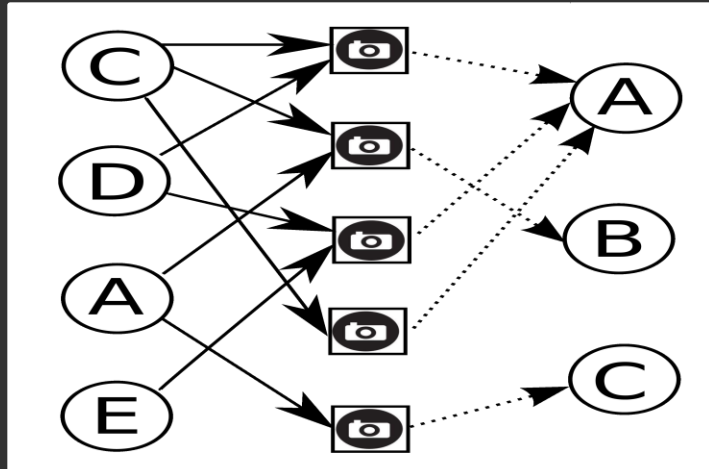


- “Active” photos (at least 1 fan)
 - More than 99% of photos owned by singleton users have no fans
 - About 95% of photos owned by MC users have no fans

Extent of Fan-Owner Interactions (II)

- ▣ Users in their roles as owners or fans of photos
 - “Active” as an owner
 - ▣ At least one posted photo with a fan
 - ▣ More the 97% of fan-owner interactions are associated with active MC owners
 - “Active” as a fan
 - ▣ At least 1 favorite photo owned by another user
 - ▣ More than 95% of fan-owner interactions are associated with active MC fans
- Vast majority (>95%) of interactions in Flickr are among active users in the MC of the friendship graph

Extent of Fan-Owner Interactions in MC(I)

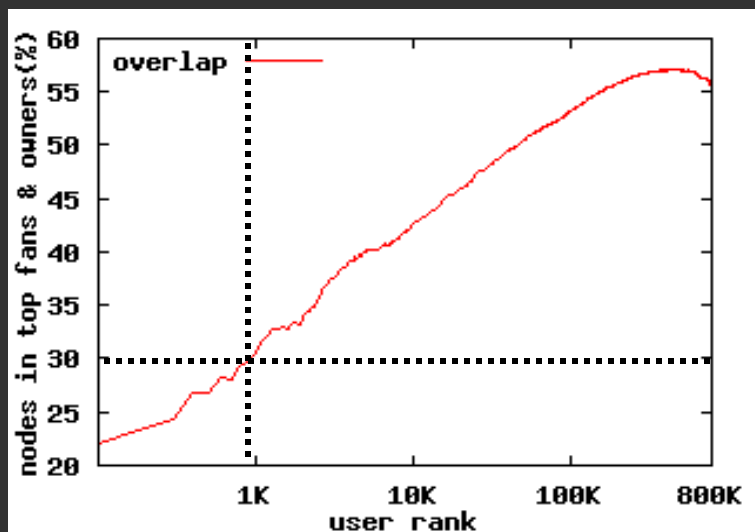


- More detailed view of active users
 - Order owners by indegree
 - Order fans by outdegree
 - Order photos by indegree

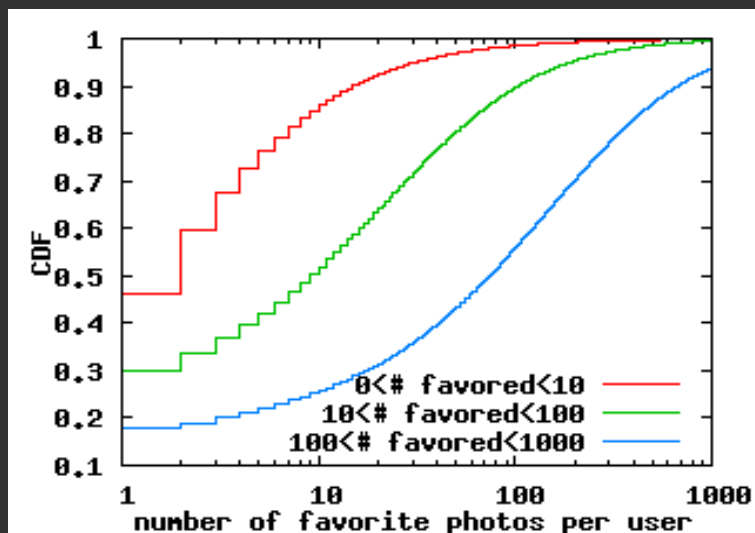
- Top 10% of fans are responsible for 80% of interactions
- Top 10% of owners are responsible for 90% of interactions
- Top 10% of photos are responsible for only about 50% of interactions

