A measurement-driven analysis of information propagation in the flickr social network
Overview

I. About Flickr
II. Data collection methodology and network topology
III. Picture popularity
IV. Evolution of pictures over time
V. Social Cascades
VI. Summary
VII. Weaknesses and Problems
Founded: 2004

Acquired by Yahoo! In 2005

Photo sharing site with social networking feature

More than 4 billion photos
I. Flickr overview – Front page

G'day yxcvbnm123mnbcxy!
Now you know how to greet people in Australian!

» Your Photostream

» Upload Photos & Video

» Your Contacts

» Explore / The Commons

Flickr is more fun when your friends use it too! Try a search or invite people so you can keep up with their latest photos.

Want to go ad-free? Get Pro!

Flickr Blog
Posted 05 Nov 09

The Flickr Collection on Getty Images has been growing and growing since it launched back in March – with a princely figure of nearly 60,000 images in...

Have you read our Community Guidelines?
It's the "Do's" and "Don'ts" of Flickr citizenship.

Holy smokes! That's cheap!
Get a year of pro for $24.35

Bookstore quality books, by you
I. Flickr overview – Front page

Über smichels pro
Fotostream

Kontakte (13)
- thais_agro
  Kein Name angegeben
- p.c.m.
  Kein Name angegeben
- maya.srikanta
  Kein Name angegeben
- Andrei Lintu
  Kein Name angegeben
- Lady_Geliebte
  E. Amada
- george_2006
  Kein Name angegeben
- chriszim24
  Kein Name angegeben
- abroschart
  Kein Name angegeben
- d.adriana
  Kein Name angegeben
- kberberi
  Klaus Berberich
- josixp
  Josiane Xavier ...

Gruppen (14)
- Jellyfish!
  5719 fotos, 1963 Mitglieder
- Baby Beluga at the Vancouver Aquarium
  112 fotos, 37 Mitglieder
- Cemeteries in Israel
  476 fotos, 30 Mitglieder
- OLPC: One Laptop Per Child
  1467 fotos, 248 Mitglieder
- Brasil/Brazil
  132747 fotos, 6253 Mitglieder

Empfehlungen

11/11/2009

Thomas Dackweiler
I. Flickr overview – Explore page

Flickr tabs have been hard at work creating a way to show you some of the most awesome content on Flickr.

We like to call it interestingness.

Explore the vigorously by choosing a point in time.

Select a month:

Choose:

You can also jump into a calendar view of this month.

Other places to explore include:

- a map of the world
- the people tags page
- Camera Finder
- viewing the slideshow or most recent uploads
- video on Flickr

11/11/2009

Thomas Dackweiler
I. Dissemination mechanisms

Search results:
Search engine provides content meta data
(Titles, tags and description of photos)

Featuring:
Front page
Explorer page

Links between content:
Links between pictures to navigate
„Sets“: Groups of similar pictures
„Pools“: Different user, but same themes

External Links:
Photos are reachable from external sites, blogs, emails,...

Social network:
Users can share contents with other users (friends, favorites,...)
II. How are informations spread over social networks?

How widely does information propagate?
How quickly does information propagate?
What is the role of word-of-mouth exchanges between friends?

→ Collect and analyze large-scale traces of information dissemination

Crawls of the favorite markings ...
... of 2.5 million users
... 33 million links (25% of the entire network)
... on 11 million photos
... over 104 consecutive days
II. Data collection methodology

Collect …

… the evolving state of the network (Part I)

… the evidence of information propagation between users (Part II)
II. **Data collection methodology**

Crawl a significant subset of the network

→ Select random user and follow all of the friends links in forward direction (snowball sample)

→ Get a **social network graph**
II. Social network graph
II. Social network graph

- Outdegree: 3
- Indegree: 2

- Flickr users
  - Seed node
  - Friends (1-hop)
  - Friends of friends (2-hop)
  - Fr. of fr. of fr. (3-hop)
Flickr social (directed) graph:

Each node = Flickr user
Each edge = Friend link
Outdegree = # Friends a user points to
Indegree = Inverted direction

68% of the links are bidirectional
Pearson's correlation coefficient: 0.76
II. Node degree distribution

55% nodes have just 1 outgoing link
Average outdegree: 14
Highest number of outgoing links: 26,342
Maximum outdegree > maximum indegree
II. Structural properties

Maximum path length between 2 nodes: 27

Average path length: 5.67

Clustering coefficient:
(How tightly the neighbors are connected)
Poorly-connected nodes: 0.05-0.10
Well-connected nodes: 0.2-0.4
II. Implications for information flow

Users with high in- and outdegree can potentially distribute informations more widely (expectation)

Most users are separated by only a few hops
→ Only short network paths

→ Good pre-conditions for wide-spread dissemination
II. Data collection methodology

How to capture the dynamics?

