A Measurement-driven Analysis of Information Propagate in the Flickr Social Network

Instructor: Augustin Chaintreau
Presenter: Fang Qian
Introduce the Flickr

- Flickr founded in 2004 and acquired by Yahoo! In 2005, is an online social network for sharing photos.

Flickr: image and video hosting website, web services suite and online community.

This service is widely used by bloggers to host images that they embed in blogs and social media.
Three Key Questions

- How widely does information propagate in the social network?
- How quickly does information propagate?
- What is the role of word-of-mouth in the network?
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
What are the features of Flickr?

- Create friend relationships with one another and share photos.
- Users can create networks of friends, join groups, send messages to other users, comment on photos, tag photos, and choose their favorite photos.
- Users have privacy control over photos they upload, allowing photos to be classified as either private, visible only to their friends or the default public.
This paper focuses on two features according to two types of links

- The list of user’s favorite photos (favorites means links to favorite photos)
- The list of a user’s contacts (contacts means the links to other users)
Data description

Crawled the Flickr network graph once per day for the period of 104 consecutive days from Nov 2- Dec 3, 2006 and Feb 3-May 18, 2007. For 2.5 million users and 33 million links, and estimated 25% of the entire network. Here, we collected information about 34 million favorite-markings over 11 million distinct photos.

Table 1: Summary of Flickr data set

<table>
<thead>
<tr>
<th>Time period</th>
<th>104 days (starting Nov 2, 2006)</th>
</tr>
</thead>
<tbody>
<tr>
<td># Links</td>
<td>17,034,807 to 33,140,018</td>
</tr>
<tr>
<td># Users</td>
<td>1,620,392 to 2,570,535</td>
</tr>
<tr>
<td># Photos</td>
<td>11,195,144</td>
</tr>
<tr>
<td># Favorite marks</td>
<td>34,734,221</td>
</tr>
</tbody>
</table>
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
2. Network Topology and Picture Popularity

- Social network topology

Figure 1: Node degree distribution

The Flickr social structure exhibits properties that promise wide-spread dissemination of popular information throughout the network.
• Picture popularity

There are three metrics for ranking picture: views, fans, or comments.

Figure 2: Picture popularity distribution
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
3. Topological Distribution of Picture Popularity

- Local versus global picture popularity
  - To well understand this we determine the most popular pictures (call this set a *hotlist*) in several local neighborhoods and compared them with the global hotlist of pictures.

<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>
For each of neighborhoods, we identified the top 100 pictures based on the number of fans from that region and compared the list with the global popular top 100 pictures. The number of photos that appear in both lists, the “overlap”.

Figure 3: Resemblance in local and global hotlists
Two key observations:

- First, the difference between global and local hotlists indicates that different pieces of information are popular among the different social network regions.
- Second, information is reachable within few hops. (focus on 4-hop neighborhood cover 36% of the entire graph.)
Distance from fans to picture uploaders

Two quantities:

(a) the fraction of fans that are located 1, 2, or 3 or more hops away from the uploaders.

(b) the fraction of nodes that become fans of the pictures.

<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>
Different sets of pictures are popular in different parts of the social network and that photo fans are closely located to the uploaders. It also means information does not propagate widely in the Flickr social network.

<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>
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
Pattern of popularity growth

• We examined the growth patterns in the 30 most popular pictures in Flickr.

• active-growth, surge-increase, and sluggish
Figure 4: Popularity growth pattern of sample photos
Long-term trends in popularity growth

- How does photo popularity evolve over a long period of time?
- Which growth pattern is dominant in a time period of a year or longer?

Ways: we examined the aggregate growth patterns of two subsets of photos: photos that are older than 1 year and photos that are older than 2 years.
Figure 5: Popularity growth patterns over a long term period
• Key observations from figures about the long-term trends in popularity growth of pictures.
  ◦ First, many photos do show an active rise in popularity during the first few days after they are uploaded.
  ◦ Second, after the first few days, most pictures enter a period of steady linear growth.

• Trends in popularity growth for less-popular pictures (pictures with fewer than 100 fans)
  ◦ Attract their limited fan population early on during their lifetime, and they become dormant after the first few months.
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding Remarks
5. Information Propagation via Social Links

- Dissemination mechanisms
  - Some important mechanisms to find pictures
    - Featuring.
    - Search results.
    - Links between content.
    - External links.
    - Social network.

  We focus on the dissemination of content via social network links in Flickr.
Identifying social cascades

- **Social cascades**: information travel widely through a social network one-hop at a time via word-of-mouth exchanges between friends in the network.

- **Definition**: user A found a photo P through the social network if and only if there exists a user B who is a friend of A such that:
  - B also marked P as a favorite,
  - B included photo P on his favorite list before A included photo P on his favorite list
  - B as a friend of A before A made photo P his favorite
• The role of social cascades

<table>
<thead>
<tr>
<th>Popularity (# Fans)</th>
<th>Total pictures</th>
<th>Total fans</th>
<th>Social cascades</th>
<th>Cascades from uploaders</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td># Photos</td>
<td>Perc.</td>
</tr>
<tr>
<td>1-5</td>
<td>2,704,806</td>
<td>4,328,609</td>
<td>1,517,550</td>
<td>56%</td>
</tr>
<tr>
<td>6-100</td>
<td>346,870</td>
<td>5,121,820</td>
<td>329,029</td>
<td>95%</td>
</tr>
<tr>
<td>101-300</td>
<td>3,502</td>
<td>499,870</td>
<td>3,502</td>
<td>100%</td>
</tr>
<tr>
<td>301-500</td>
<td>154</td>
<td>54,773</td>
<td>154</td>
<td>100%</td>
</tr>
<tr>
<td>501+</td>
<td>29</td>
<td>20,113</td>
<td>29</td>
<td>100%</td>
</tr>
<tr>
<td>Total</td>
<td>3,055,361</td>
<td>10,025,185</td>
<td>1,850,264</td>
<td>61%</td>
</tr>
</tbody>
</table>

• Cascades from uploaders:
  ◦ The fraction of social cascade-based favorite markings that are just 1-hop away from the uploaders.
Peer pressure in photo favorite marking
  ◦ If one has many friends who declare that they like a given picture, is the user more likely to mark the picture as a favorite in the future?

Figure 7: The probability of becoming a fan as a function of the number of friends who are already fans of the same picture.
- Time taken for social cascade hops
  - How long it takes for information to propagate along each hop of the social cascade.

Table 6: 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>

Observations indicate that favorite marking information takes a long time to spread across each link in the social network.
Content

1. Measurement Methodology
2. Network Topology and Picture Popularity
3. Topological Distribution of Picture Popularity
4. Temporal Evolution of Picture Popularity
5. Information Propagation via Social Links
6. Concluding
6. Concluding Remarks

- Social links are the dominant method of information propagation.
- Information spreading is limited to individuals who are within close proximity of the uploaders.
- Spreading takes a long time at each hop.

- As a result, content popularity is offer localized in the network and popularity of pictures steadily increases over many years.
THANK YOU!!!