SIGCOMM WOSN 2008

Characterizing Social Cascades in flickr

Meeyoung Cha
MPI-SWS

Alan Mislove
MPI-SWS

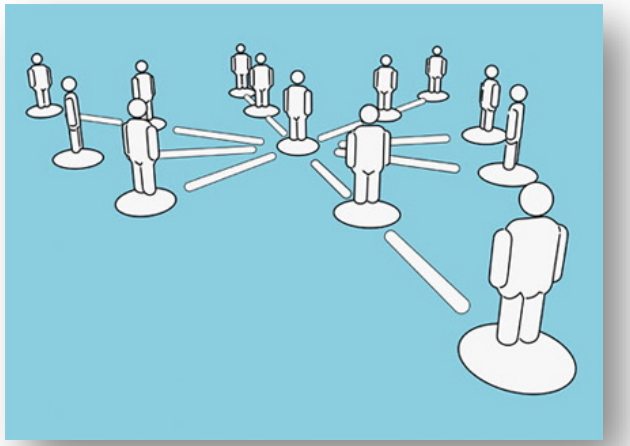
Ben Adams
MPI-INF

Krishna P. Gummadi
MPI-SWS

meeyoung.cha@gmail.com

Online social networks

- OSN websites are popular, e.g., Flickr, Facebook, Orkut
- Used for a variety of information propagation purposes
  - Viral marketing, political campaign, content sharing, launch of movie trailers, product promotions, etc.



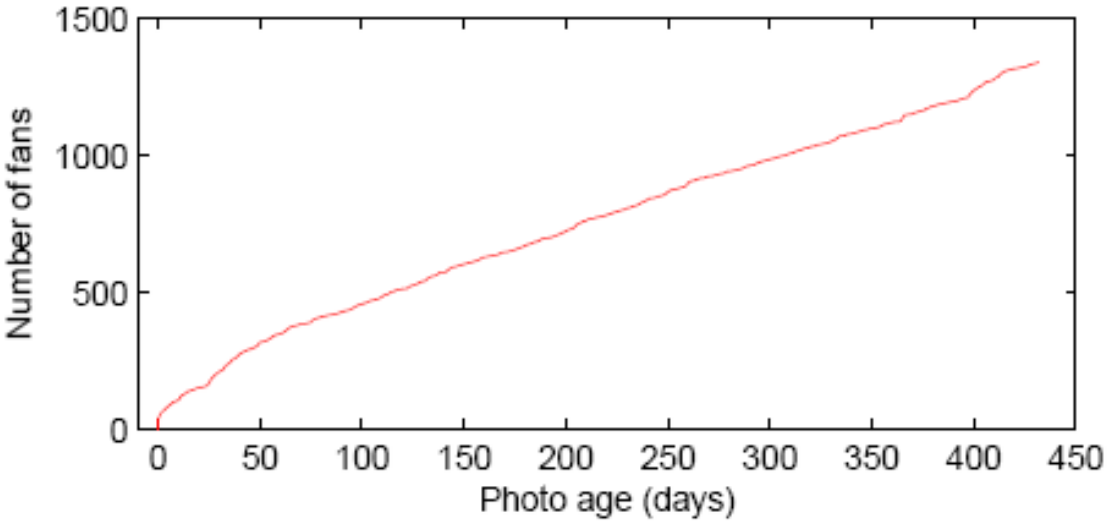
- **How does information propagate in OSNs?**

cnet.com

Information propagation in Flickr

- Growth of fans of a popular Flickr photo

Fire Canoe #2



- **How did the fans get to know of this picture?**

Fire Canoe #2 by Peter Bowers

Mechanisms of information propagation

- Featuring (front page, hotlists)
- External links
- Search results
- Links between content
- Online social links

Key challenge: Gathering the data

- **Crawled a substantial fraction of Flickr social network**
  - 2.5*M* users and 33*M* friend links
    (in its largest weakly connected component)
  - Repeated the crawls for 100 consecutive days

- **Gathered Flickr users' bookmarked pictures**
  - Users bookmark their favorite pictures
  - 34*M* bookmarks of 11*M* distinct photos uploaded by users

Part1.
Measurement
methodology

Part2.
Analysis of
spreading patterns

Part3.
Modeling
social cascades

How to identify information flow through social links?

- **Did a particular bookmark spread through social links?**
  - **No:** if a user bookmarks a photo and if *none* of his friends have previously bookmarked the photo
  - **Yes:** if a user bookmarks a photo *after* one of his friends bookmarked the photo

What role do social links play?

