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

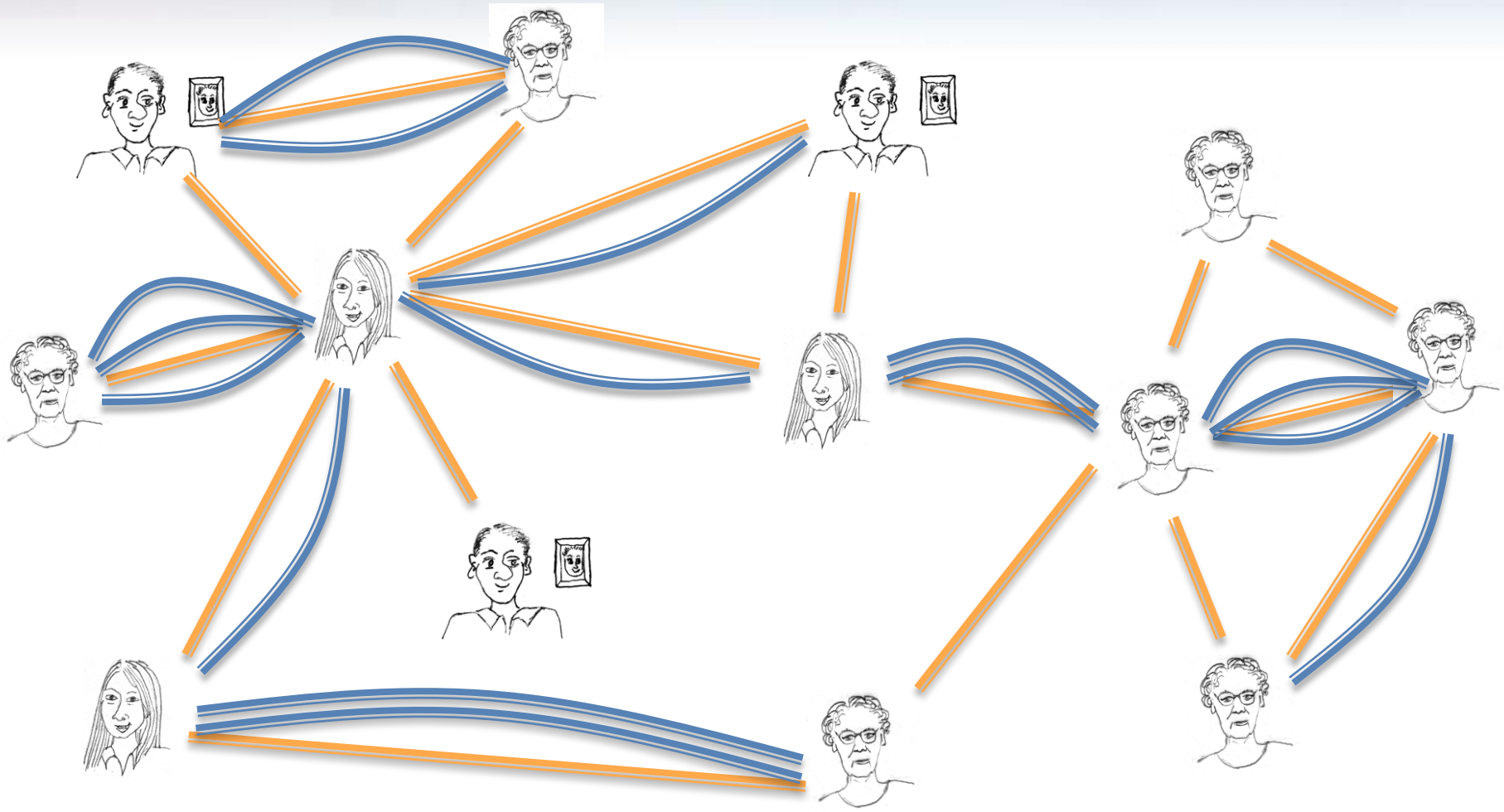
New Lenses on Behavior & Relation

- 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



Scale

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

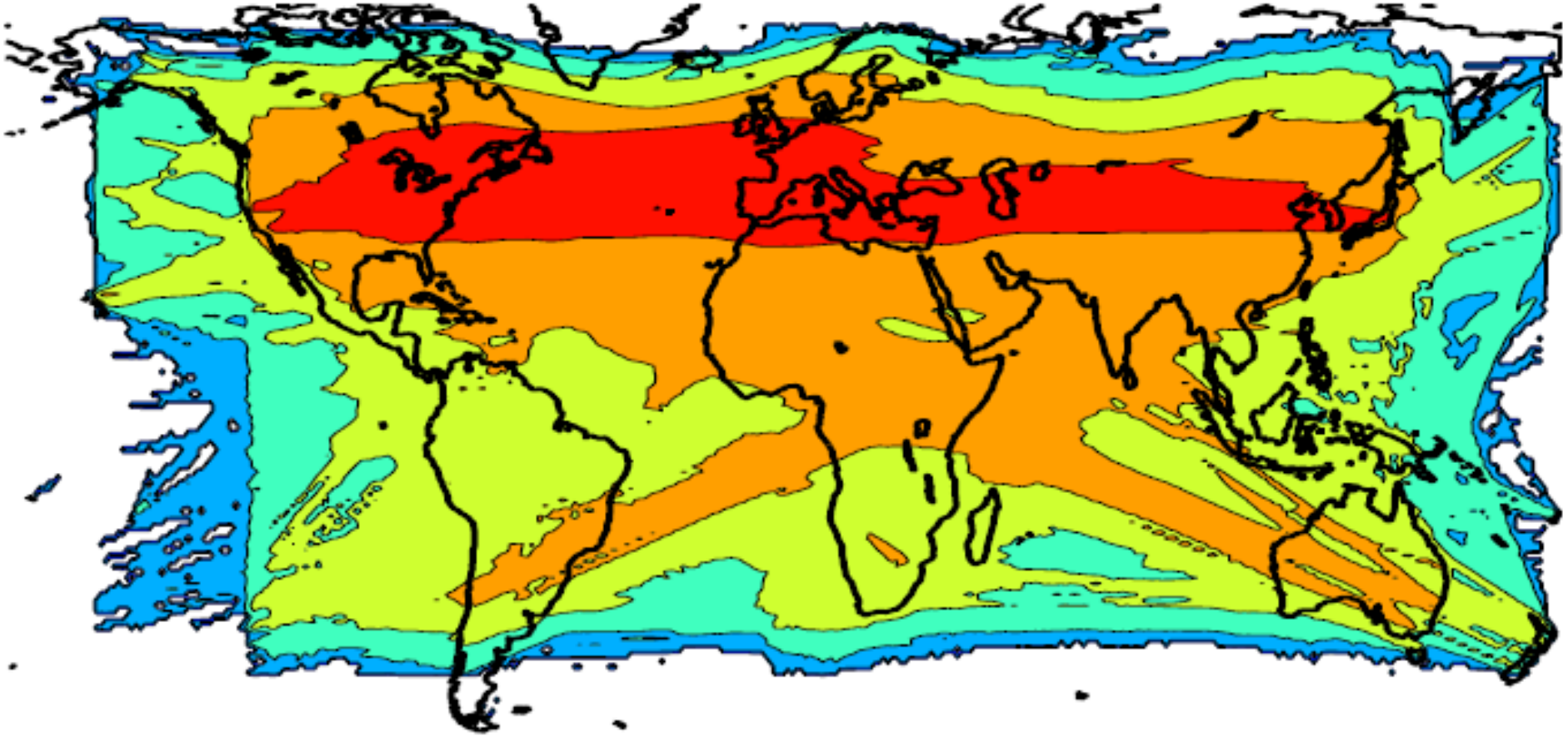
All data anonymized. No message text.

