Through the Lens of a Large Instant-Messaging Network: Planetary-Scale Views on Behavior

Eric Horvitz
Joint work with Jure Leskovec

Princeton University
April 2009
New Lenses on Behavior & Relation

- Anonymized data from wide-scale communication systems
- Structural properties of human communication graph
- Insights about people and groups, influences of demographics
Instant Messaging as Network

- **Buddy**
- **Conversation**
One month of data
- 245 million users logged in
- 180 million users engaged in conversations

Communication graph (two-way)
- > 30 billion conversations
- > 255 billion messages exchanged
- 1.3 billion edges

4.5 terabytes
Data Attributes

For every conversation: list of participants:
- User ID
- Time joined, time left
- Num. of messages sent, received

Demographic data (self-reported):
- Age
- Gender
- Location (Country, ZIP)
- Language

All data anonymized. No message text.
Research questions:

- What are key structural properties of the communication graph?
- How do geographic relationships affect communication?
- How are communication patterns influenced by demographics (age, sex, language, country)?
For each conversation between geo points (A,B) we increase the intensity on the line between A and B.
Communication Density

Each point represents number of users at location
Per Capita Analysis

Users per capita
Who Talks to Whom: Number of conversations
Is it a Small-World (after all)?

- Small-world experiment [Travers & Milgram '67]
  [Omaha, Wichita] → [Boston]
  - 296 letters (64 make it), avg num hops 6.2
Small World Studied on Larger Scale

Avg. path length 6.6
90% of others can be reached in < 8 hops
Communication: Geo distance

- Longer links used more
Random pairs of people are 6631 km apart on the average (7317 km median)
Geo-length of Shortest Paths

Shortest paths are about 15,000 kilometers longer than what they could be
Geo-length of Shortest Paths
Age: Number of conversations

- Young people communicate with same age
Age: Conversation duration

- Older people have longer conversations
Age: Messages per conversation

- Older people exchange more messages per session.
Age: Messages per unit time

Young people converse more quickly
Communication: Gender

- Influence of gender
  - Number of conversations: ~chance
  - Cross-gender: Longer, more messages

![Diagram showing gender influence on conversations, duration, and messages/conversation]
Opportunities to study behaviors in the large

- Patterns of communication
- Influence of demographics

Investigation of structure of network
- Well-connected small world

Multiple directions of ongoing research