- Conducted preliminary analysis for very popular photos

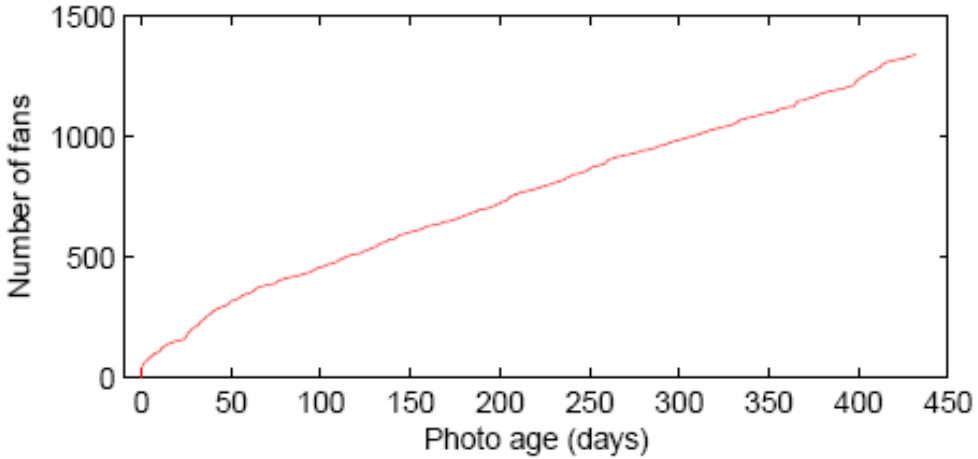
	Total	Through social links
# photos	*1,180*	
# bookmarks	*171,131*	

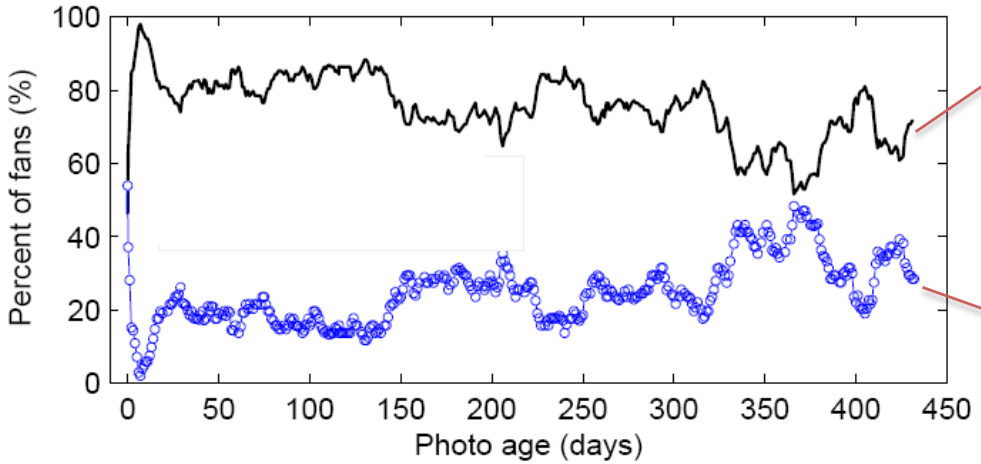
  - On-going work on further analysis of the data

- **42% of bookmarks** propagate through social links
- **The role of friend links in information spread crucial**