Behavioral Studies at Planetary Scale

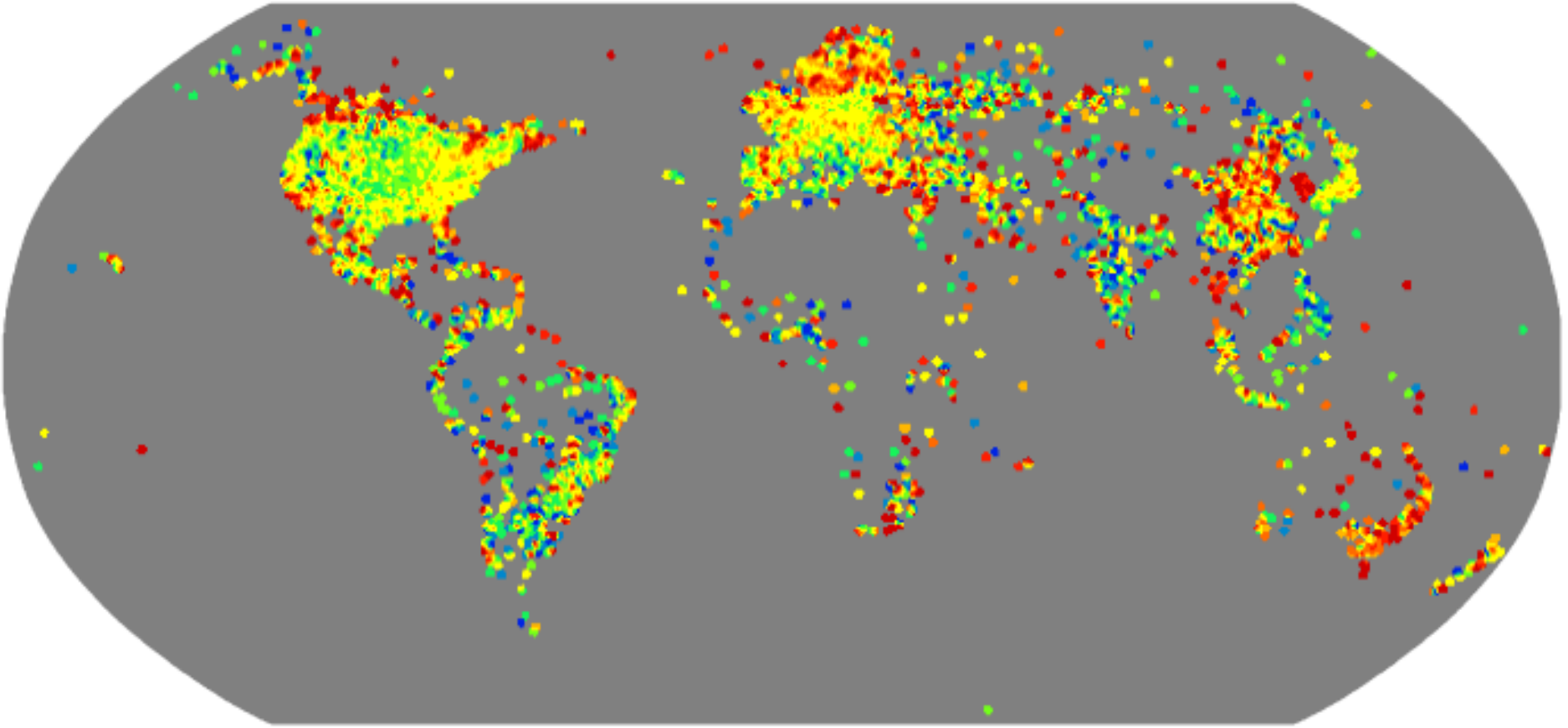
Research questions:

- 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)?

