A FEW CHIRPS ABOUT TWITTER

Balachander Krishnamurthy – AT&T Labs--Research
Phillipa Gill – University of Calgary
Martin Arlitt – HP Labs/University of Calgary
Outline

- What are micro-content networks?
- Methodology
- Characterization
- Conclusions
Micro-content networks

- An average YouTube video is large, 10 MB
- Micro-content network messages are very small (typically < 1 KB)
- One to many communication possible
- Often a publish-subscribe system with control on subscribers
- Senders and recipients can choose how to send/receive messages
Twitter

- Started Oct. 2006
- Allows users to send short messages ("tweets")
  - Max length of 140 characters (compatible with SMS)
- Micro-blogging
- Notion of following (friends) and followers (subscribers) - with permission
- Used to transmit messages during the 2007 California fires, and riots in Kenya
Interfacing with Twitter

Twitter Input Methods
- User Applications
- Facebook
- IM
- SMS
- Web Interface

Twitter API

Web Interface
Twitter API
Facebook

Twitter Output Methods
- Statuses
- SMS
- IM
- RSS
- User Applications
Outline

- What are micro-content networks?
- Methodology
- Characterization
- Conclusions
Methodology

- **Constrained crawl** (67,527 users)
  - Constrained by Twitter API rate limiting
  - Limited to collecting partial set of each user’s friends

- **Metropolized random walk** (31,579 users)
  - Used to validate constrained crawl
  - Previously used for unbiased sampling of peer to peer networks [Stutzbach et al. IMC 2006]

- **Public Timeline data** (35,978 users)
  - Timeline of most recent messages available on demand.
Outline

- What are micro-content networks?
- Methodology
- Characterization
- Conclusions
High order results

- Following vs. followers
  - Relationships not always symmetric
- Different classes of users
  - Not all human
- Number of tweets varies significantly
- Geographic patterns vary
  - Few countries dominate
Characterization

- User relationships
- Properties of tweets
  - What tools are used to post tweets?
  - When are Twitter users active?
  - How many tweets do users have?
- Other properties of Twitter users
  - UTC offsets in the datasets
  - Geographical spread of Twitter
Characterizing user relationships

- “Followers” (people who subscribe to receive your tweets)
- “Following” (people whose tweets you subscribe to)
- Relationships are not necessarily symmetric
User relationships
User relationships - Broadcasters

- News outlets, radio stations
- No reason to follow anyone
- Post playlists, headlines
User relationships - Acquaintances

- Similar number of followers and following
- Along the diagonal
- Green portion is top 1-percentile of tweeters
User relationships - Odd

- Some people follow many users (programmatically)
- Hoping some will follow them back
- Spam, widgets, celebrities (at top)
Characterizing user tweets

- Where do tweets come from?
- When are people tweeting?
- How many tweets do users have?
Where do tweets come from?

<table>
<thead>
<tr>
<th>Crawl</th>
<th></th>
<th>Timeline</th>
</tr>
</thead>
<tbody>
<tr>
<td>%</td>
<td>tweets</td>
<td>source</td>
</tr>
<tr>
<td>61.7</td>
<td>40,163</td>
<td>Web</td>
</tr>
<tr>
<td>7.5</td>
<td>4,901</td>
<td>txt (mobile)</td>
</tr>
<tr>
<td>7.2</td>
<td>4,674</td>
<td>IM</td>
</tr>
<tr>
<td>1.2</td>
<td>792</td>
<td>Facebook</td>
</tr>
<tr>
<td>22.4</td>
<td>14,566</td>
<td>custom applications</td>
</tr>
</tbody>
</table>
When are people tweeting?

- Steady activity during the day with drop-off during late night hours.
Number of tweets per user

![Graph showing the number of tweets per user](image)
Other properties of Twitter users

- UTC offsets
- Geographical spread of users
Comparison of UTC offsets of users between datasets

![Graph showing comparison of UTC offsets]

- X-axis: Offset (Hours) from GMT
- Y-axis: % of Users

Legend:
- Crawl
- Timeline
Geographical presence of Twitter

![Graph showing the geographical presence of Twitter users, with lines representing different regions: Asia-Pacific, South America, Japan, Europe, and North America. The x-axis represents User ID (millions), and the y-axis represents Number of Users (000s).]
Summary

- One of the first large characterizations of Twitter
- Diversity of access methods
- Presence of interesting user-communities (e.g., broadcasters)
- Distinct properties compared to larger OSNs
I'm so glad you agreed to meet in person. There are some things that just can't be said in 140 characters.