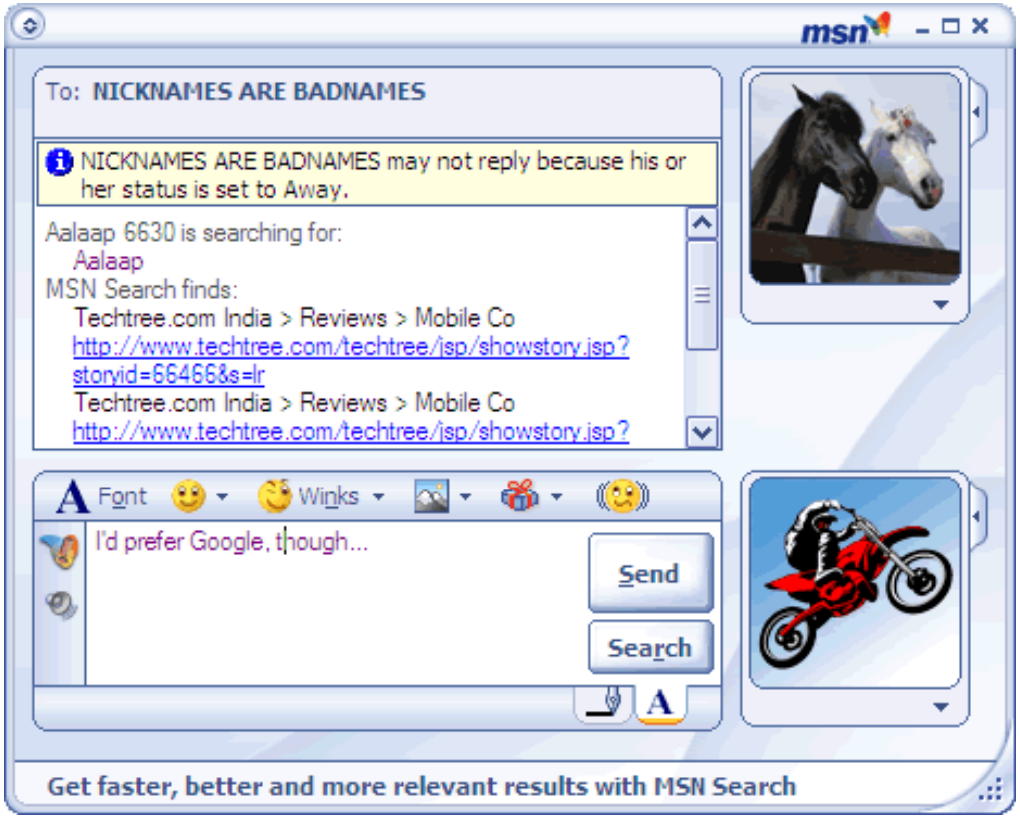
Small world phenomena
6-degrees of separation
Navigation in networks

CS 322: (Social and Information) Network Analysis

Jure Leskovec
Stanford University

Instant Messaging



- Contact (buddy) list
- Messaging window

IM – Phenomena at planetary scale

- Observe social and communication phenomena at a ***planetary*** scale
- **Largest social network analyzed to date**

Research questions:

- How does communication change with user demographics (age, sex, language, country)?
- How does geography affect communication?
- What is the structure of the communication **network**?

Data statistics: Total activity

- Data for **June 2006**
- Log size:
  150Gb/day (compressed)
- Total: 1 month of communication data:
  4.5Tb of compressed data
- **Activity over June 2006 (30 days)**
  - 245 million users logged in
  - 180 million users engaged in conversations
  - 17,5 million new accounts activated
  - More than 30 billion conversations
  - More than 255 billion exchanged messages

Data statistics: Typical day

Activity on a typical day (June 1 2006):

- 1 billion conversations
- 93 million users login
- 65 million different users talk (exchange messages)
- 1.5 million invitations for new accounts sent

Communication: Demographics

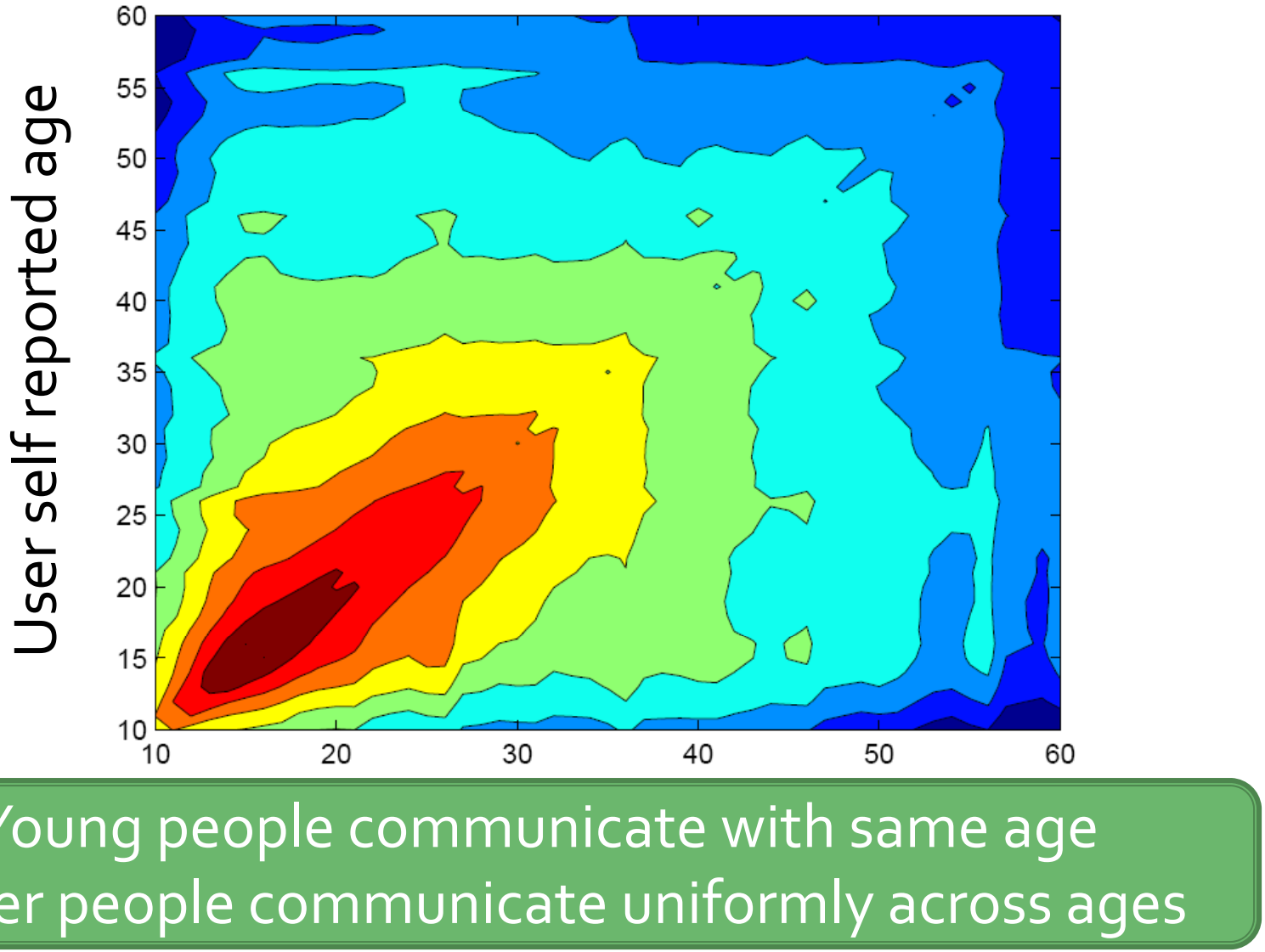
- How do people's attributes (age, gender) relate to communication?