Visualizing World Communication Axis

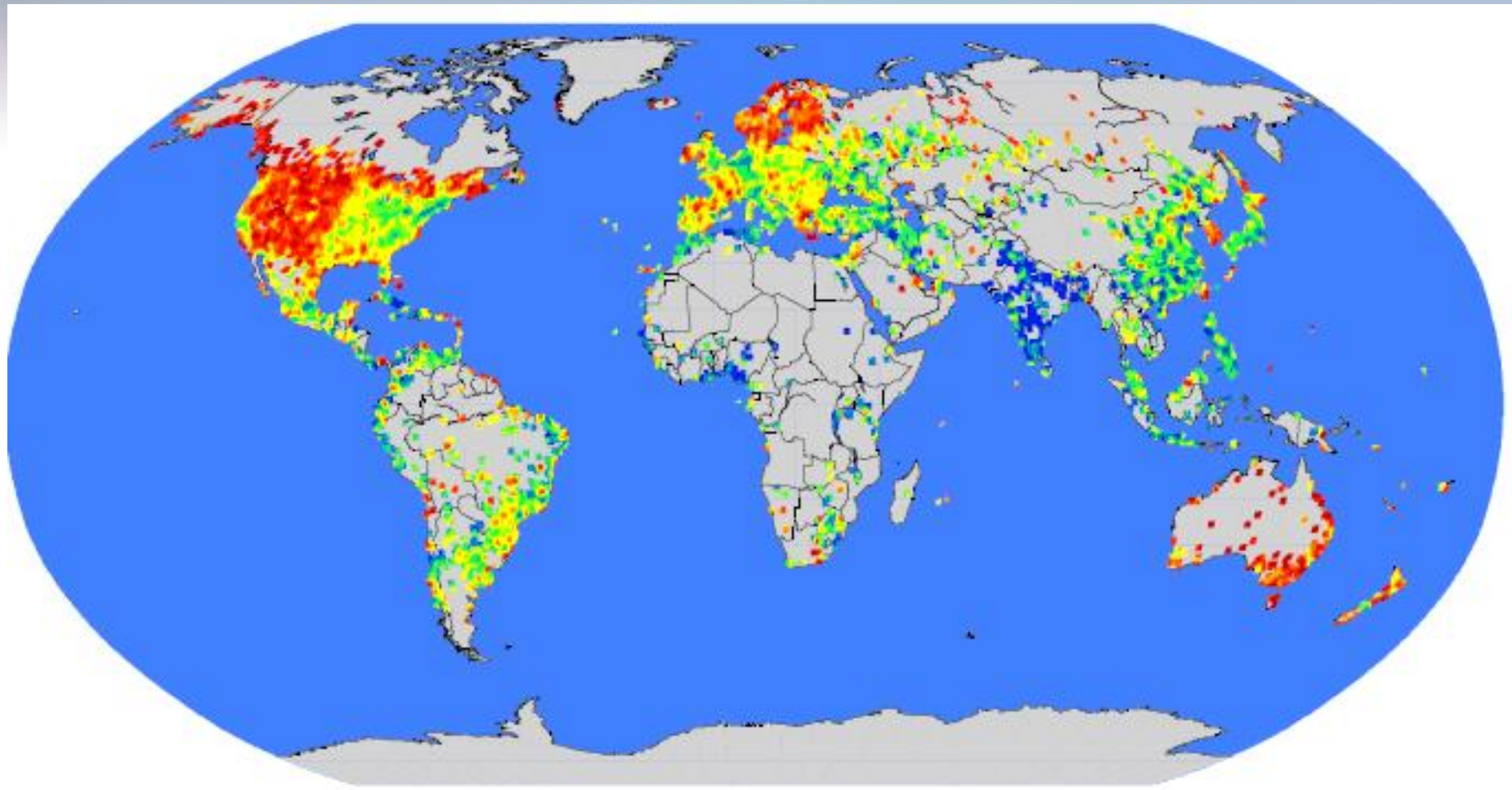


Communication Density



Each point represents number of users at location

Per Capita Analysis

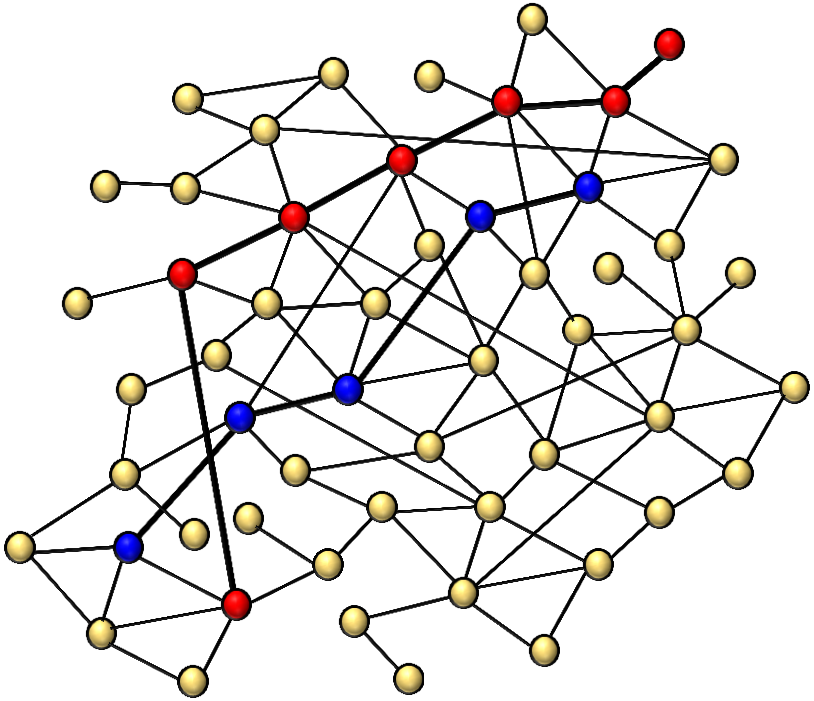


Users per capita

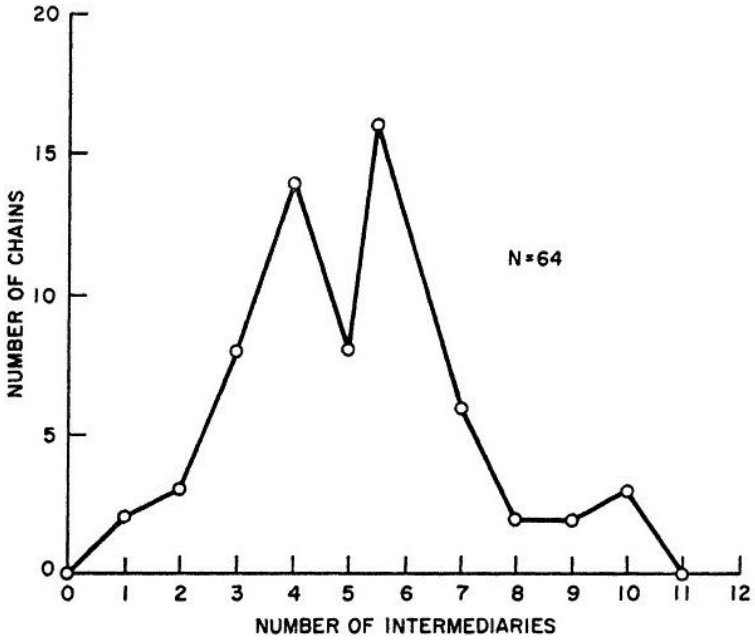
Who Talks to Whom: Number of conversations

Is it a Small-World (after all)?

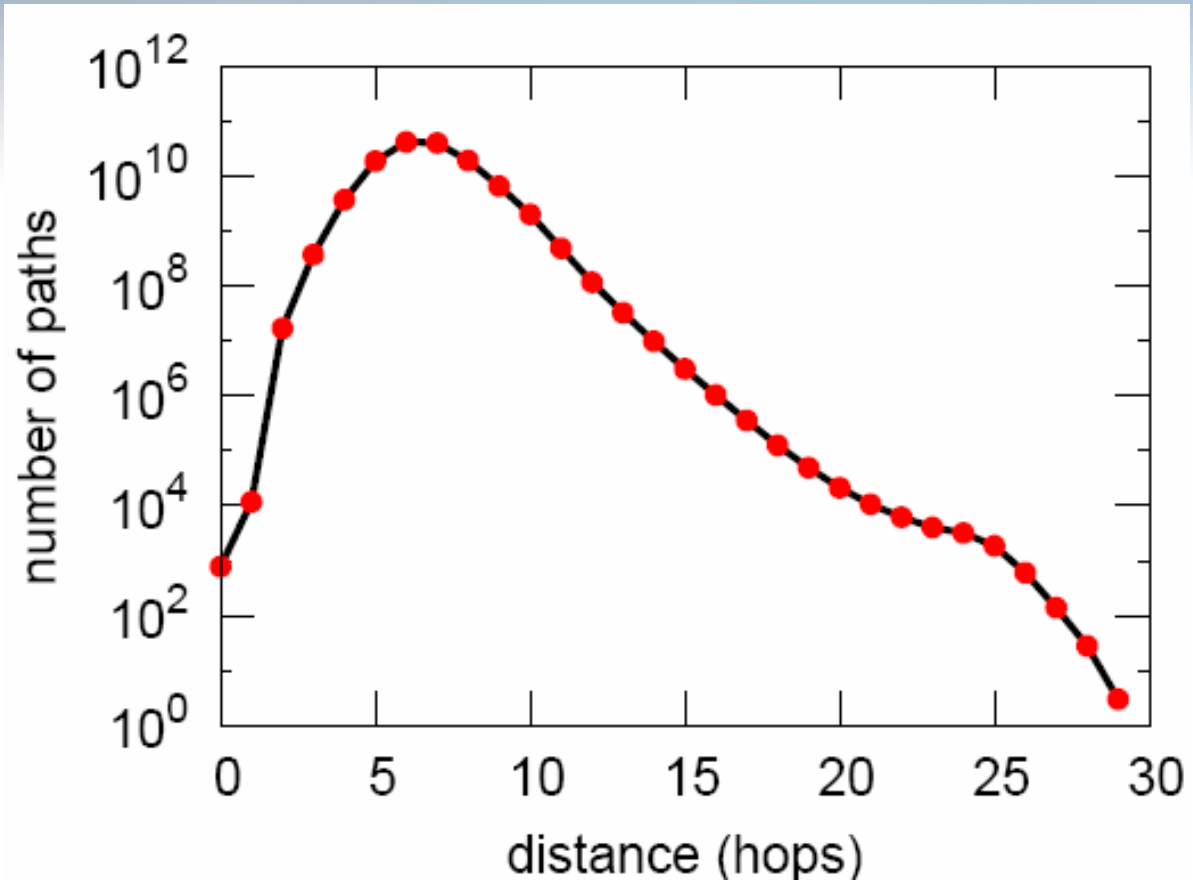
- Small-world experiment [Travers&Milgram ’67]
  [Omaha, Wichita] → [Boston]
  - 296 letters (64 make it), avg num hops 6.2



