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

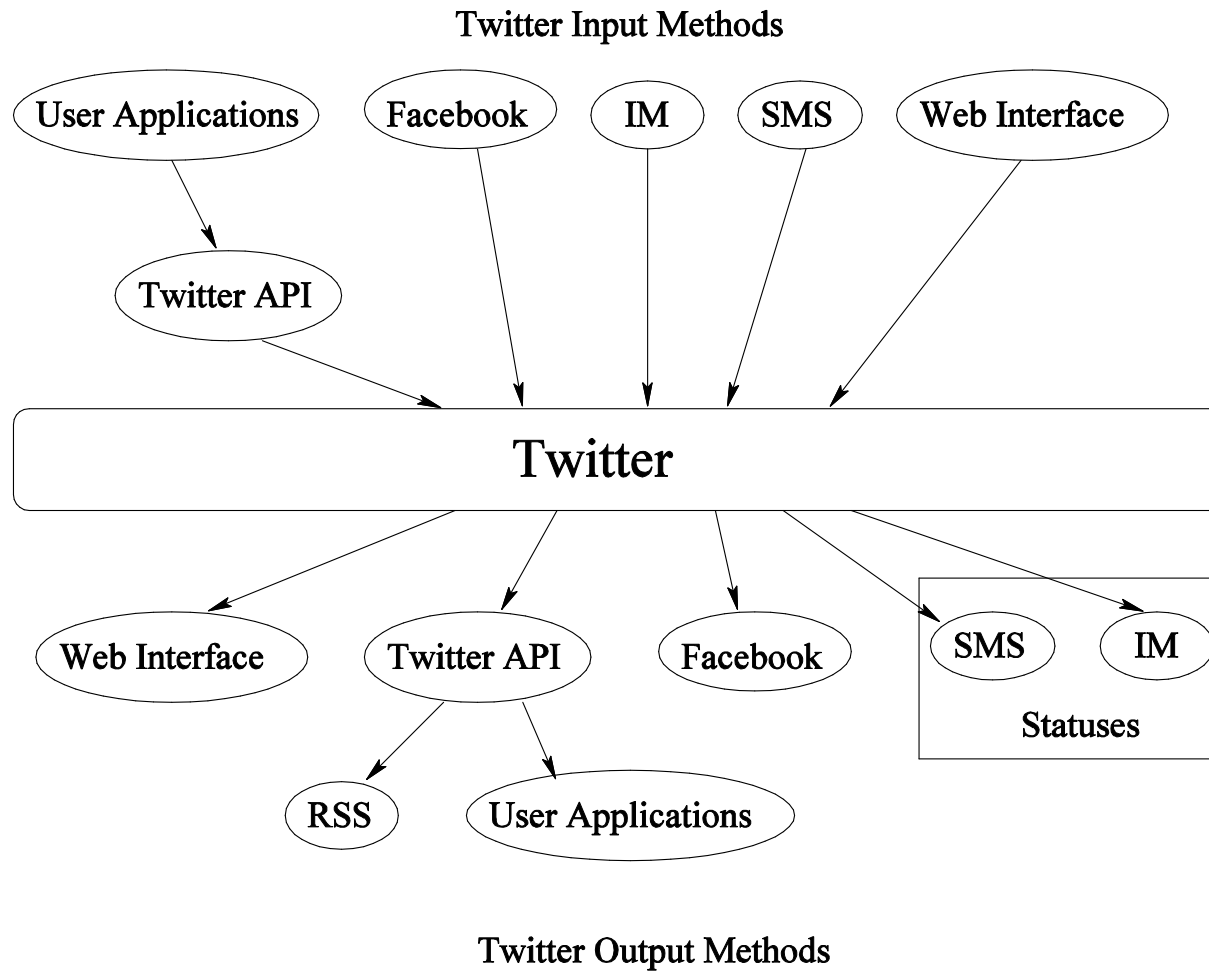- Characterization

- 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

Twitter Output Methods

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
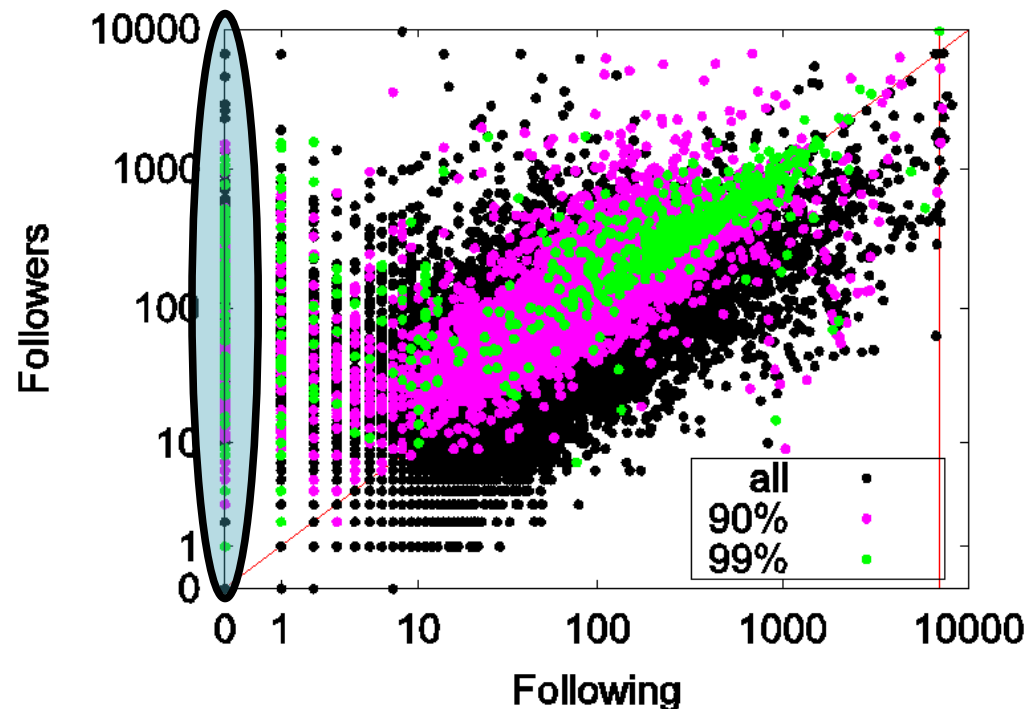 - Geographical spread of Twitter