➤ The top 10% fans/owners are responsible for most interactions

Extend of Fan-Owner Interactions in MC (II)



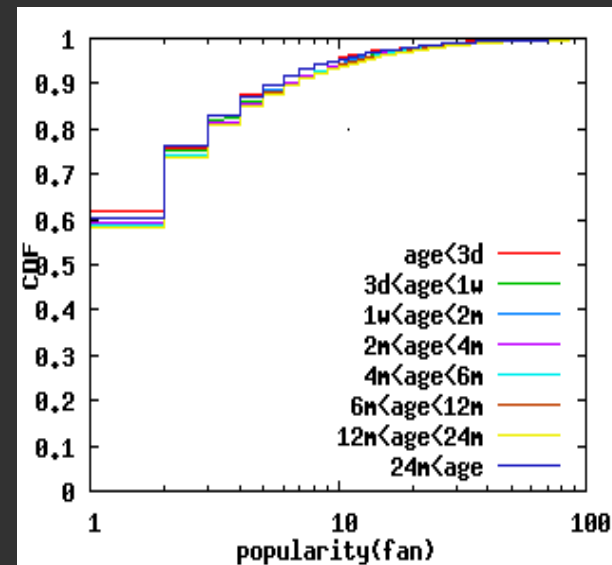
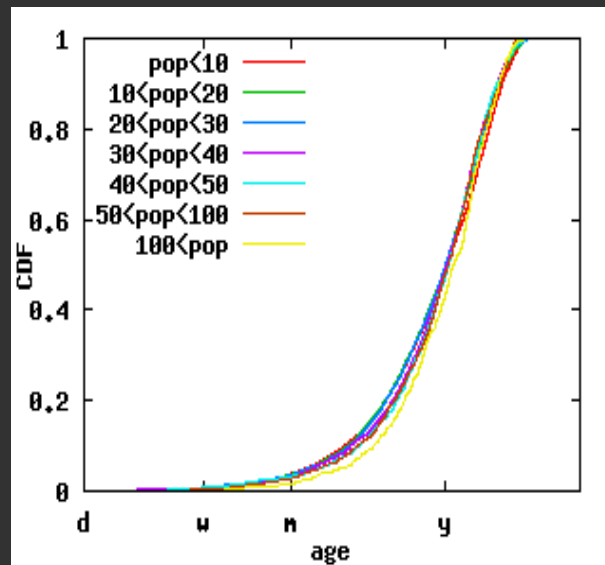
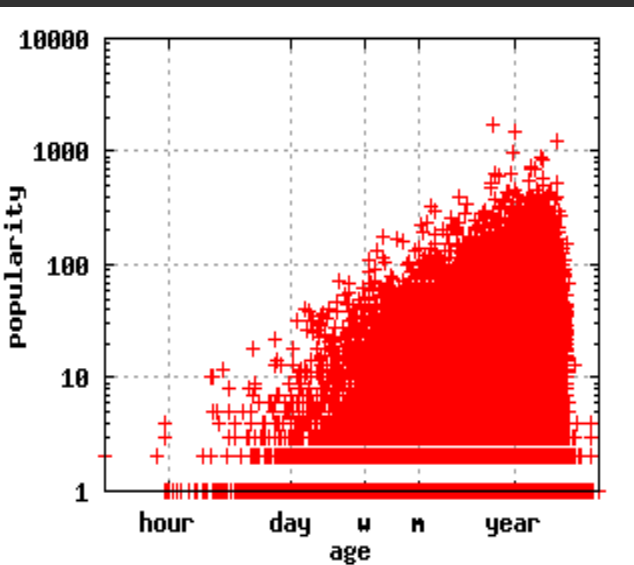
- On the overlap between top active fans and top active owners?
 - E.g., 30% of the top 1K fans are among the top 1K owners
 - Percentage of overlap reaches max of around 57% for top 200K fans



- On the correlation between the level of activity of a user as a fan and as a owner?
 - The most active fans are more likely to be among the most active owners, and conversely.