Milgram’s small world experiment



Small World Studied on Larger Scale

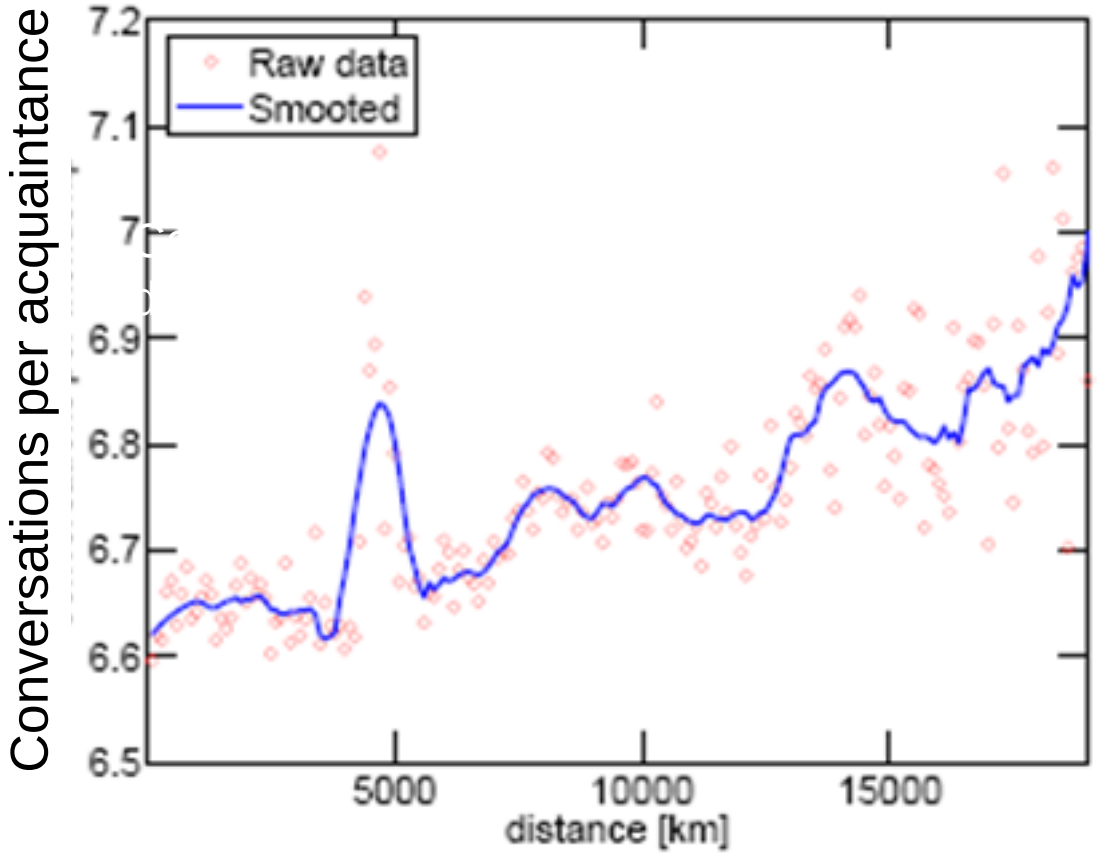


Avg. path length 6.6

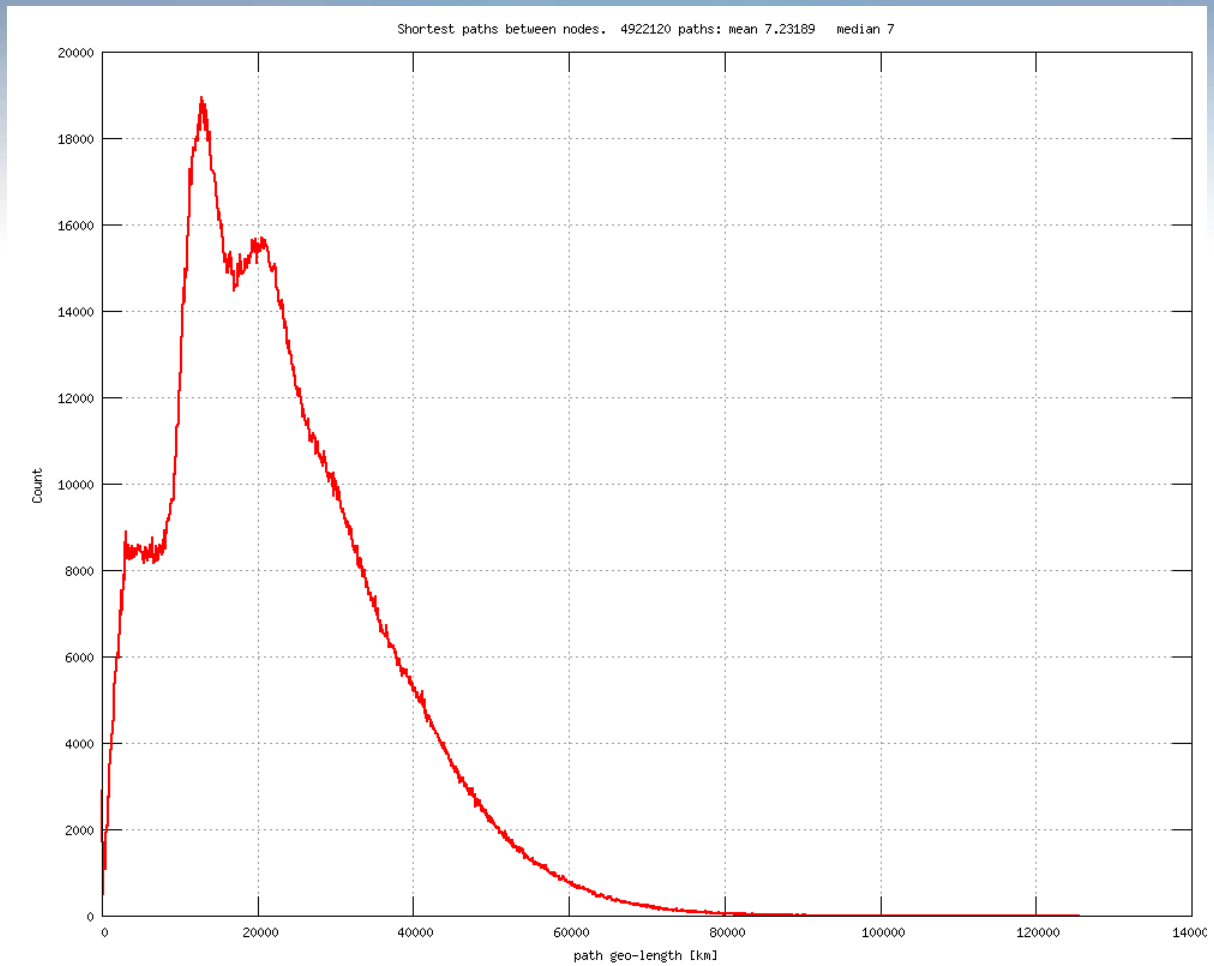
90% of others can be reached in < 8 hops