Characterizing user relationships

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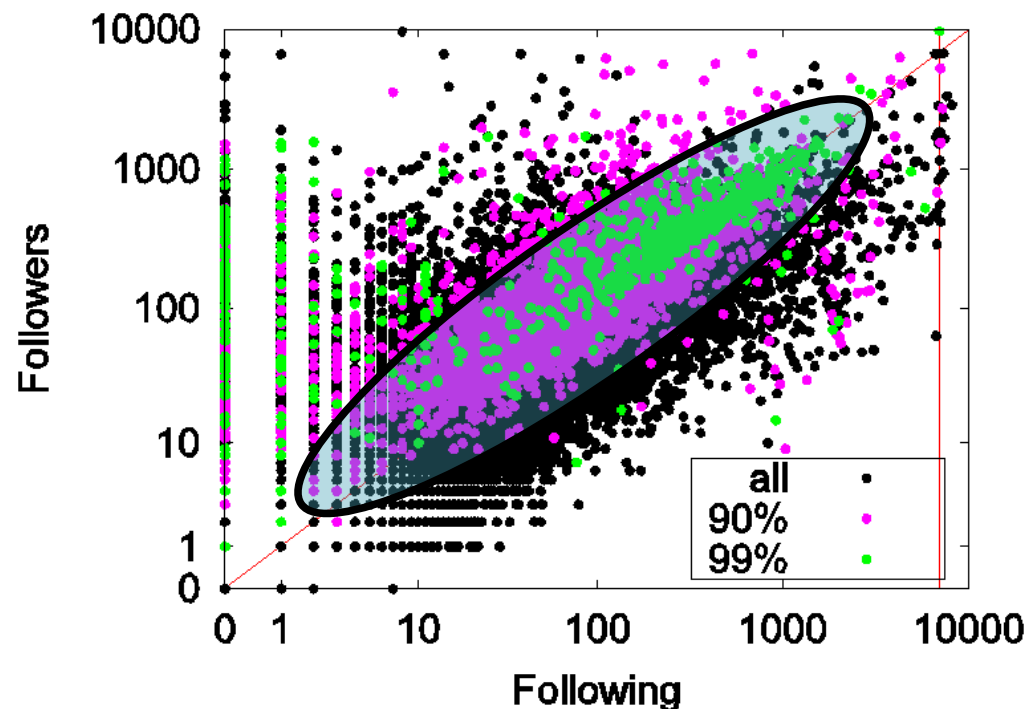