Pattern 1: steady increase

- **75%** of bookmarks through social links





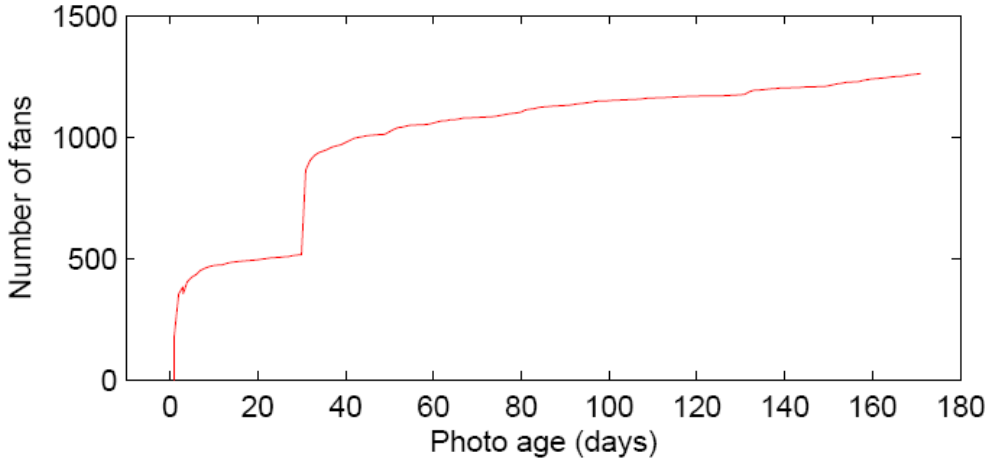
Found through social links

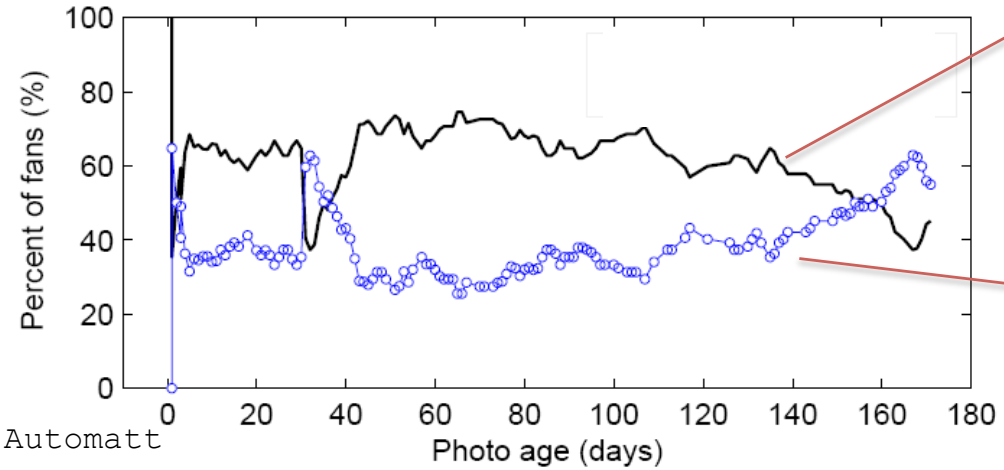
Through other mechanisms

Fire Canoe #2 by Peter Bowers

Pattern 2: surge increase

- **60%** of bookmarks are through social links
- At surges, more bookmarks are from other mechanisms





Found through social links

Through other mechanisms

Midtown Shadow by Automatt

Bookmarks cascade through OSN

- Popularity evolves over time with **different patterns**
- **Significant bookmarks** are through **social links**

- We call the information propagation through social links over time as the **social cascade**

Part1.
Measurement
methodology

Part2.
Analysis of
spreading patterns

Part3.
Modeling
social cascades

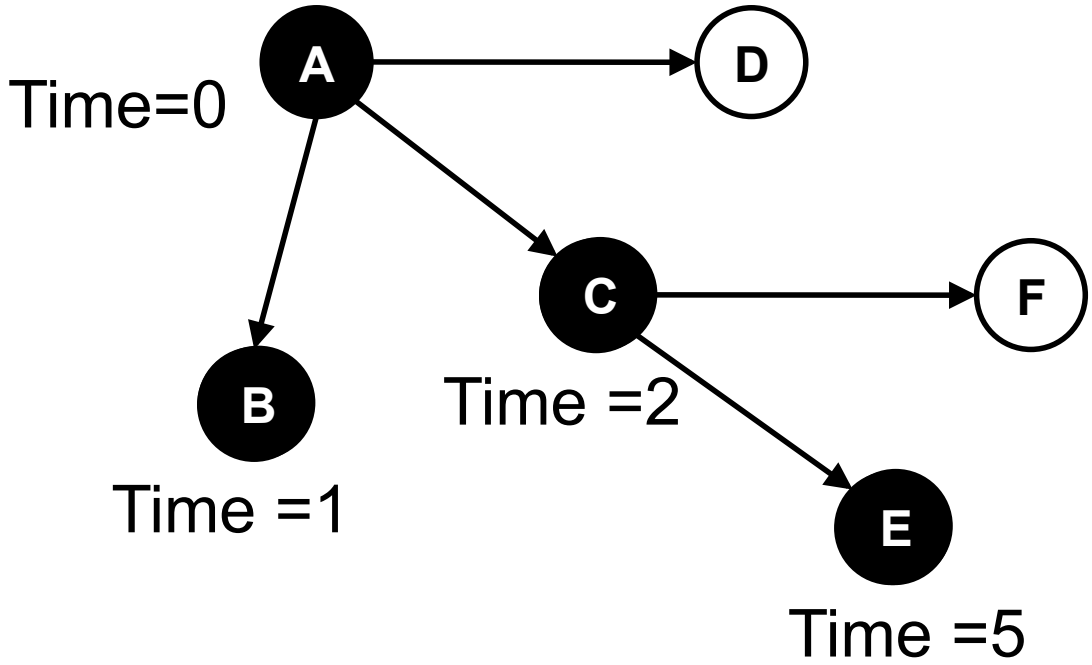
Modeling social cascades

- **Why do modeling?**
  - Help us understand how information spread better
  - Can predict and estimate near-future trends
  - Useful for viral marketing

- **Can existing models characterize social cascade?**

Can existing epidemiological models describe social cascade?