Attribute	Random	Communicate
Age	0.030	0.162
Gender	0.434	0.426
ZIP	0.001	0.23
County	0.046	0.734
Language	0.030	0.798

Probability that users share an attribute

People tend to talk to similar people (except gender)

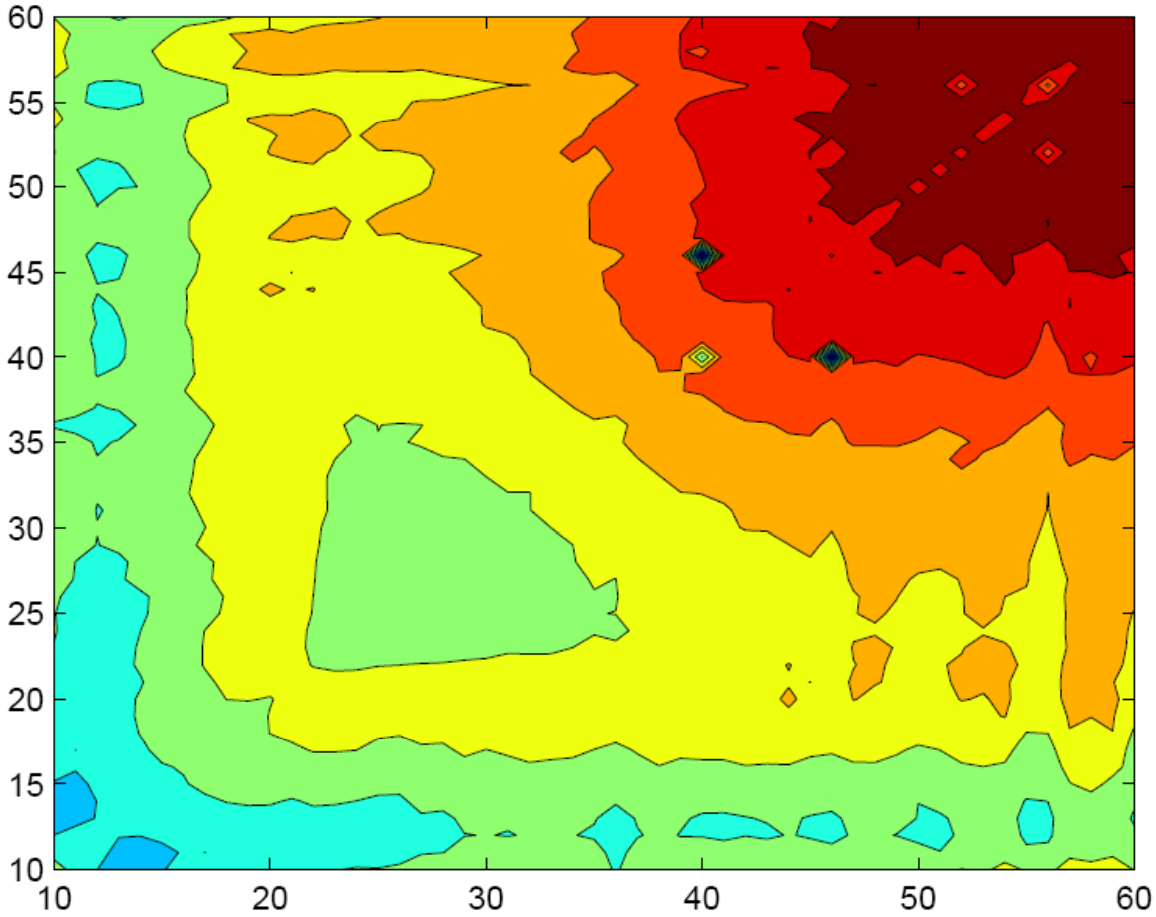
Age: Number of conversations



1) Young people communicate with same age
2) Older people communicate uniformly across ages