User relationships
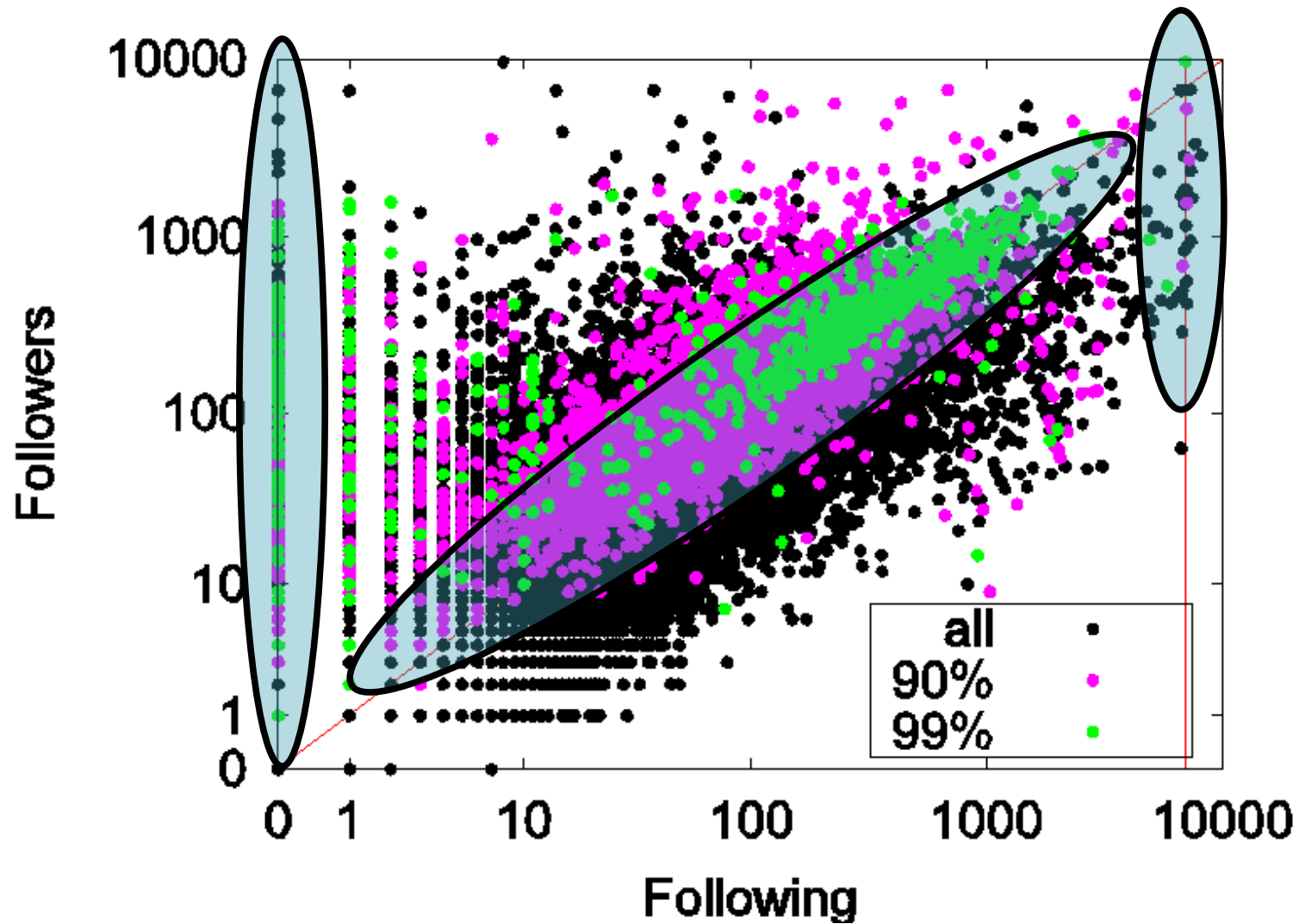
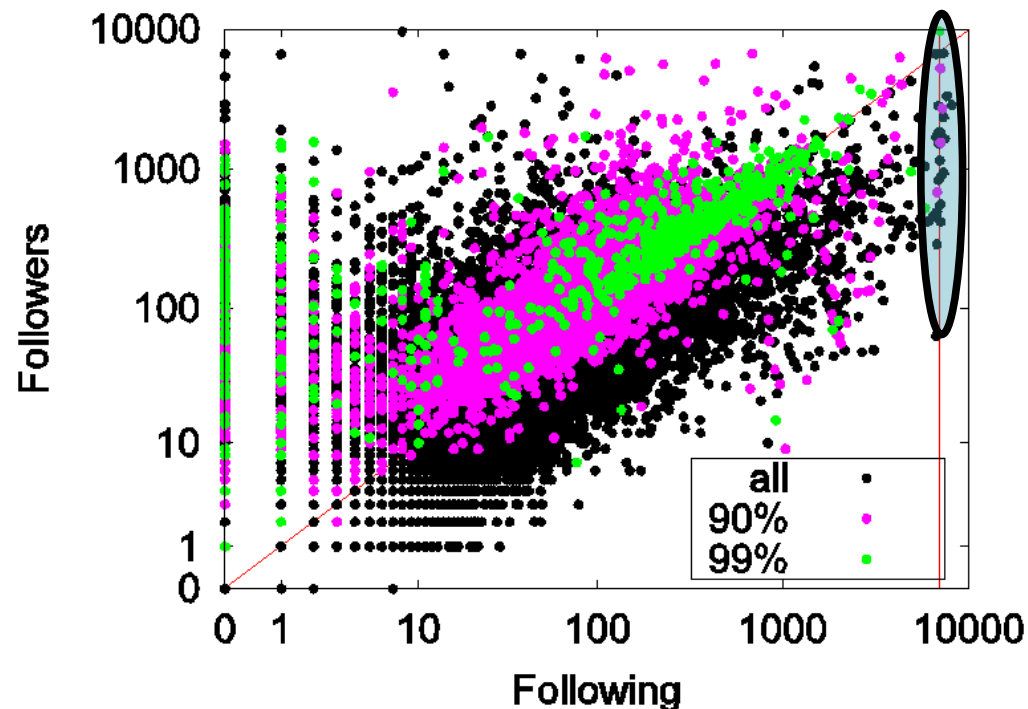# User relationships - Broadcasters