Communication: Geo distance

- Longer links used more

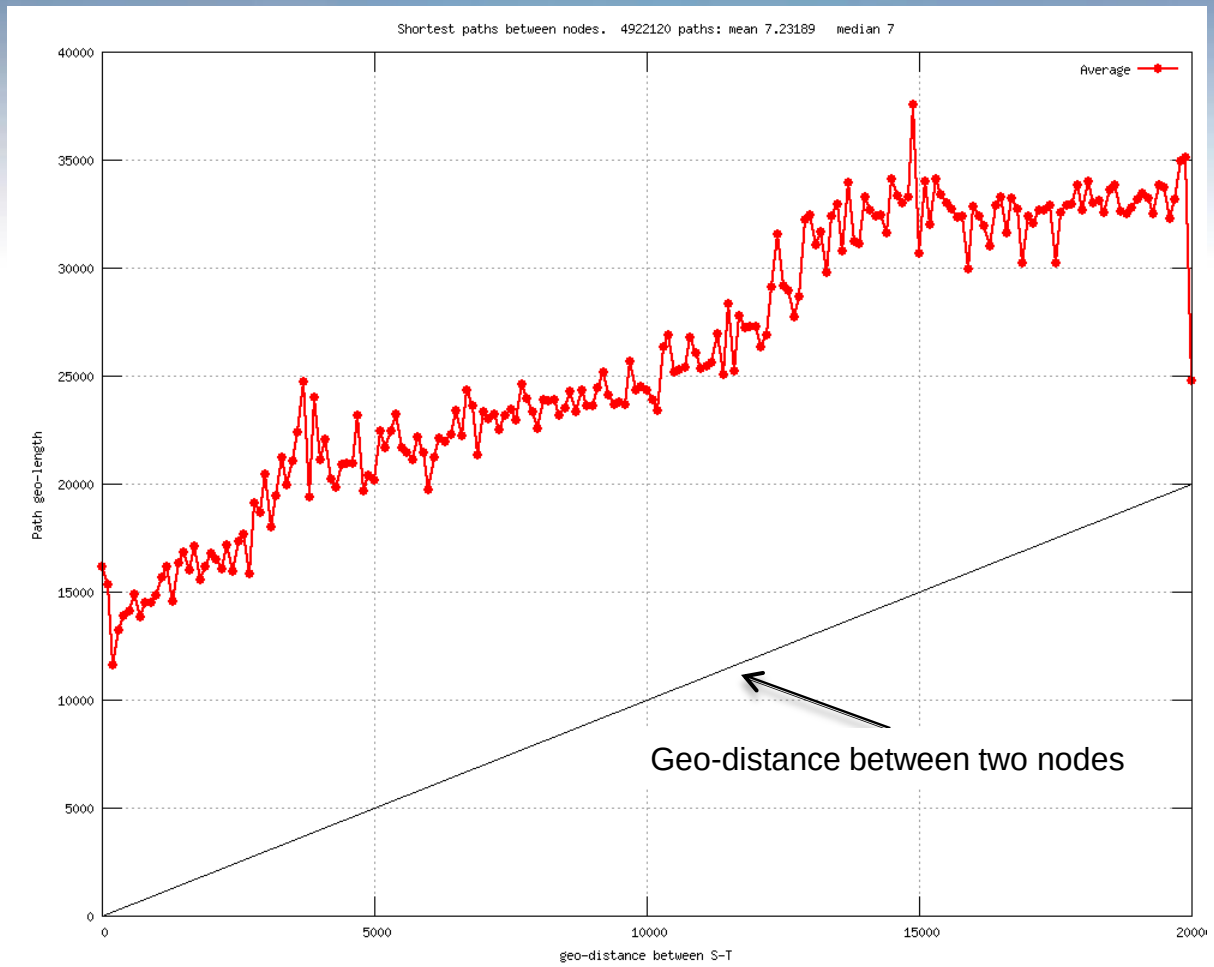


Geographic Separation



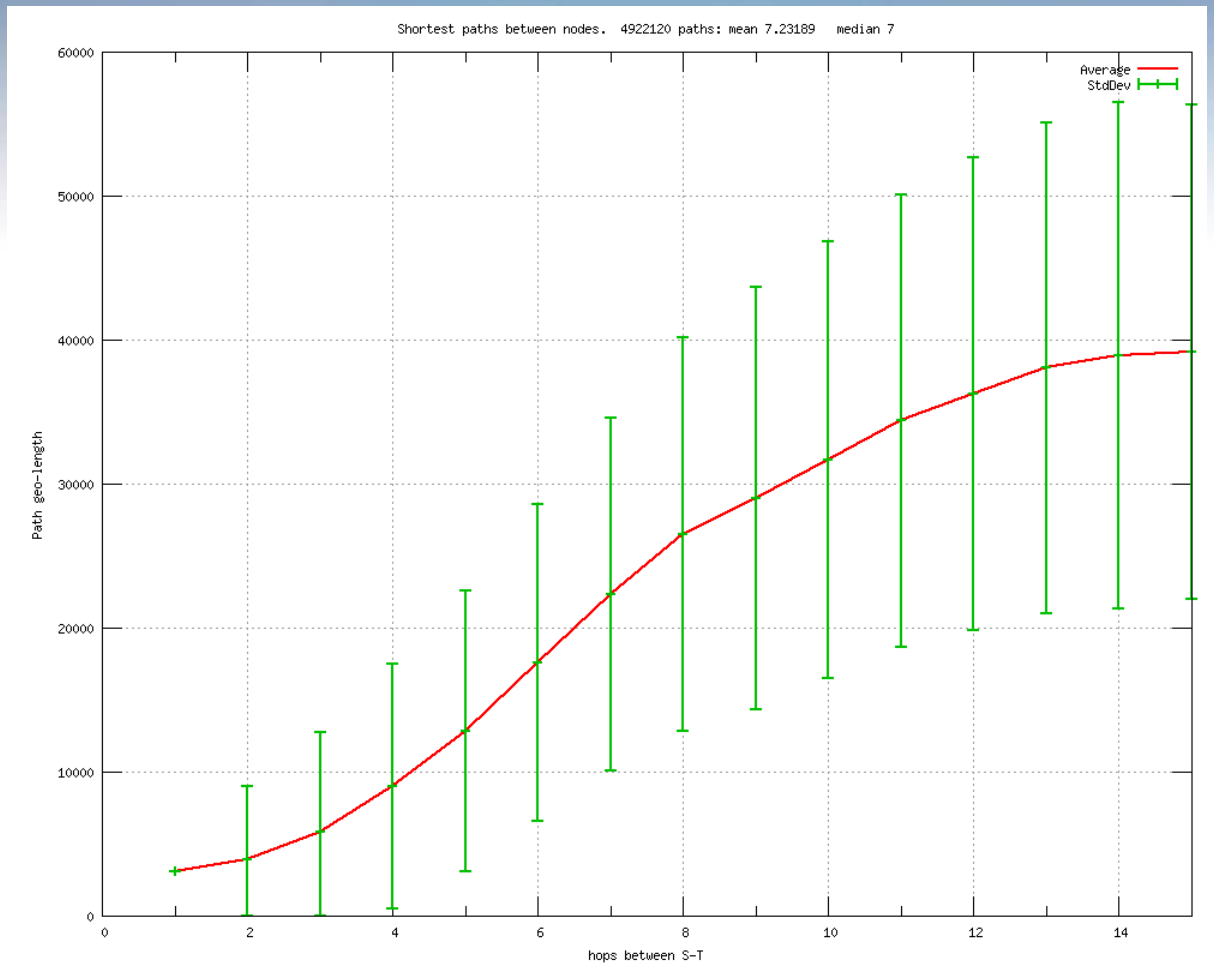
Random pairs of people are 6631 