Age: Total conversation duration

User self reported age



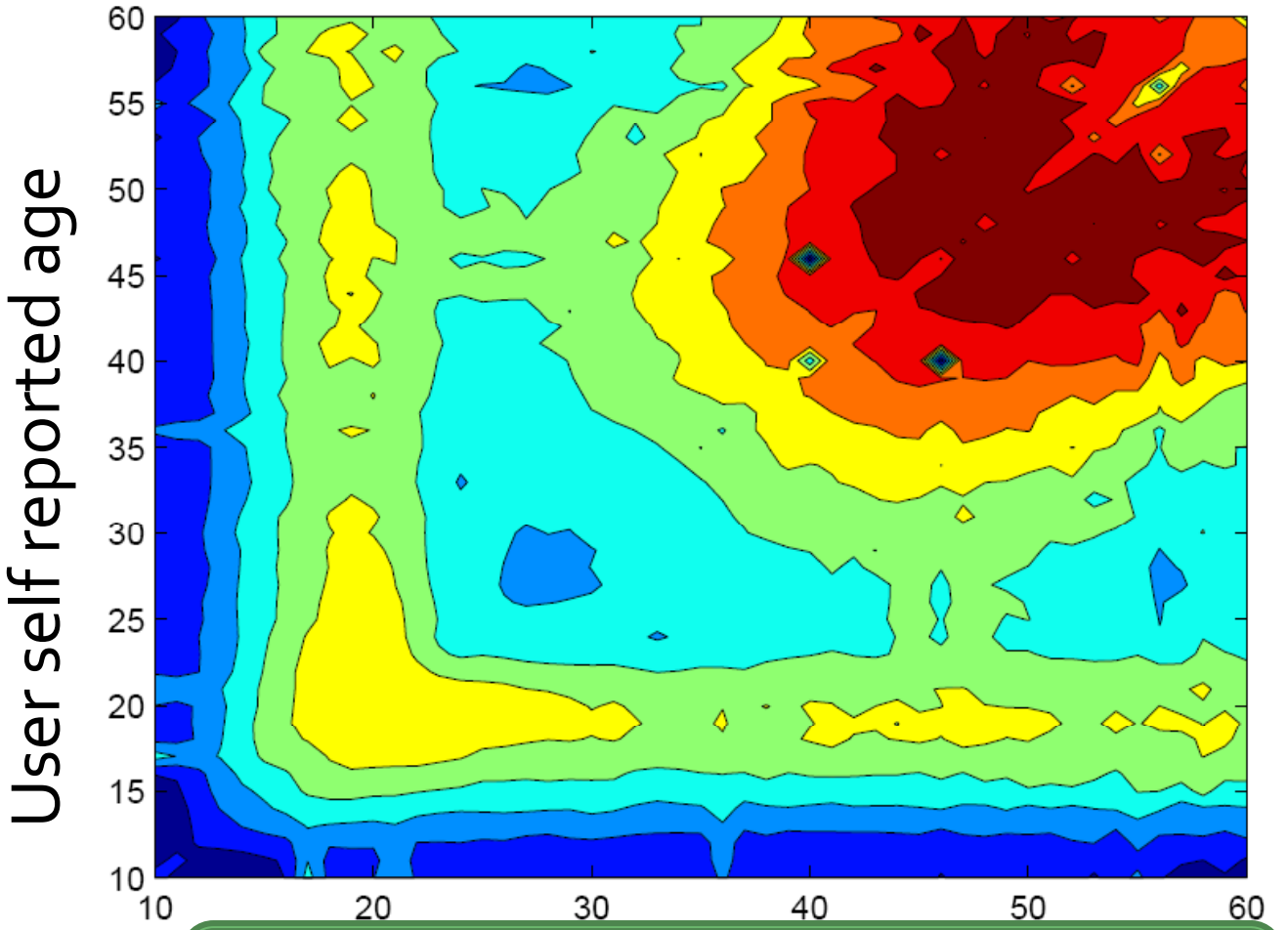
High

Low

1) Old people talk long

2) Working ages (25-40) talk short

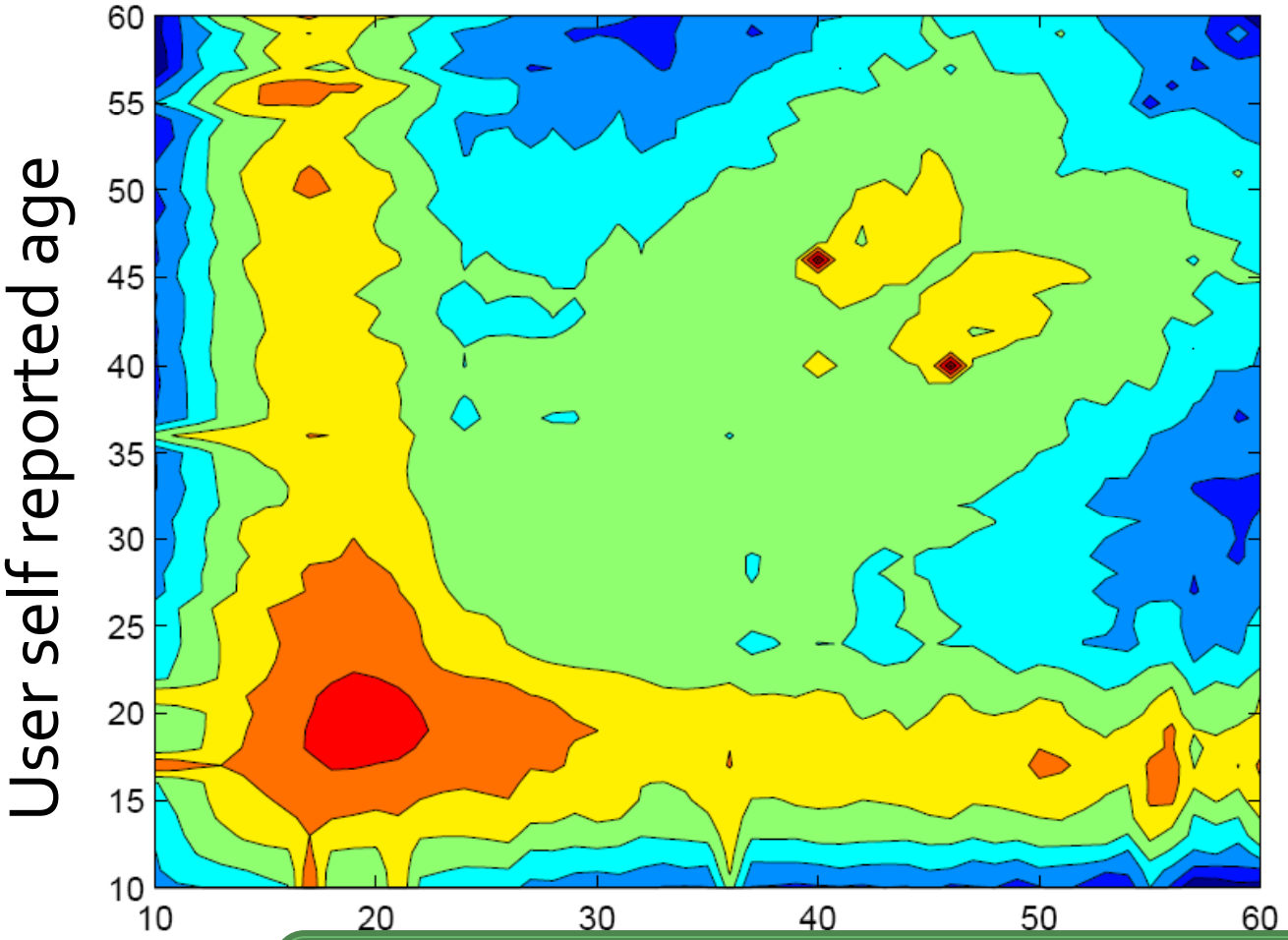
Age: Messages per conversation



1) Old people talk long

2) Working ages (25-40) talk quick

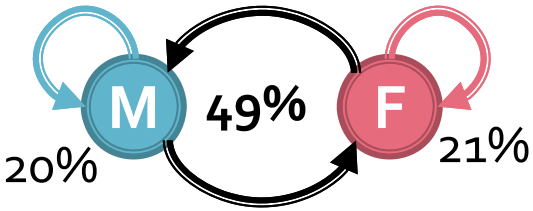
Age: Messages per unit time



1) Old people talk slow
2) Young talk fast

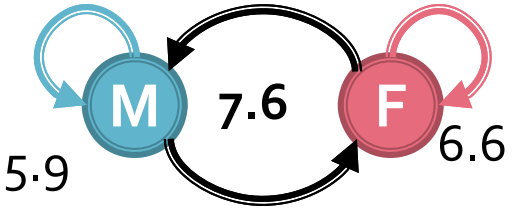
Communication: Gender

- Is gender communication biased?
  - **Homophily:** Do female talk more among themselves?
  - **Heterophily:** Do male-female conversations took longer?
- **Findings:**
  - Num. of. conversations is not biased (follows chance)
  - Cross-gender conversations take longer and are more intense (more attention invested)