- 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)
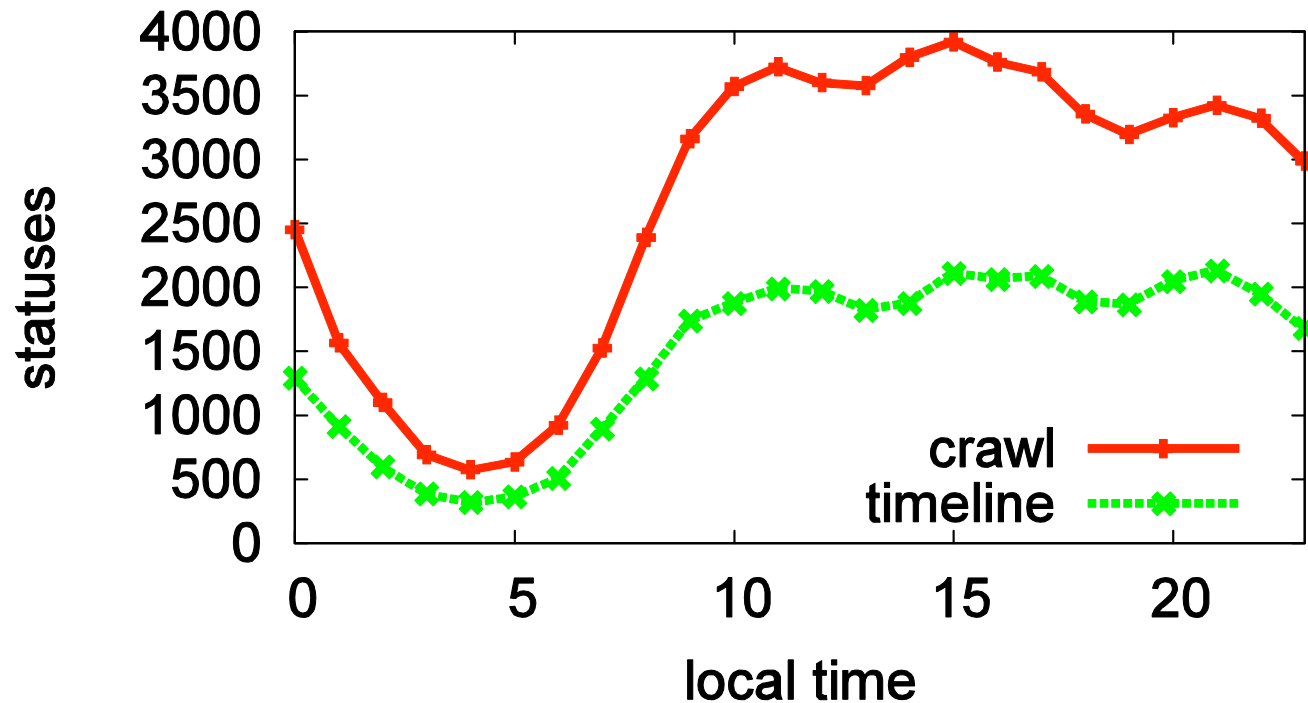
Characterizing user tweets