km apart on the average (7317 km median)

Geo-length of Shortest Paths



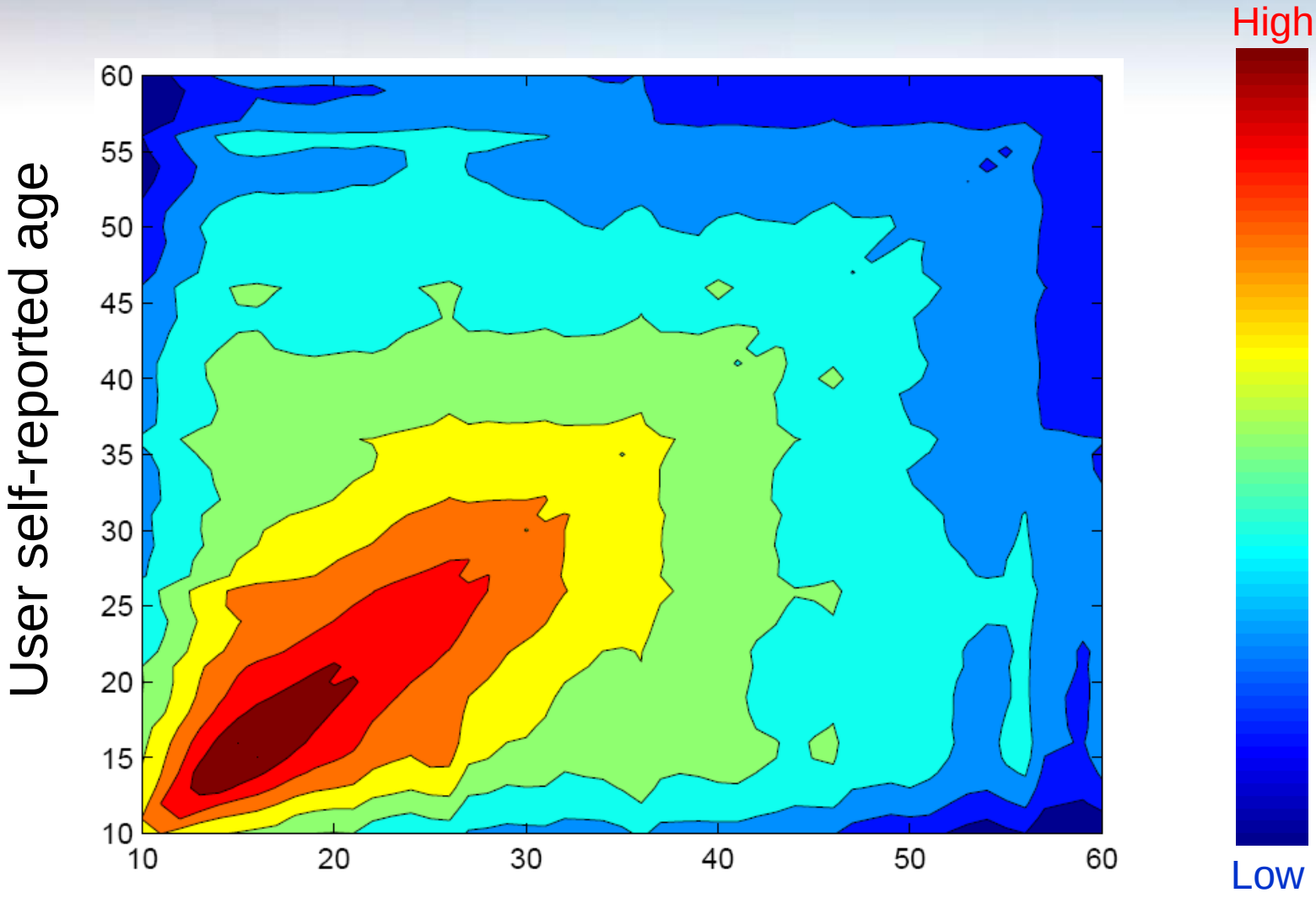
- Shortest paths are about 15,000 kilometers longer than what they could be