Launch a complete crawl of this graph every day

→ Newly created/removed friend links or users are recorded
II. Data collection methodology

How to collect evidence of information flow?

Get information on the favorite photos (exact timestamp) → Recreate favorite marking events

Also known: State of the network at the time the marking took place → Filter the influencing factors
II. Problems

- This methodology does not consider deleted favorite marking
- How did the user come upon this favorite-marking?
- Network can not be manipulated: No changes possible and no tests in a controlled environment
What are the various popularity metrics?
III. Picture popularity distribution

→ Only a few pictures achieve high popularity and are spread widely
Examine the relationship between number of views, comments and fans:

Correlation coefficient (from -1 to 1):

-1 → Negative linear relationship
0 → No linear relationship
1 → Positive linear relationship
III. Correlation coefficient

# views ↔ # comments: 0.13 (not strongly correlated)

# views ↔ # fans: 0.23

# comments ↔ # fans: 0.60 (highly correlated)
Correlation between views and fans for popular pictures:

Decreases from 0.23 to 0.21 for pictures with >100 fans and from to 0.13 for pictures with >1,000 fans

→ Weaker correlation

Reasons:
• User find many pictures uninteresting
• Favorite-marking or comment only as registered user possible
III. Picture popularity

How widely does favority-marking propagate through the social network?
III. Picture popularity

→ Compare the most popular pictures in local neighborhoods with a global hotlist of pictures (Part 1)

→ Examine the distribution of fans as a function of their distance from the uploaders (Part 2)
III. **Compare local and global hotlist (Part 1)**

Pick 250 users randomly who have at least 1 photo

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>1-hop</td>
<td>6</td>
<td>1,377</td>
<td>1,379</td>
<td>2,816</td>
</tr>
<tr>
<td>≤ 2 hops</td>
<td>2,785</td>
<td>199,330</td>
<td>174,100</td>
<td>290,671</td>
</tr>
<tr>
<td>≤ 3 hops</td>
<td>283,001</td>
<td>1,050,400</td>
<td>938,880</td>
<td>1,159,636</td>
</tr>
<tr>
<td>≤ 4 hops</td>
<td>880,051</td>
<td>1,625,482</td>
<td>1,563,500</td>
<td>1,667,054</td>
</tr>
</tbody>
</table>

→ Identify the top 100 pictures from their neighborhood (4-hops max.) = local neighborhood
III. Compare local and global hotlist (Part 1)

Compare:

Top 100 pictures based on the number of fans from that region

↔

Globally popular top 100 pictures

→ Determine the „overlap“
III. „Overlap“

1-hop neighborhood:
233 out of 250 local regions had no overlap between both hotlists
largest overlap was 19 pictures
III. „Overlap“

Avg. overlapped pictures:

≤ 2-hops: 8
≤ 3-hops: 39
≤ 4-hops: 70

→ Wider neighborhood boundaries → More overlaps
Different pieces of information are popular among the different regions

Close neighborhood $\rightarrow$ Pictures localized
Wider neighborhood $\rightarrow$ High overlap

$\rightarrow$ Information are reachable within a few hops
III. Distance from fans to uploaders (Part 2)

• Fraction of fans that are 1, 2, 3 or more hops away from the uploader

• Fraction of nodes that become fans of the pictures
### III. Percentage of fans in k-hops distance from uploaders

<table>
<thead>
<tr>
<th># Fans</th>
<th>1-hop away</th>
<th>2-hops away</th>
<th>3+-hops away</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>60</td>
<td>31</td>
<td>9</td>
</tr>
<tr>
<td>6-100</td>
<td>55</td>
<td>32</td>
<td>13</td>
</tr>
<tr>
<td>101-300</td>
<td>43</td>
<td>42</td>
<td>15</td>
</tr>
<tr>
<td>301-500</td>
<td>37</td>
<td>46</td>
<td>17</td>
</tr>
<tr>
<td>501-</td>
<td>32</td>
<td>49</td>
<td>19</td>
</tr>
</tbody>
</table>

(3 million pictures and 10 million favorite-markings)
Problem: High differences in k-hop neighborhood sizes → Hard to compare the distribution for different photos

Solution: Calculate the fraction of each k-hop neighborhood that became fans

How?: Visit each user who is k-hops from the uploaders and count how many of them have marked the picture as a favorite
III. Percentage of fans for uploaders out of k-hop friends

<table>
<thead>
<tr>
<th># Fans</th>
<th>1-hop away</th>
<th>2-hops away</th>
<th>3-hops away</th>
</tr>
</thead>
<tbody>
<tr>
<td>101-300</td>
<td>1.77</td>
<td>0.08</td>
<td>0.001</td>
</tr>
<tr>
<td>301-500</td>
<td>1.39</td>
<td>0.12</td>
<td>0.004</td>
</tr>
<tr>
<td>501-</td>
<td>1.14</td>
<td>0.17</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Propagation of favorite marked photos is limited and photos are mostly spread in direct neighborhood.
Different sets of photos are popular in different parts of the network