➤ The top active users form a core of the Flickr interaction graph

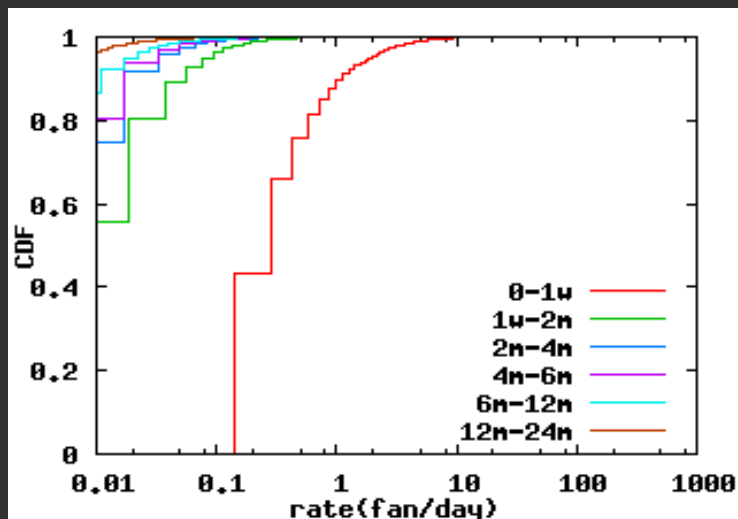
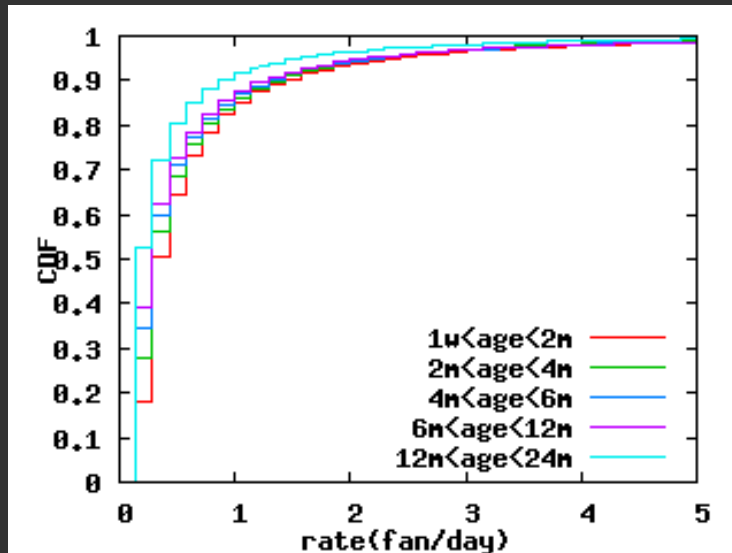
Temporal Properties of Interactions (I)



- Age of a photo vs. popularity
 - Range of popularity widens with age
 - Distribution of photo age does not the photo's popularity
 - The distribution of the popularity of a photo does not depend on its age
- Explanation?

Temporal Properties of Interactions

(II)



- In terms of fan arrival rate of photos, what matters is not the age of the photo ...
 - Age of the photo does not have much effect on the distribution of fan arrival rate
 - ... but when during the photo's lifetime the fans arrived
 - Fan arrival rate in the first week is an order of magnitude larger than during other periods
- Most photos receive most of their fans during the first week after their posting

Conclusions

- ▣ Discussed 2 measurement methodologies for collecting fan-owner interactions in the Flickr OSN
- ▣ Presented initial study of fan-owner interaction in Flickr
 - Most of the users are inactive (as defined in this work)
 - More than 95% of interactions occur in MC of the friendship graph
 - Top 10% of owners (fans) in MC cause 90% (80%) of all interactions
 - There is significant overlap between the top owners and top fans and these users form a core of the Flickr interaction graph
 - Most photos receive most of their fans early on (during first week)
- ▣ Bad news – good news
 - Inferred friendship graphs say little about user interaction/dynamics
 - Observed concentration of “activity” is promising for measurements and studying dynamics

Future Work

- ▣ Leverage the observed concentration in the user interaction graph for measurements
- ▣ Characterization of other types of interactions in other OSNs
 - Messaging in Twitter
 - Video-tagging in YouTube
- ▣ More detailed study of user interaction patterns and their dynamics
 - Multi-scale (in time and space) analysis of interaction graphs
 - Idea: slow (temporal) dynamics at coarse (spatial) scales
- ▣ Understanding underlying causes for observed interaction patterns

Thanks!

Questions?

Website

<http://mirage.cs.uoregon.edu/OSN>

Contact for code and data:

Masoud Valafar

masoud@cs.uoregon.edu