Conversations

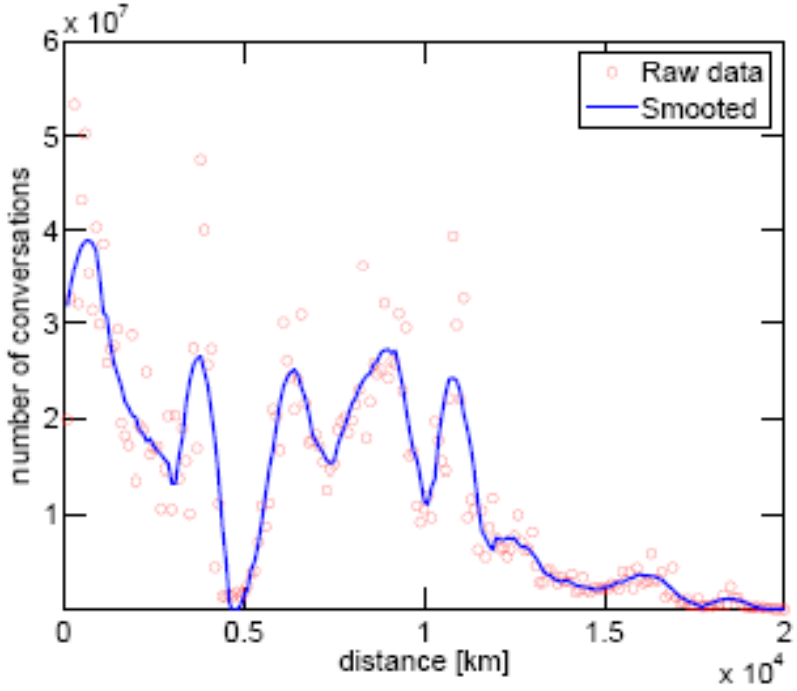
Duration



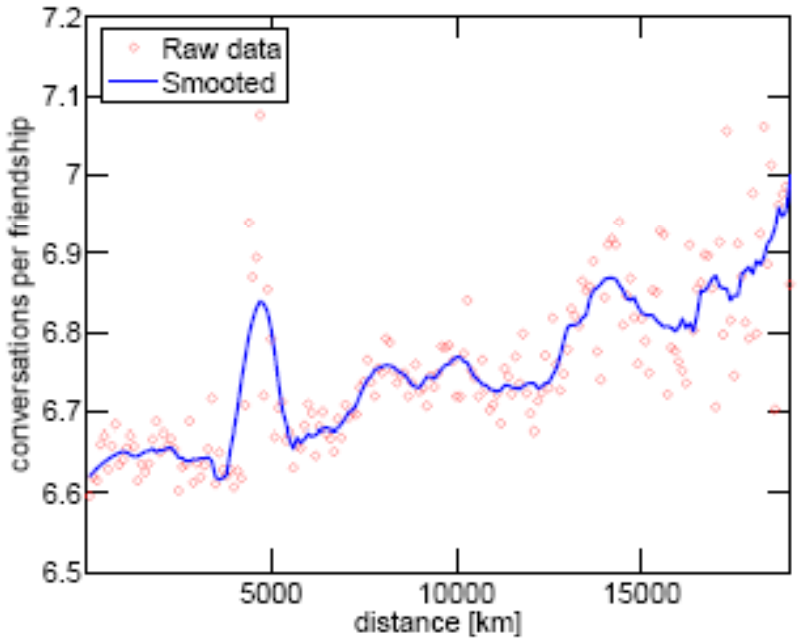
Messages/conversation

Communication: Geo distance

- Longer links are used more

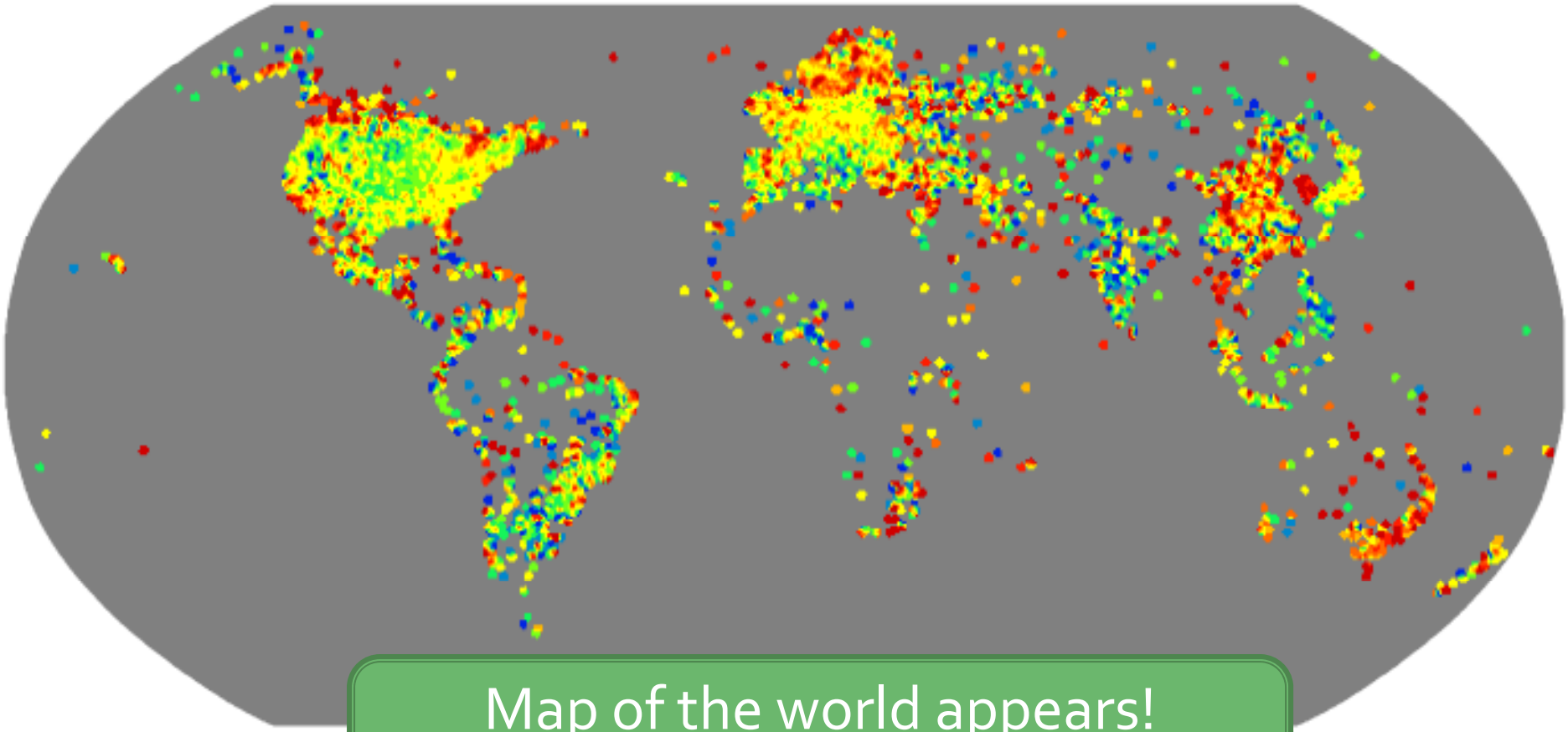


(a) Number of conversations



(b) Conversations per link

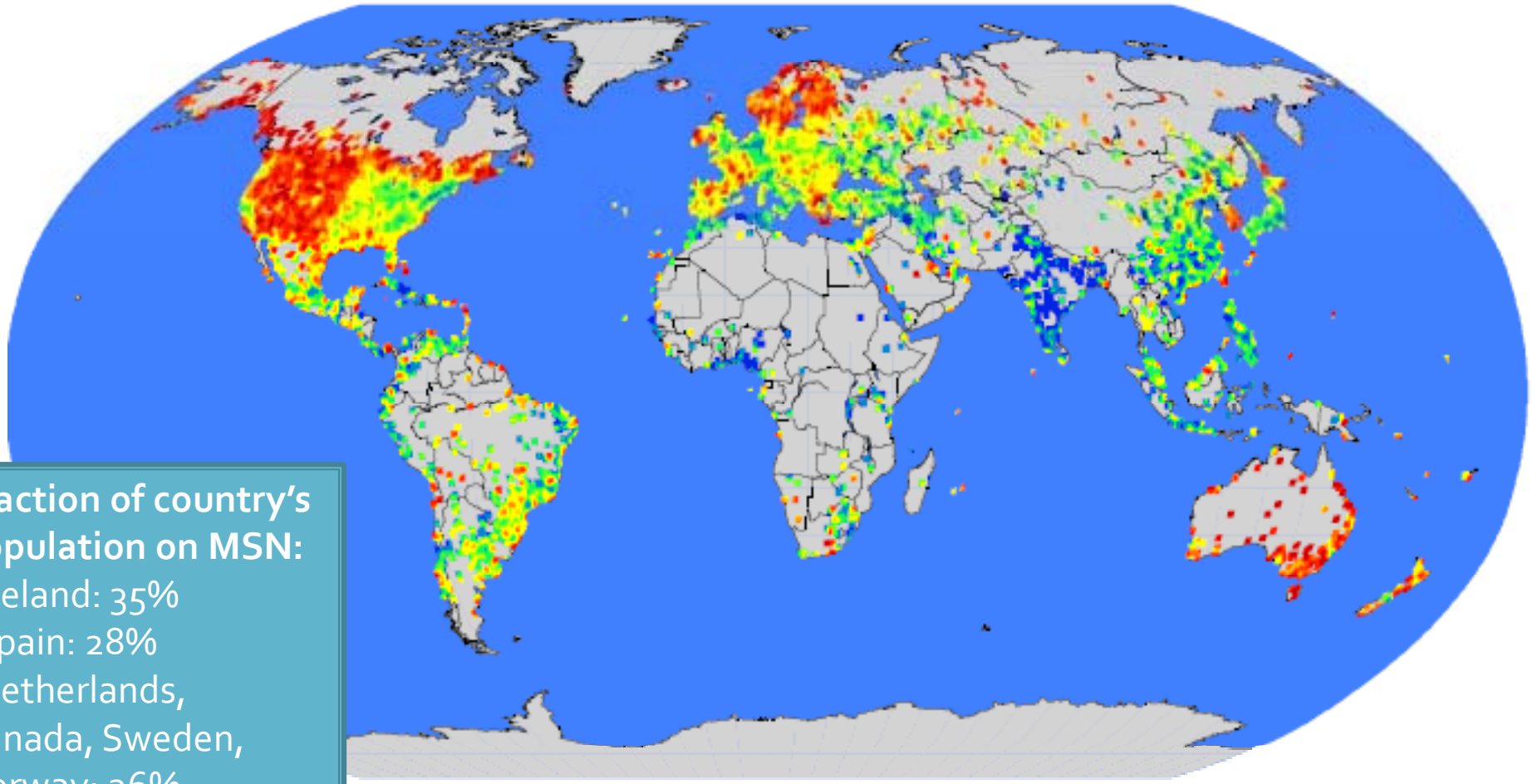
Communication: Geography (1)



Map of the world appears!
Costal regions dominate

Each dot represents number of users at geo location

Communication: Geography (2)