☐ Where do tweets come from?

☐ When are people tweeting?

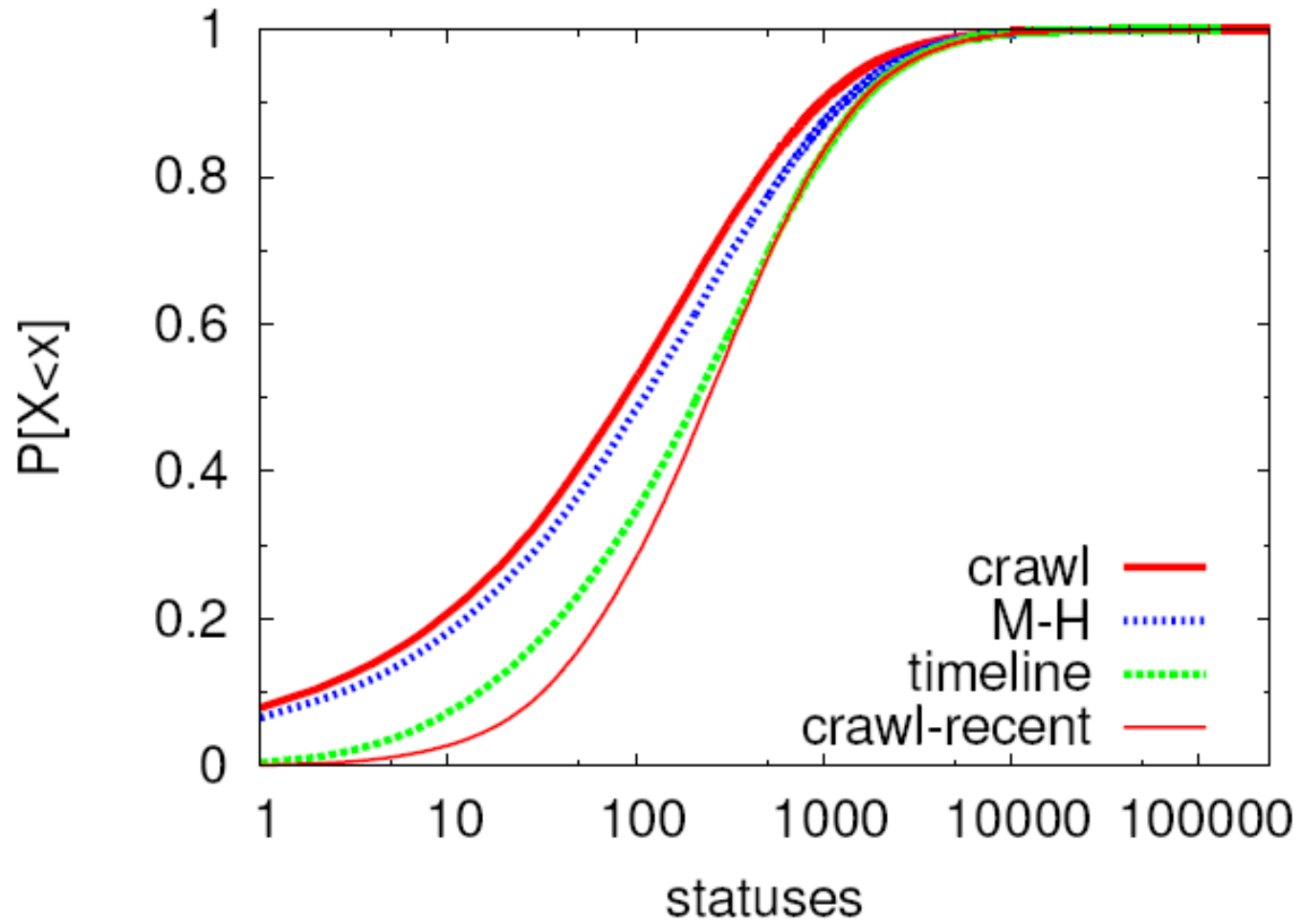☐ How many tweets do users have?