- **Photos propagate through OSN—like diseases spread over offline human contact network**



Epidemiological Framework

- **The basic reproduction number** or **R0**

  - the expected number of new infections by the origin
    - If R0>1, disease spreads out
    - If R0<1, disease fizzles out
    - If R0=1, critical epidemic threshold

- Known R0s: HIV [2,5], Measles [12,18]
- R0>1 is a success case in viral marketing

Tested if epidemiology can be applied to social cascade

- **Empirical counting of R0**

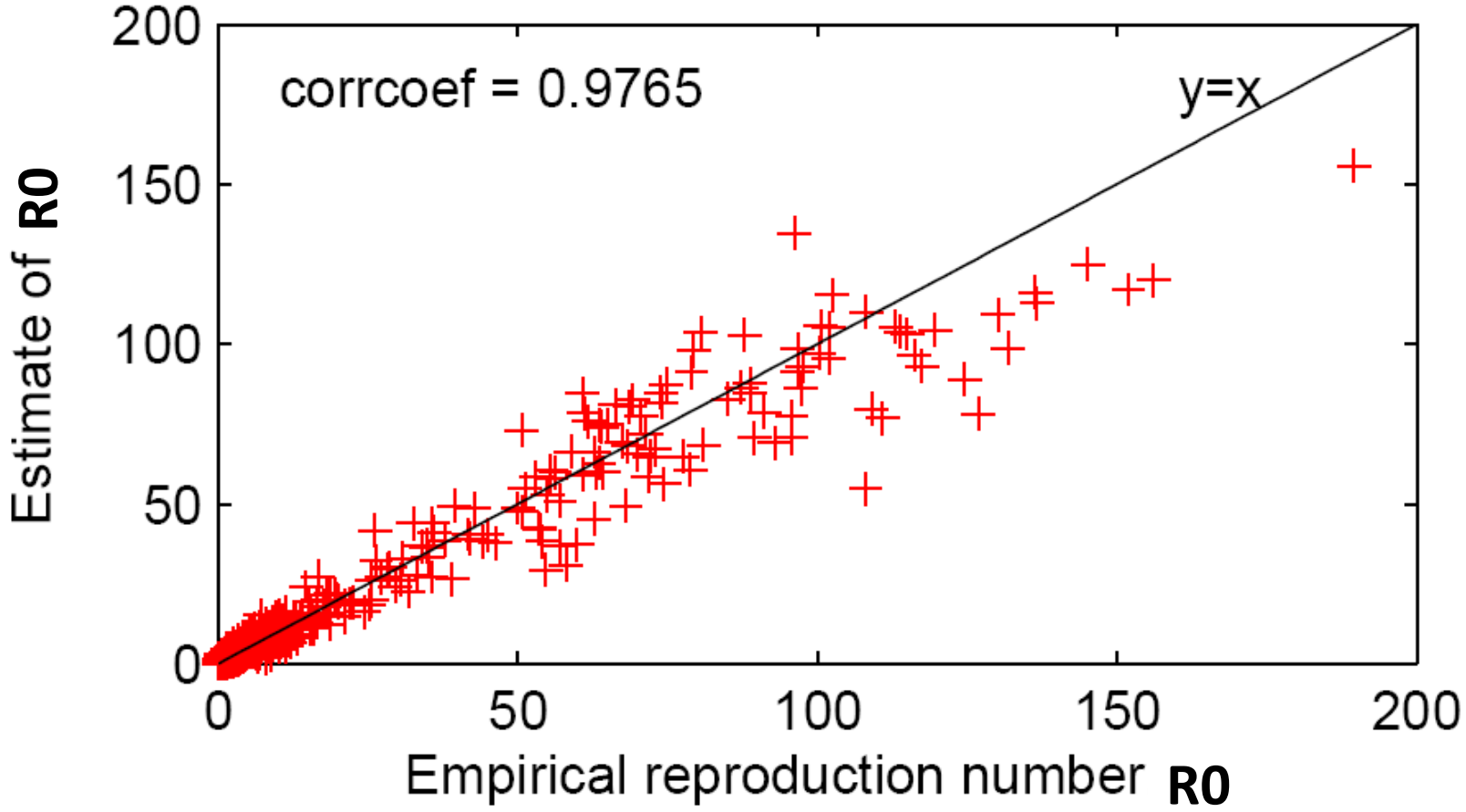
  - For each fan, count how many friends further bookmark the same photo. Average the count.

- **R0 from existing theory** (May-Lloyd-2001)

  - Premise: diseases have unique infection probabilities

Online cascade like infectious diseases

- Existing framework fits perfectly for popular photos



Social cascade has a strong correlation to epidemiology

- Finding: offline spreading of diseases can describe online information propagation through social links

- Potential uses: Potential to predict the spread of photos in other online social networks like Facebook and Orkut

Summary

- **The first work to investigate the role of OSN in information propagation using real traces**
- **Significant fraction of bookmarks from social cascades**
- **Epidemiological framework to be used to model social cascade and make prediction for marketing purposes**