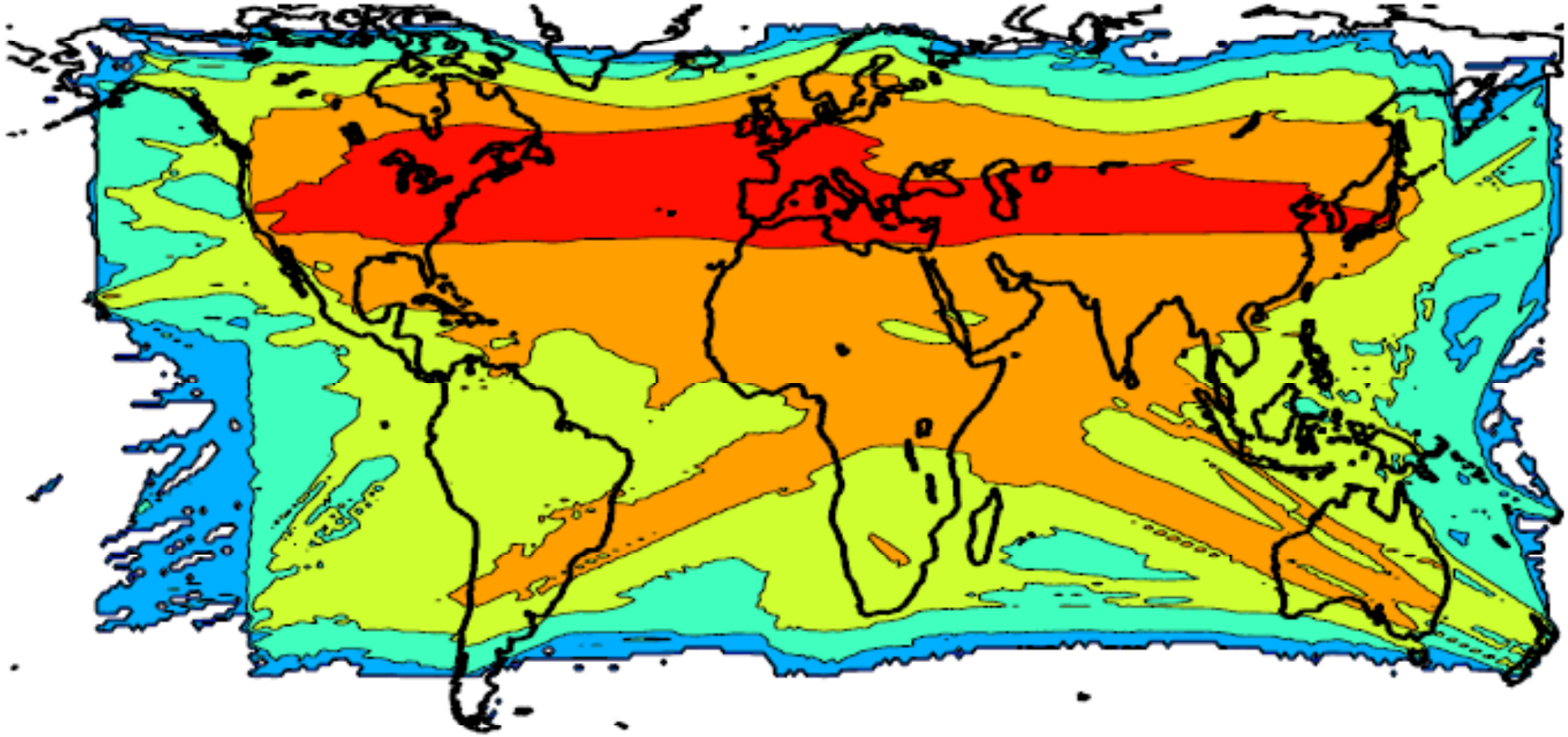


Fraction of country's population on MSN:

- Iceland: 35%
- Spain: 28%
- Netherlands, Canada, Sweden, Norway: 26%
- France, UK: 18%
- USA, Brazil: 8%