Where do tweets come from?

Crawl		source	Timeline	
%	tweets		%	tweets
61.7	40,163	Web	57.0	20,510
7.5	4,901	txt (mobile)	7.4	2,667
7.2	4,674	IM	7.5	2,714
1.2	792	Facebook	0.7	261
22.4	14,566	custom applications	27.3	9,821

When are people tweeting?

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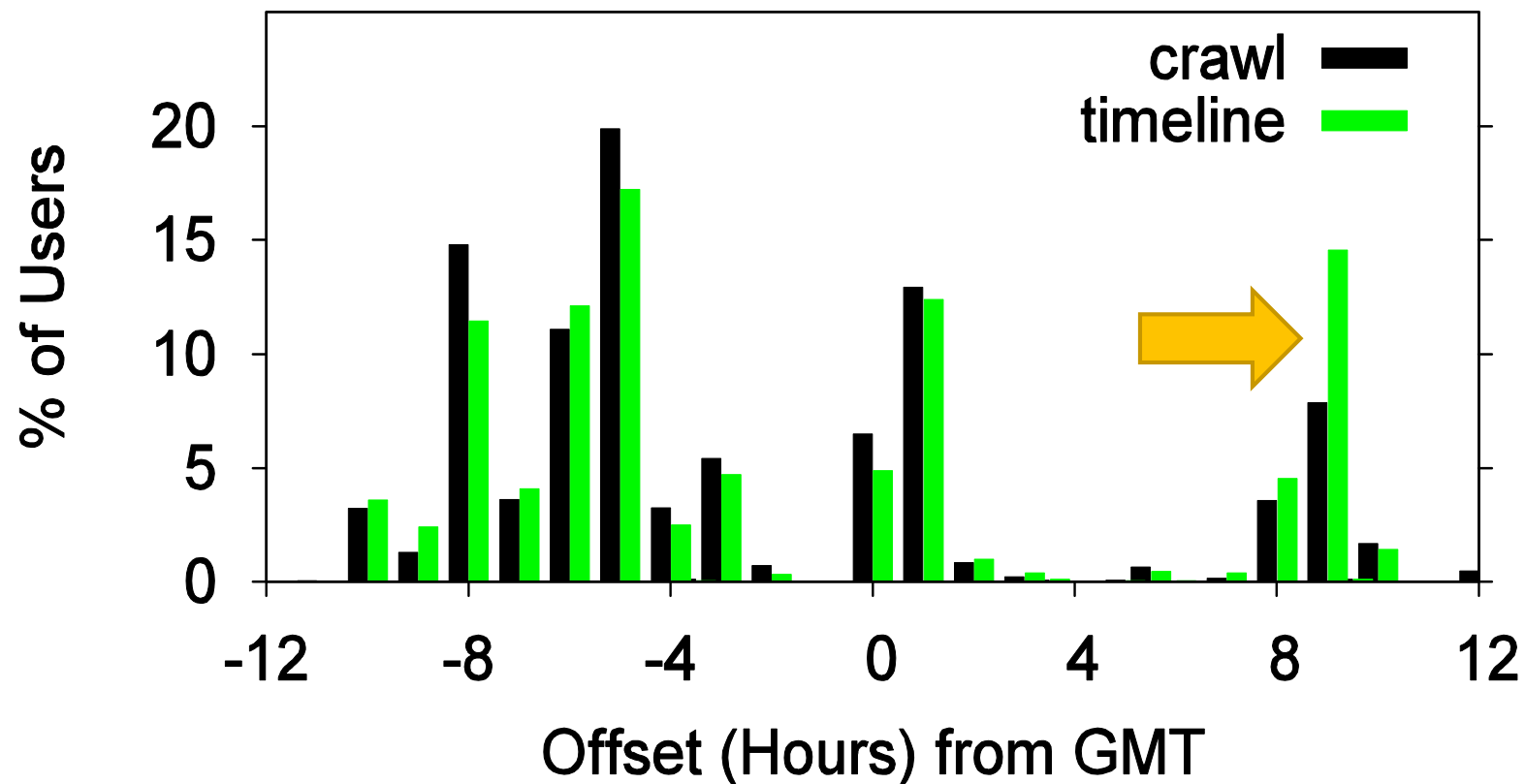