Geo-length of Shortest Paths



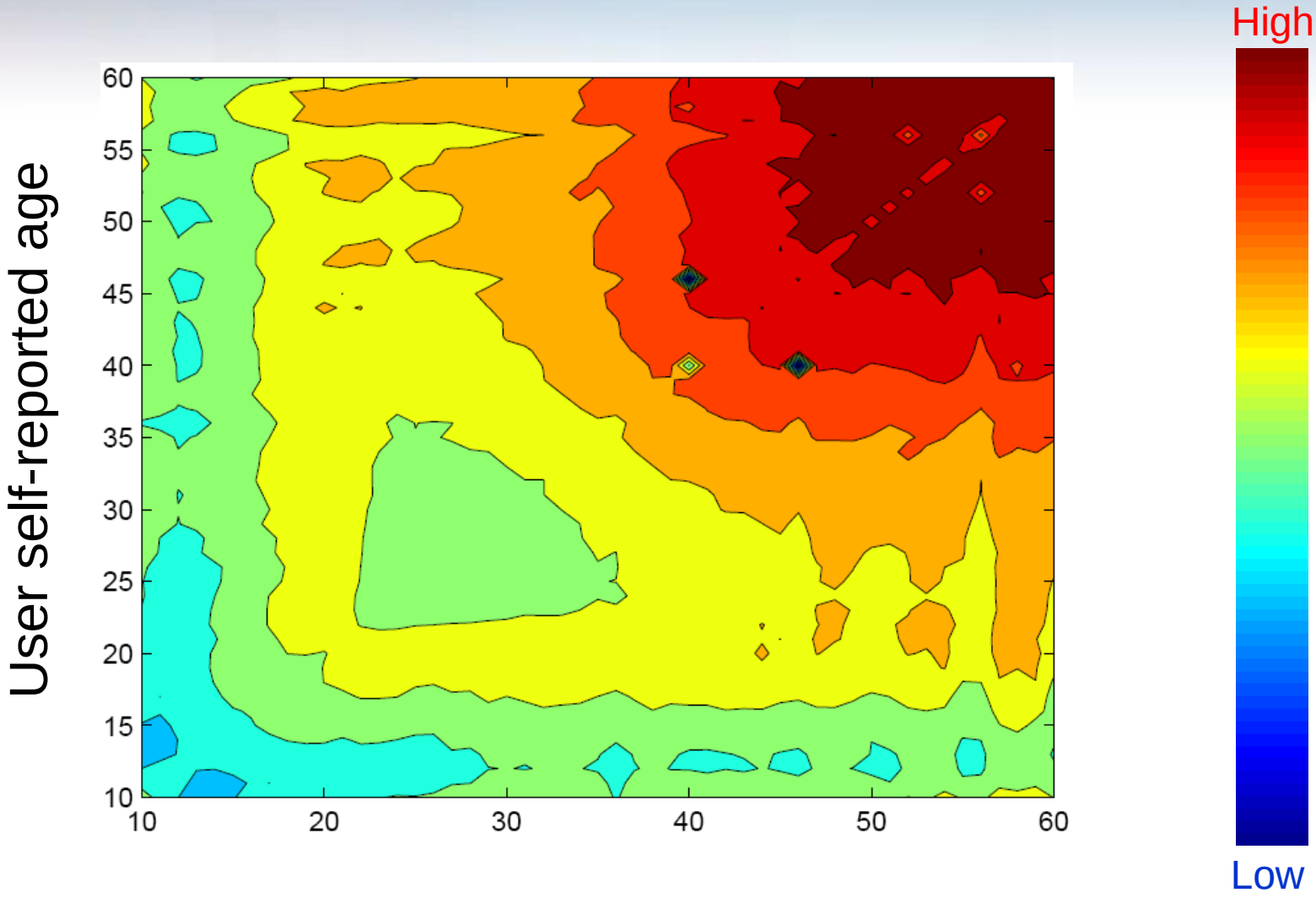
Age: Number of conversations

- Young people communicate with same age



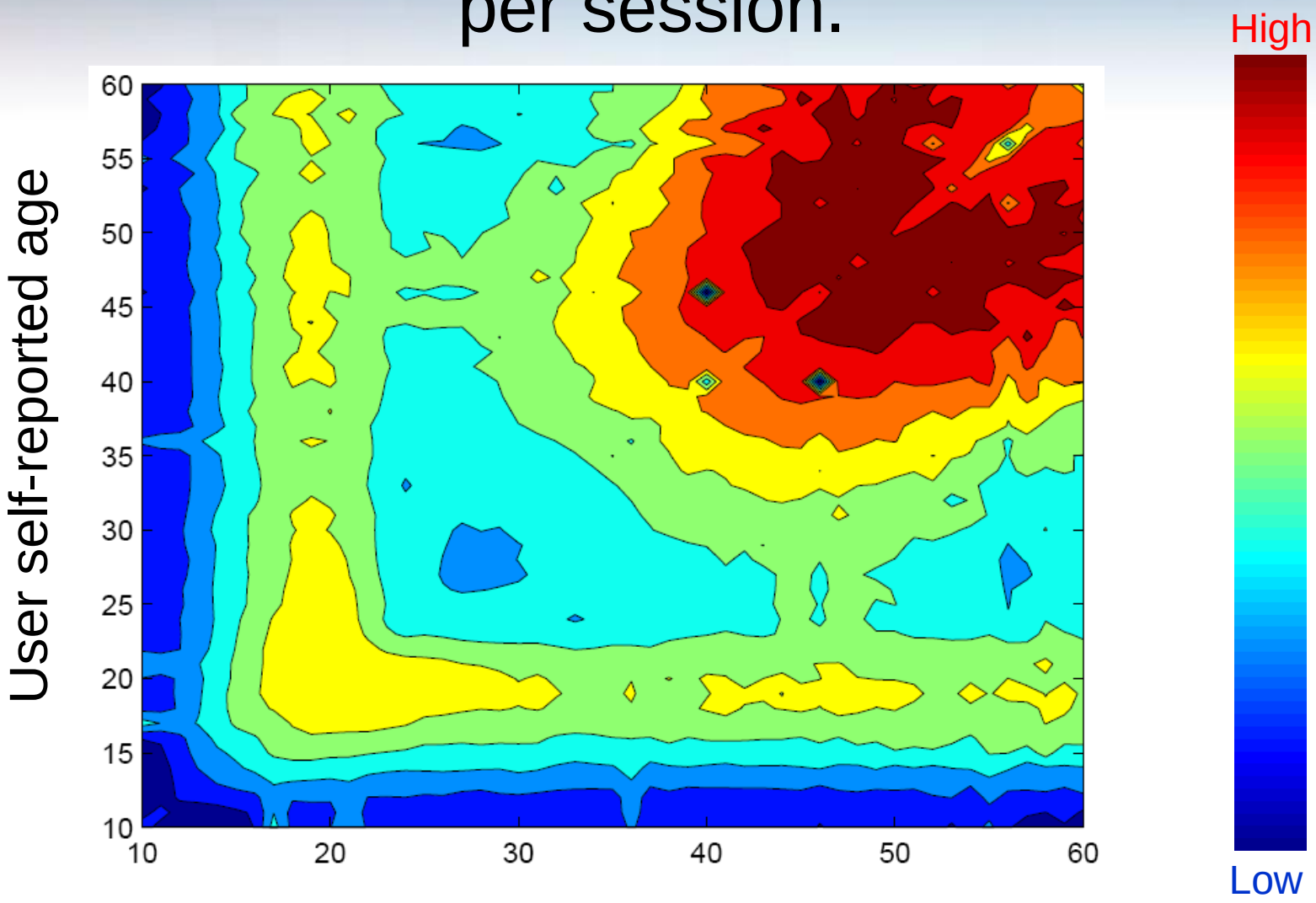
Age: Conversation duration

- Older people have longer conversations



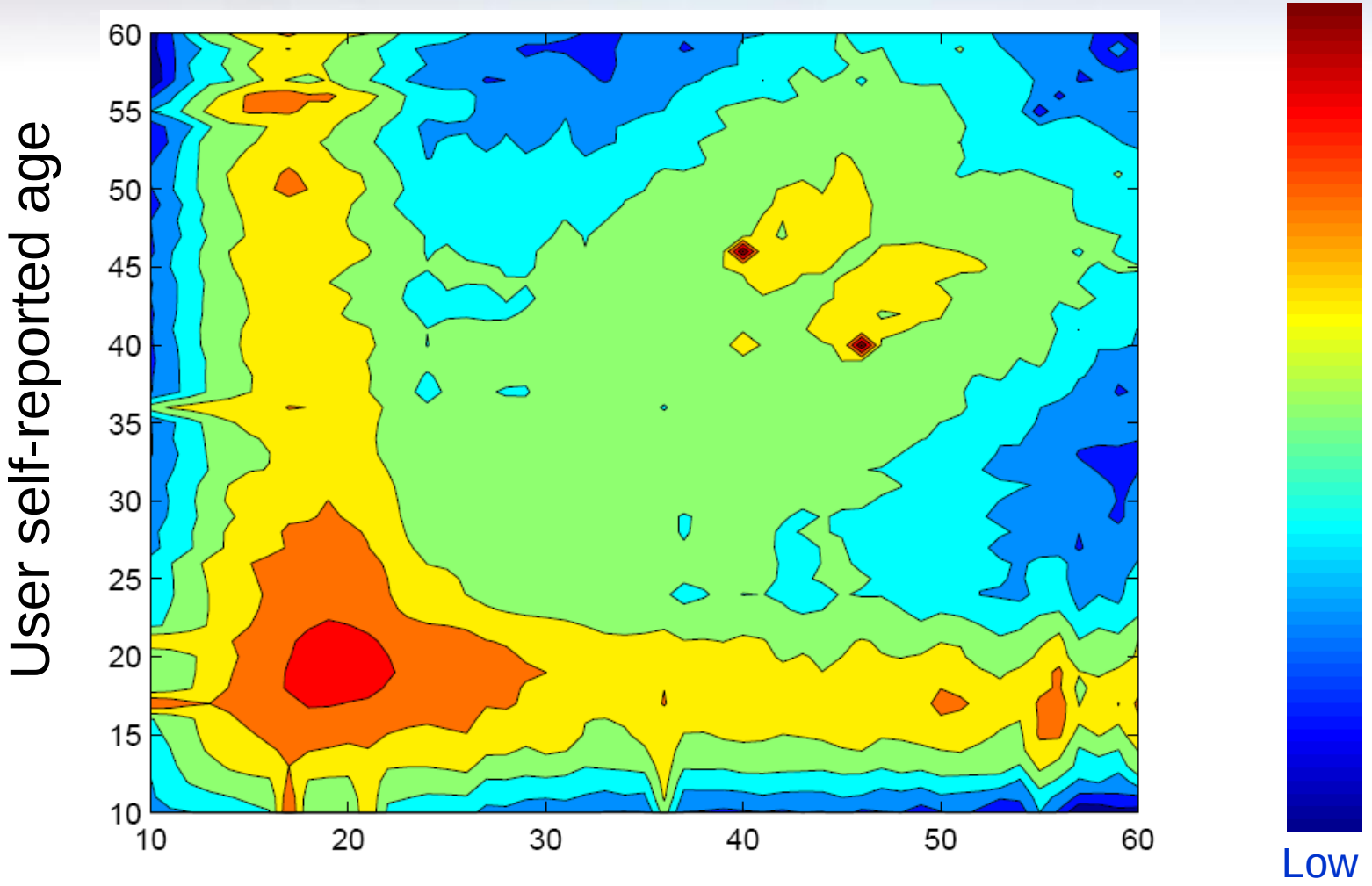
Age: Messages per conversation

- Older people exchange more messages per session.



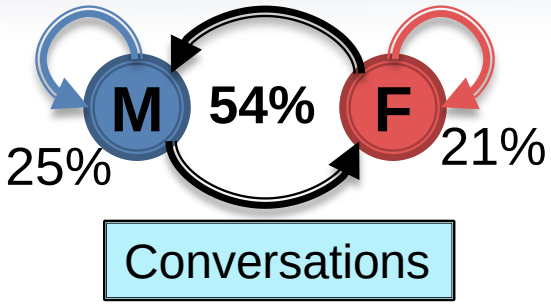
Age: Messages per unit time

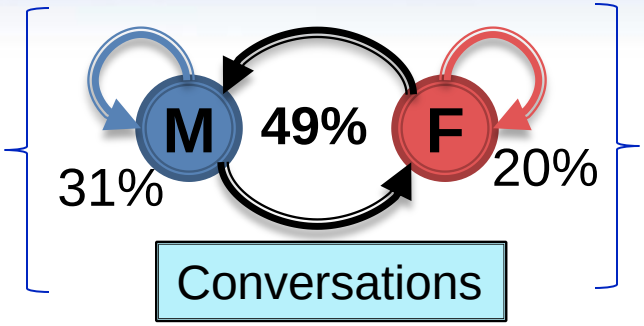
- Young people converse more quickly



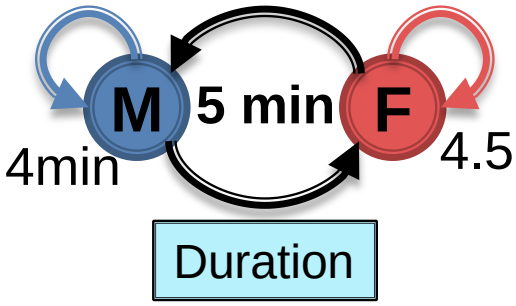
Communication: Gender

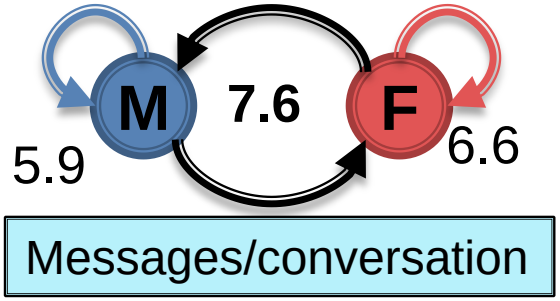
- Influence of gender
  - Number of conversations: ~chance





  - Cross-gender: Longer, more messages





Summary

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

More information

- J. Leskovec and E. Horvitz. Worldwide Buzz: Planetary-Scale Views on an Instant-Messaging Network, *Microsoft Research Technical Report MSR-TR-2006-186*, Microsoft Research, June 2007.

- J. Leskovec and E. Horvitz. Planetary-Scale Views on a Large Instant-Messaging Network, *Proceedings of WWW 2008*, Beijing, China, April 2008.