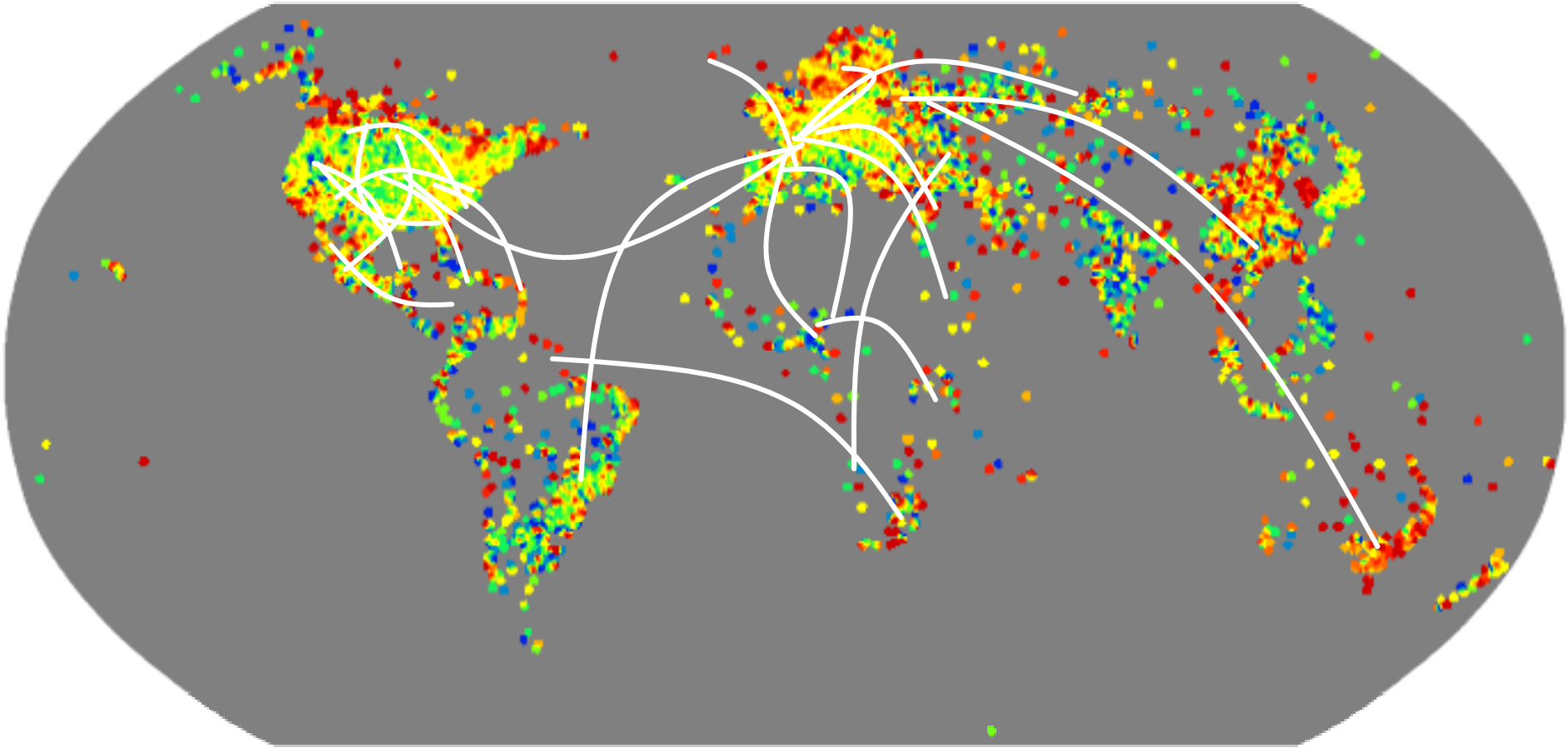
Users per capita

World communication axis

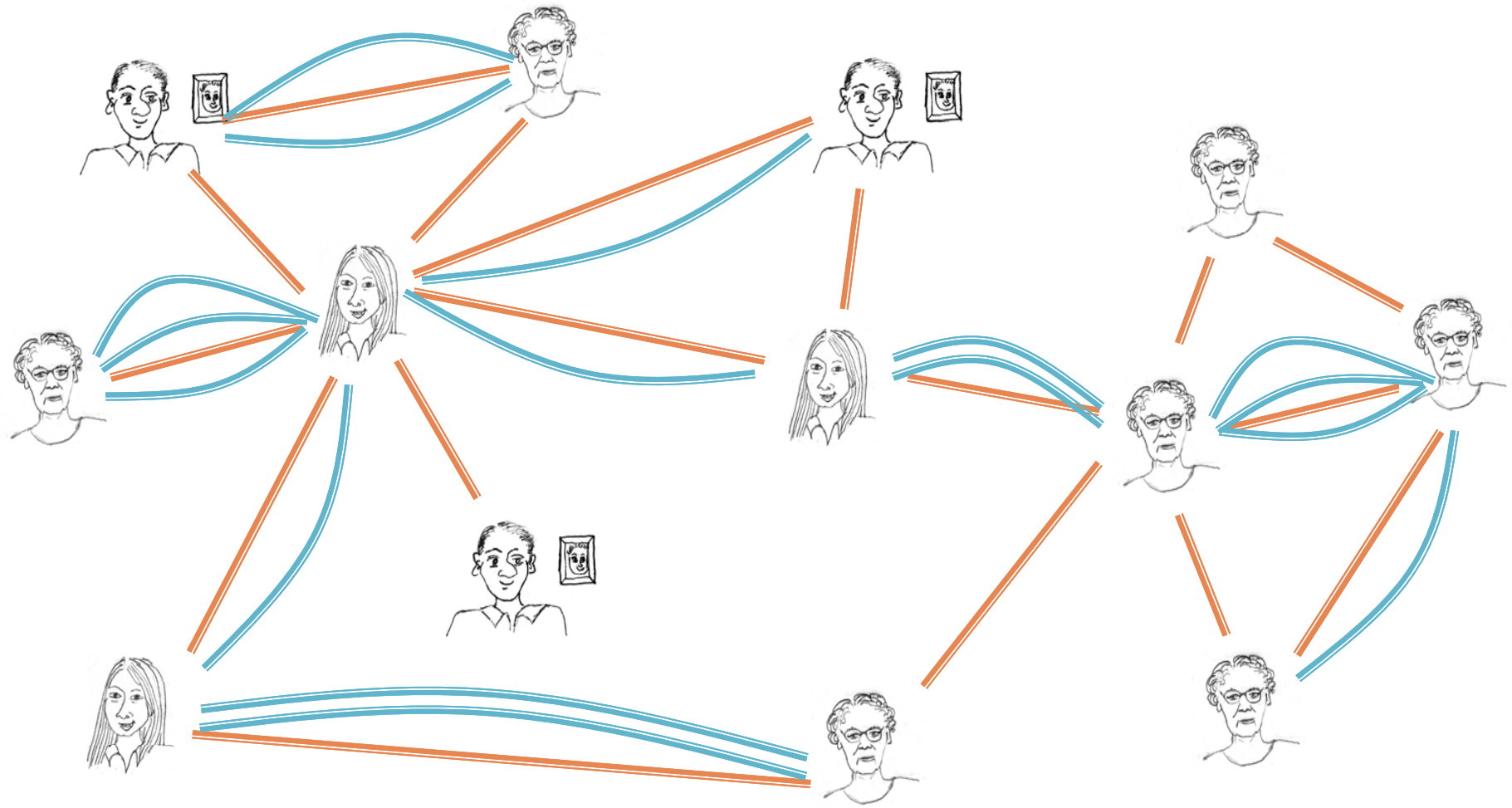


For each conversation between geo points (A,B) we increase the intensity on the line between A and B

Messaging as a Network



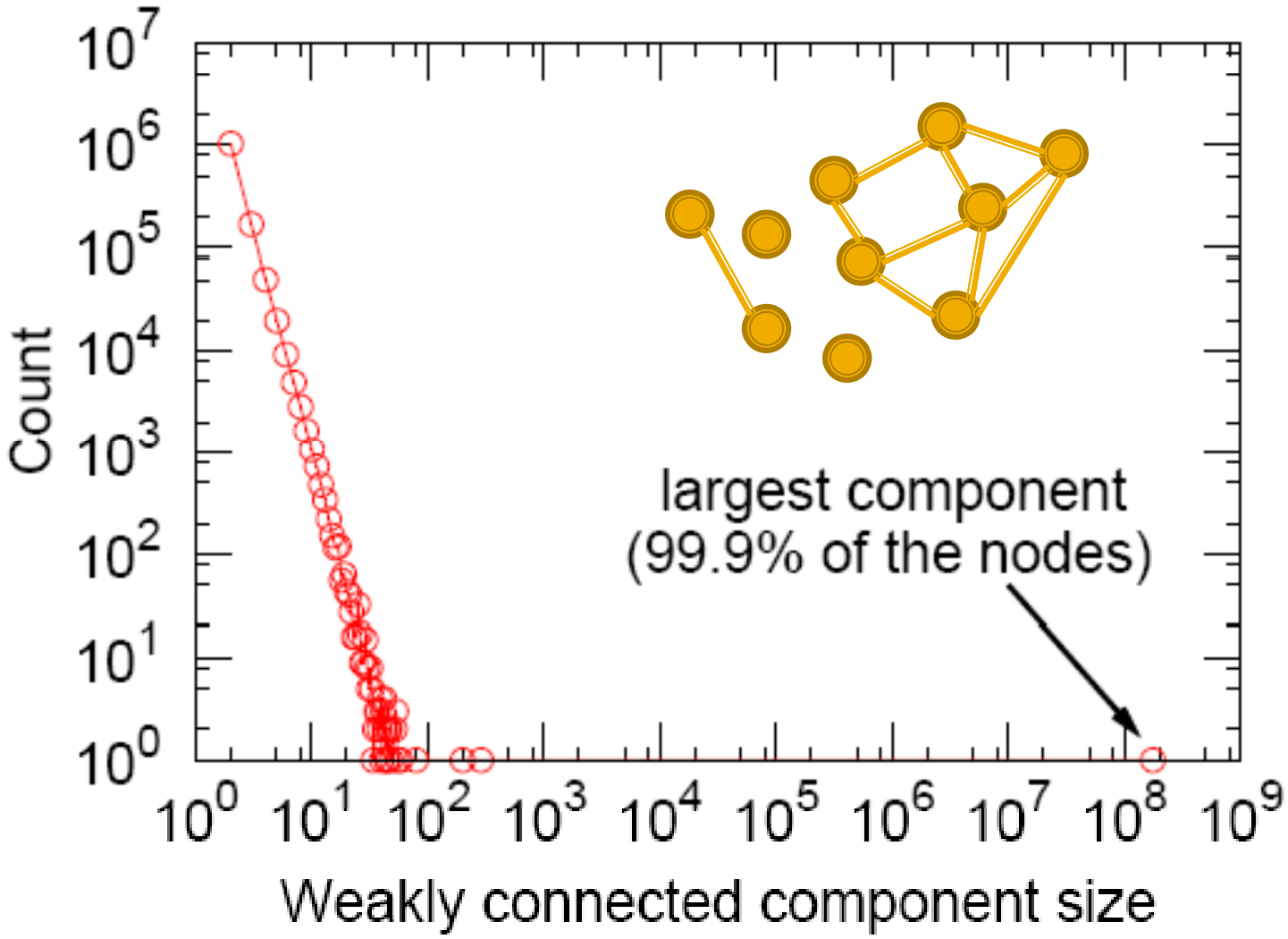
Messaging as a Network



IM Communication Network

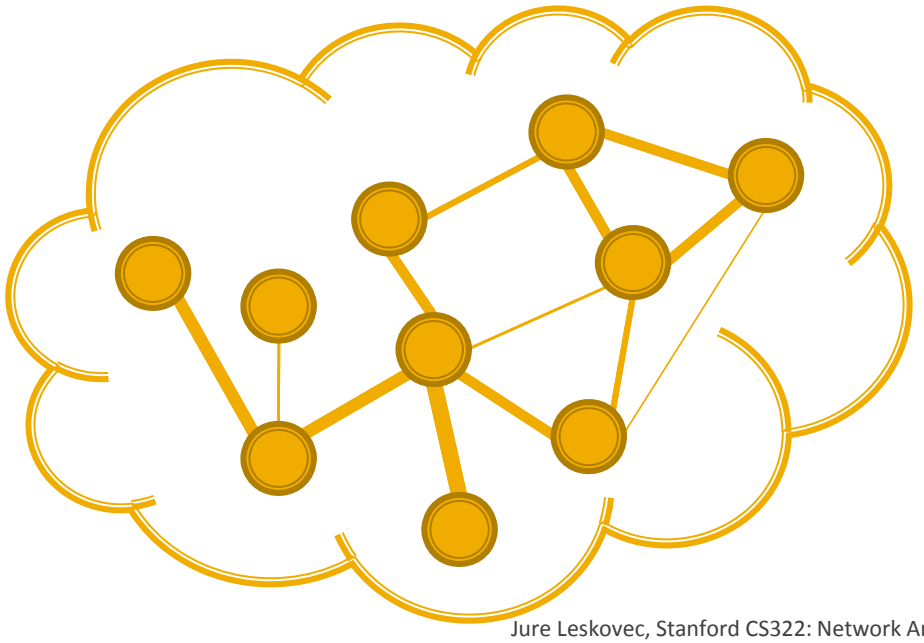
- **Buddy graph**
  - 240 million people (people that login in June ’06)
  - 9.1 billion buddy edges (friendship links)
- **Communication graph** (take only 2-user conversations)
  - Edge if the users exchanged at least 1 message
  - 180 million people
  - 1.3 billion edges
  - 30 billion conversations