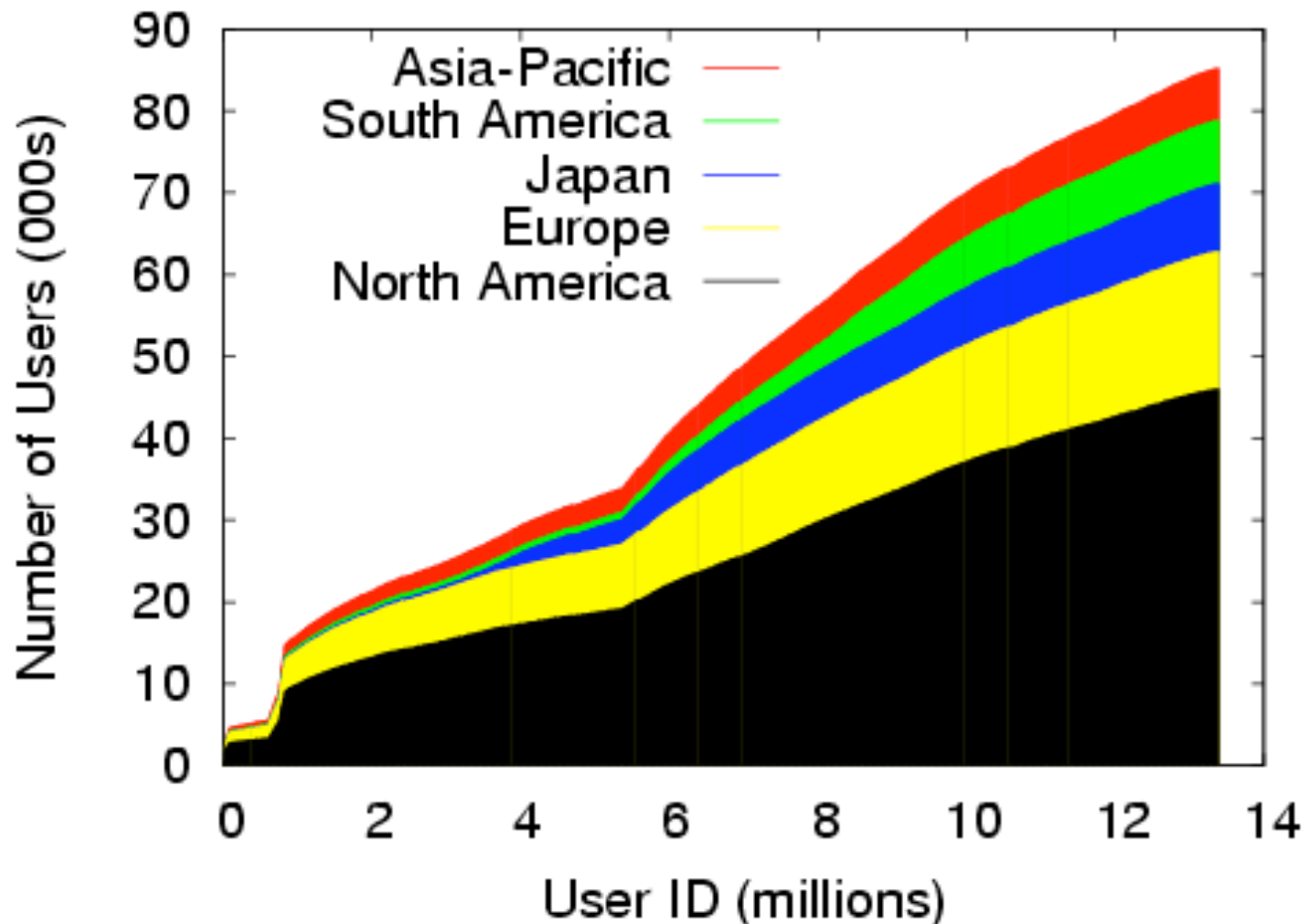# Number of tweets per user

Other properties of Twitter users

- UTC offsets
- Geographical spread of users

Comparison of UTC offsets of users between datasets

Geographical presence of Twitter

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

QUESTIONS?