Fans of a picture are closely located to the uploaders

Information do not propagate widely in this social network
How quickly do fans mark photos after their upload?
IV. Four different patterns of growth in popularity

3 distinct growth phases: active-growth, surge-increase and sluggish
Consider photos which are older than 1 year and photos which are older than 2 years

→ Filter those pictures with more than 100 fans

→ 5,346 and 897 photos, respectively
IV. Long term trends in popularity growth

→ Active rise after a few days
→ After 10-20 days: Steady linear growth
→ Almost 40% fans acquired after the first 100 days
→ Steady growth instead of exponential
→ Different ways to get popular
V. Social Cascades

Focus on the dissemination of content via links in Flickr

Social cascade = Information exchange via word-of-mouth

Problem:
No knowledge about how the user found photos
→ Use of heuristics
v. Social Cascades

Definition: „User A found photo P from user B“

... if there exists a user B who is a friend of A such that:

- B also marked P as a favorite
- B included P on his favorites before A
- B was a friend of A before A made photo P his favorite

→ Photo is propagated from B to A via a social link
v. Social Cascades

(Time X before time Y)
v. Social Cascades

2 Problems:

- Multiple friends of A could have found the picture → A received the information from all of them

- Uploaders can not be fan of their own photos → They are fans by default
v. Social Cascades

Use the data from the 104 days crawl

Only those photos that were uploaded during this time
(entire popularity history)

→ 10,025,185 favorite markings and 3,055,361 pictures
v. Social Cascades

How many favorites were marked by social links?
v. Social Cascades

→ Uploaders play an important role in the social cascades of less popular pictures
v. Peer pressure in photo marking

Favorite-marking influenced by the number of friends who already have marked that photo?
v. Peer pressure in photo marking

Probability of becoming a fan increases with the number of friends who already marked that picture

→ Peers influence favorite marking
v. Time taken for social cascade hops

How long takes it to propagate information along each hop of the social cascade?

Table: Exposure time in days prior to favorite marking

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>3,685</td>
<td>190,353</td>
<td>0</td>
<td>60</td>
<td>140</td>
<td>904</td>
</tr>
</tbody>
</table>

• 35% of fans found their favorite marked picture within a week after friends' favorite marking (not shown in table)

• 50% took over 60 days

• 140 days is the average delay

→ It takes long time to spread across each link in the network
v. Social cascades: Observations

→ >50% of the favorite marking through social links

→ Individuals take a long time (3 to 5 months) to mark a photo as a favorite that was already marked by a friend
Study:

2 different pictures with diverse growth pattern

Object:
Look for evidence of social cascades in the growth of popularity

v. Social cascades: Related Work

"Characterizing Social Cascades in Flickr“, Meeyoung Cha,
First workshop on Online social networks (2008)
v. Social cascades: Related Work

(a) Growth of fans, photo A

(b) Growth of fans, photo B

(c) Breakdown of new fans, photo A

(d) Breakdown of new fans, photo B
v. Social cascades: Related Work

Observations:

„Social cascade” group accounts for over half of new fans for both photos

→ social network plays a significant role in content dissemination

Dominance of the “social cascade” group over the “other” group switches during the two popularity surges exhibited by photo B

→ Due to other mechanisms (i.e., linking from external sites or featuring)
Highly connected nodes are more likely to disseminate pictures

But: They are also likely to replace that pictures with a different favorite very quickly

→ High transmission rate, but short duration of „infection“
Vice versa for poorly connected nodes

Most efficient: Intermediate Connectivity
vi. Overall summary

- Most fans are only a few hops from the uploader away
- Pictures spread slowly throughout the network
- Even popular photos do not spread widely throughout the network
- Over 50% of users find their favorite pictures from their friends

→ Contradict the expectations!

Explanations? ...
Model of viral marketing:
Uploader is often the only seed who distributes a photo and beyond his neighborhood there are no further distributor

Homophily:
People who like each other's photos tend to become friends and people who are friends tend to like each other's photos
Delay:
Related to the rate at which users are exposed to the new pictures friends marked

→ Only small number of updates of recently uploaded pictures of friends

→ Also depends on the login-frequency
VII. Weaknesses and Problems

Use of heuristics:
We do not know exactly which mechanisms are responsible for which users' favorite-marking

We do not know WHEN the user viewed a photo
(We only know the total view counts per photo)
Rarely focus on pictures with less than 100 fans

→ Different growth pattern
  → Fast dissemination
  → Limited fan population in the early stage, dormant after the first few month

→ No conflict with their initial expectation concerning pictures with less fans
VII. Weaknesses and Problems

No focus on other dissemination mechanisms: „Front page“, „Explore page“, search results or external links
Thank you for your attention!