Network: Connectivity



Network: Tie-strength

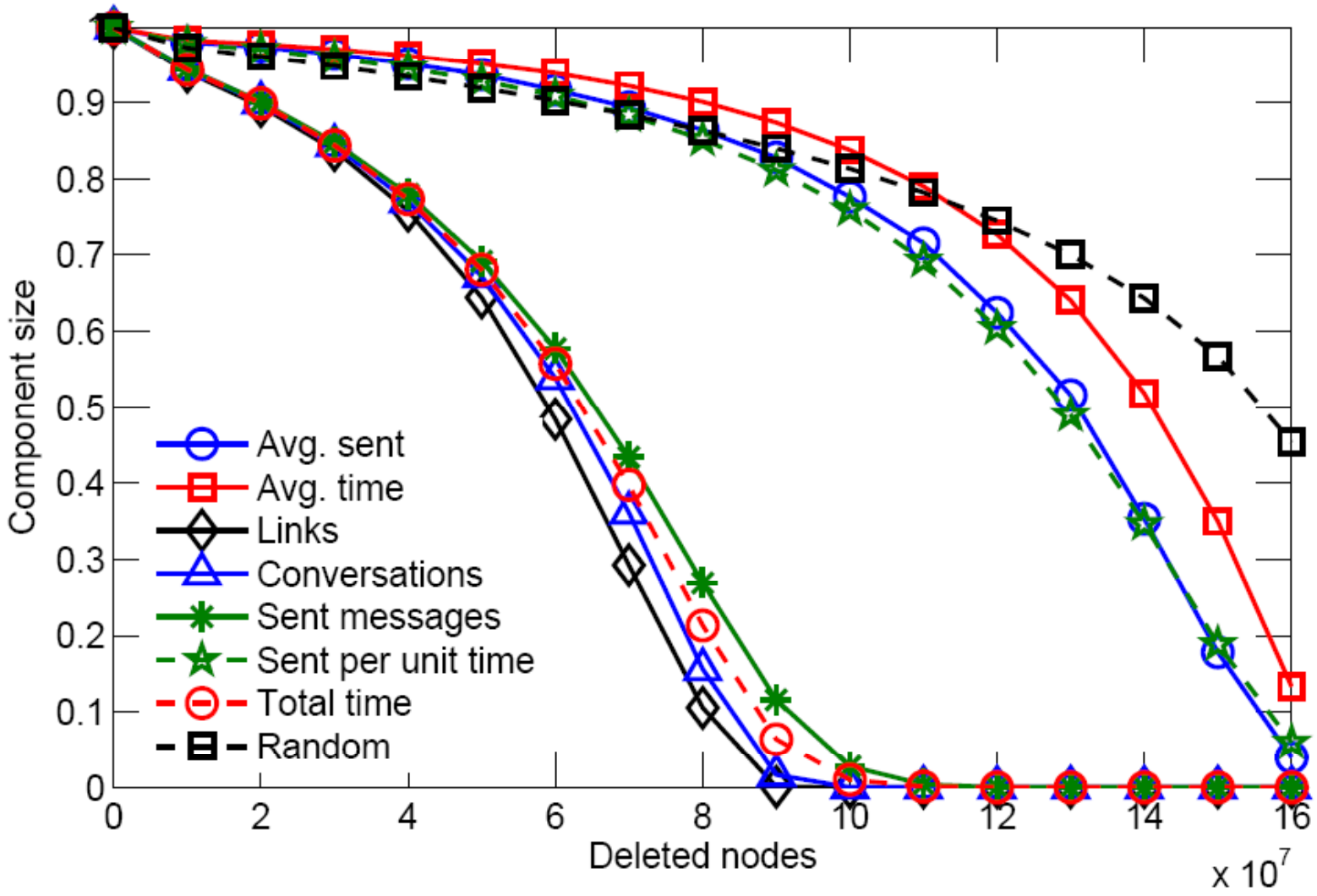
- Remove nodes (in some order) and observe how network falls apart:
  - Number of edges deleted
  - Size of largest connected component



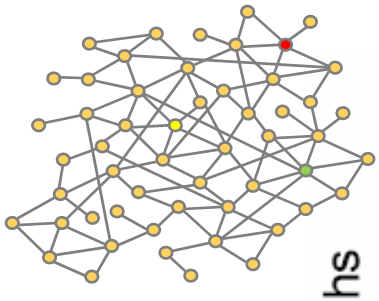
Order nodes by:

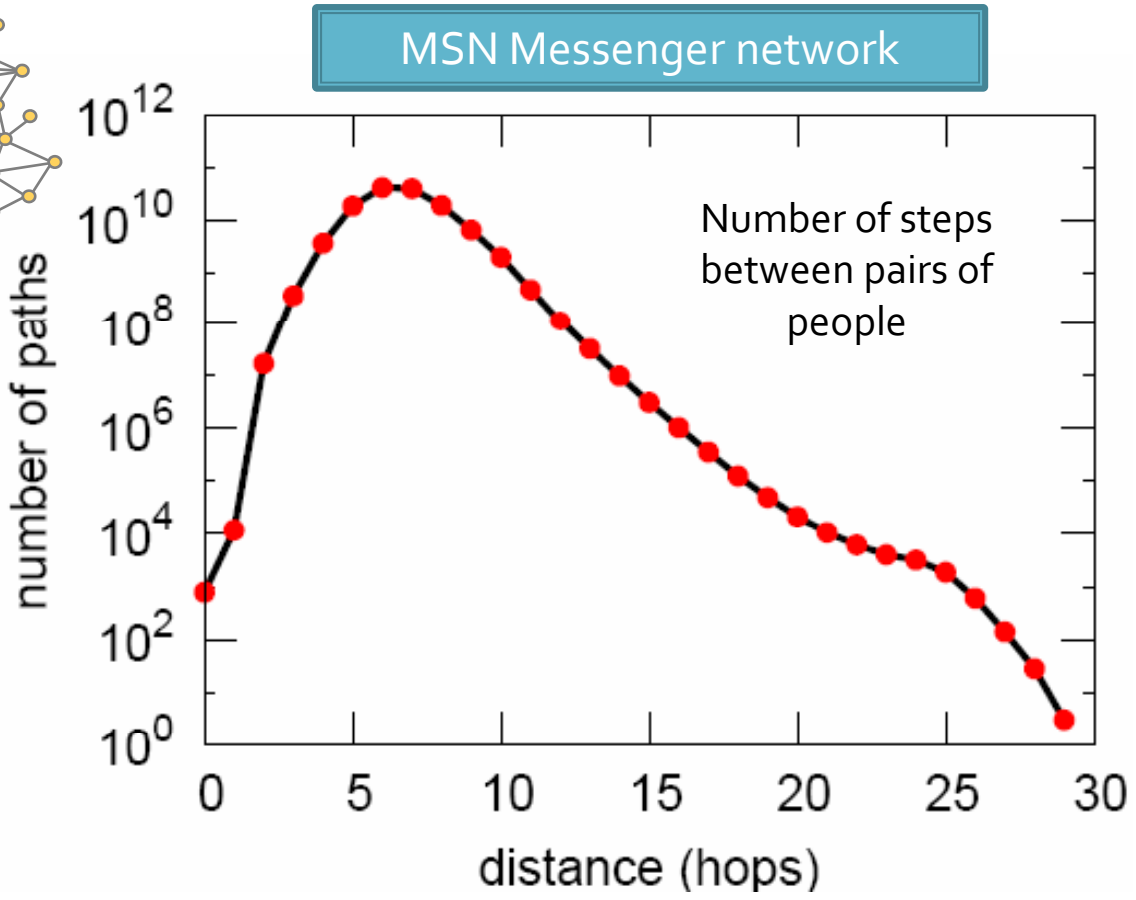
- Number of links
- Total conversations
- Total conv. Duration
- Messages/conversation
- Avg. sent, avg. duration

Strength: Connectivity



Network: Small world



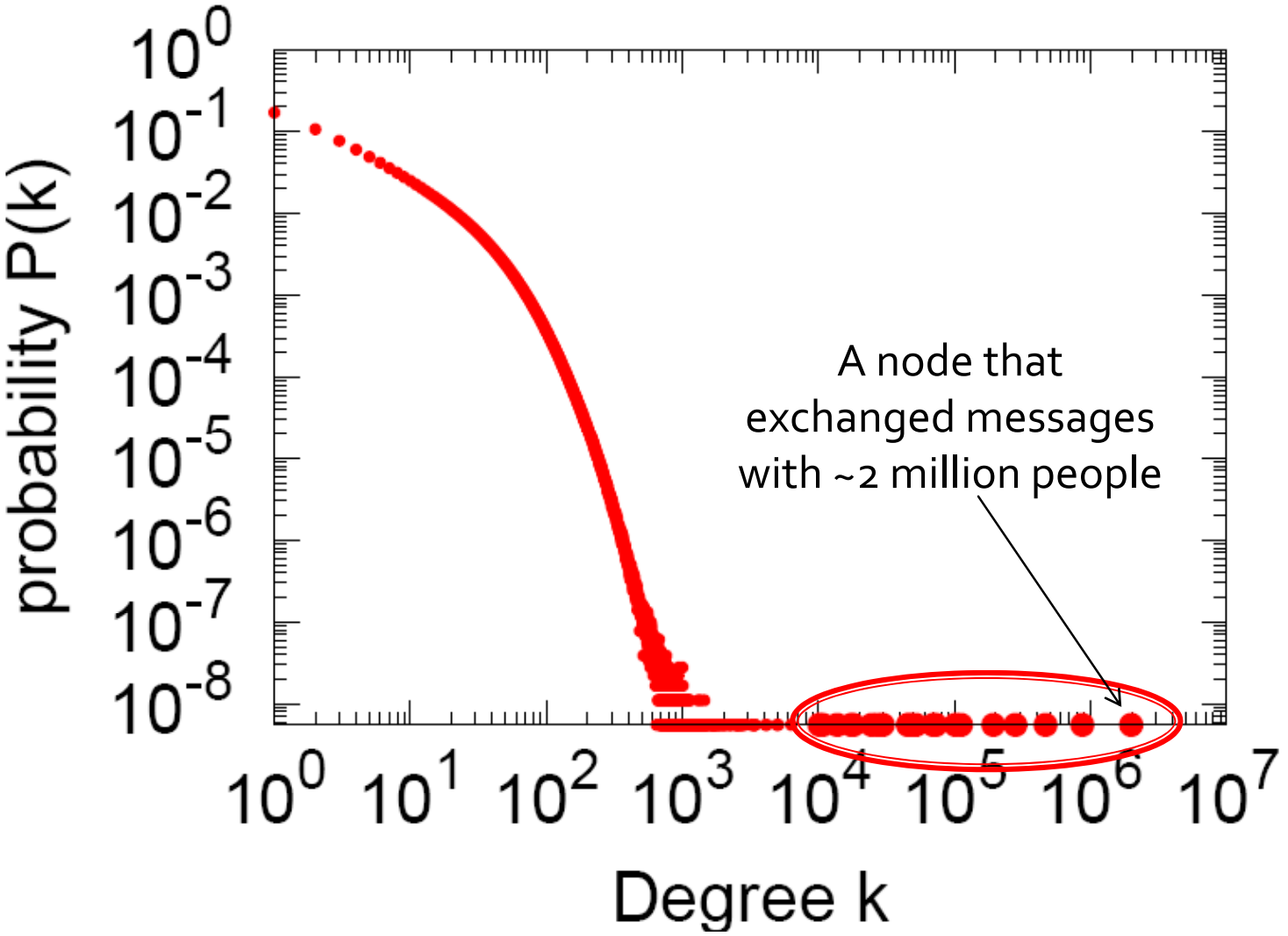


Avg. path length **6.6**

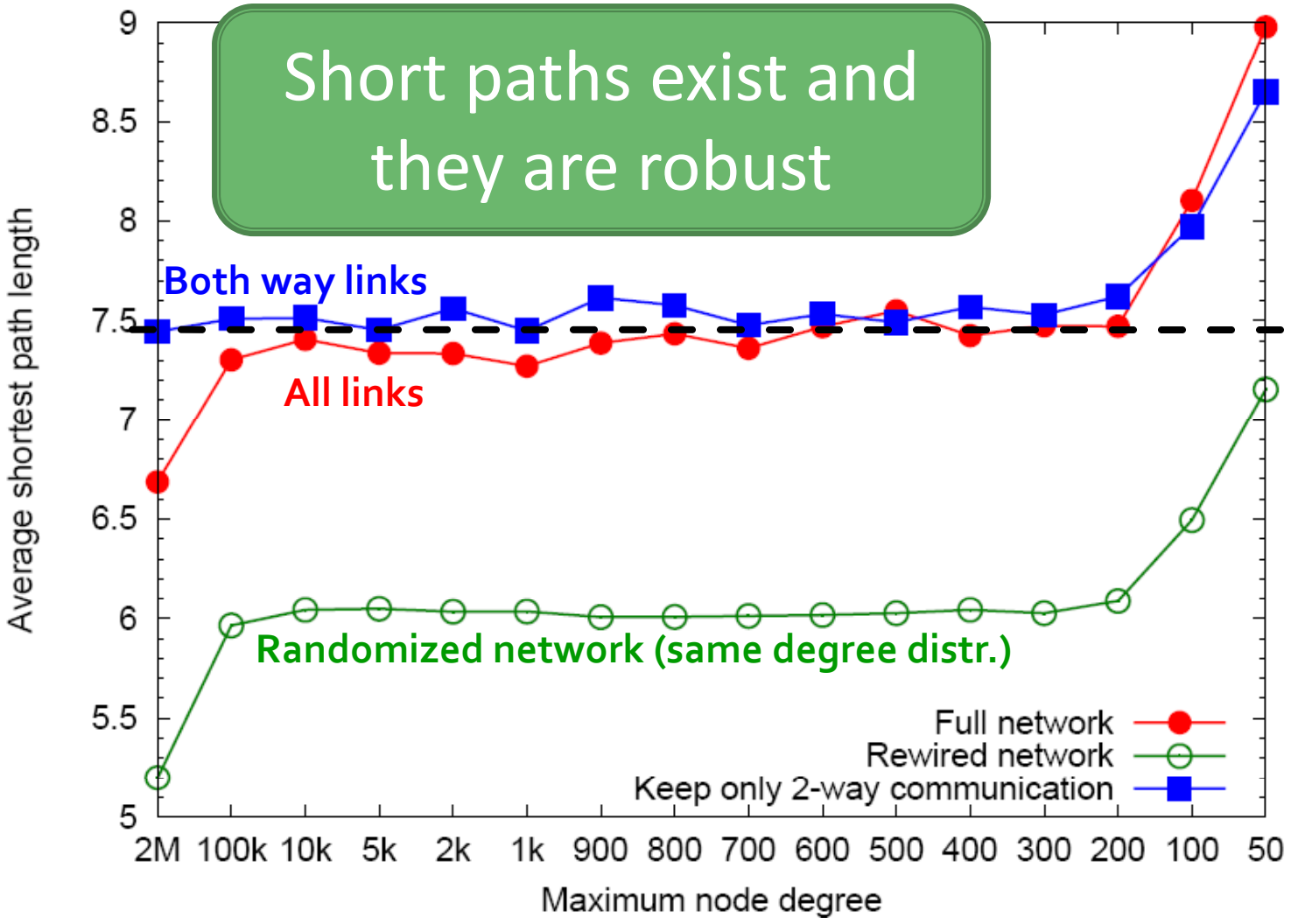
90% of the people can be reached in < 8 hops

Hops	Nodes
0	1
1	10
2	78
3	3,96
4	8,648
5	3,299,252
6	28,395,849
7	79,059,497
8	52,995,778
9	10,321,008
10	1,955,007
11	518,410
12	149,945
13	44,616
14	13,740
15	4,476
16	1,542
17	536
18	167
19	71
20	29
21	16
22	10
23	3
24	2
25	3

Number of connections



Robustness of shortest paths



Compact nations

County	Country	Avg Path Len [hops]
Turkey	Turkey	5.18
Brazil	Brazil	5.60
Belgium	Belgium	5.63
United Kingdom	United Kingdom	5.63
Spain	Spain	5.72
Mexico	Mexico	5.72
France	France	6.03
China	China	6.38
United States	United States	6.96

- Degrees of separation (avg. shortest path length) inside the country

USA: Degrees of separation

County	Country	Avg Path Len [hops]
United States	Lebanon	6.17
United States	Australia	6.22
United States	Norway	6.23
United States	Albania	6.24
United States	Malta	6.24
United States	United Kingdom	6.28
United States	Bahamas	6.29
United States	Sweden	6.37
United States	Bahrain	6.37
United States	Canada	6.38

Top "close" countries

County	Country	Avg Path Len [hops]
United States	Bulgaria	7.28
United States	Poland	7.39
United States	Russia	7.42
United States	Romania	7.48
United States	Lithuania	7.57
United States	Slovakia	7.84
United States	Korea, South	8.03
United States	Czech Republic	8.05
United States	Japan	8.85

Top "far" countries

Six Degrees of Kevin Bacon

- Bacon number:
  - Create a network of Hollywood actors
  - Connect two actors if they co-appeared in the movie
  - Bacon number: number of steps to Kevin Bacon
- As of Dec 2007, the highest (finite) Bacon number reported is 8
- Only approx. 12% of all actors cannot be linked to Bacon