Directions in Pervasive and Social Computing: *On Activities, Privacy, and the Crowd*

Eric Horvitz
Microsoft Research

PerCom 2011
Seattle, WA

Learning from Activity Data

What might we learn about human behavior and of world phenomena from logs of the activities of people?

Privacy and Sensing

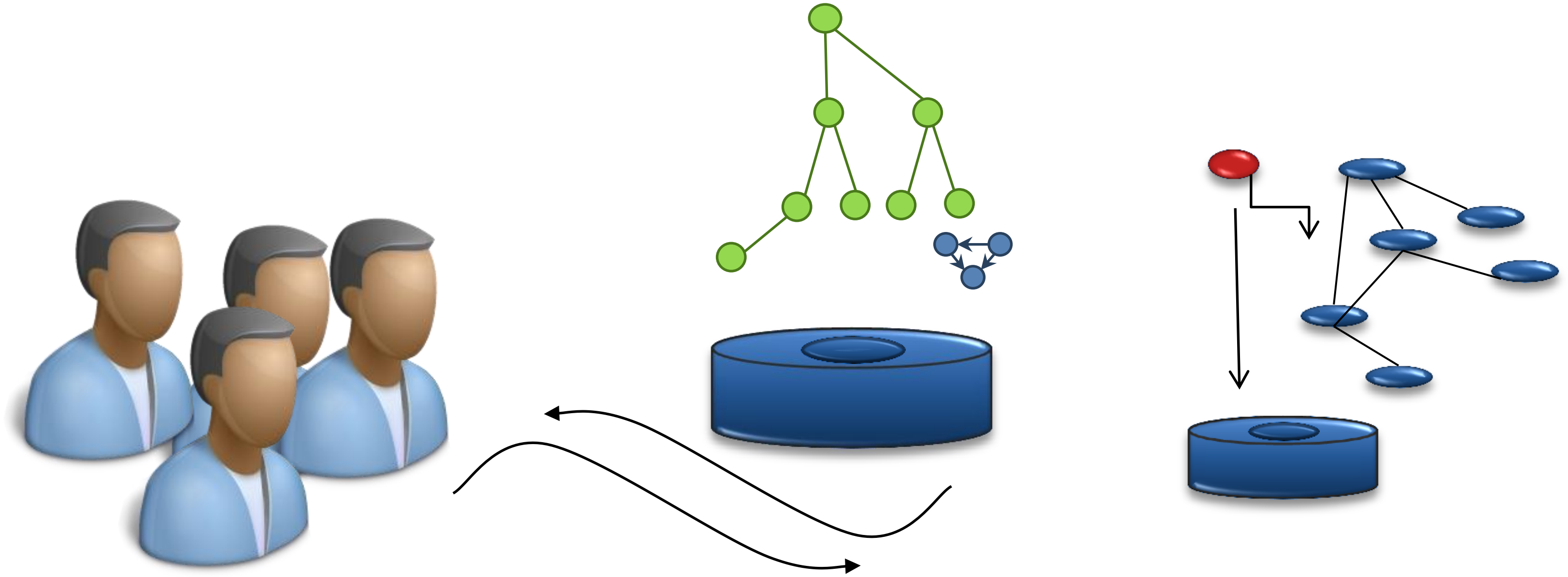
How can we harness data about people while acting in accordance with preferences about privacy?

Crowdsourcing & Collective Intelligence

How might we leverage social connectivity and awareness in the crowdsourcing of solutions?

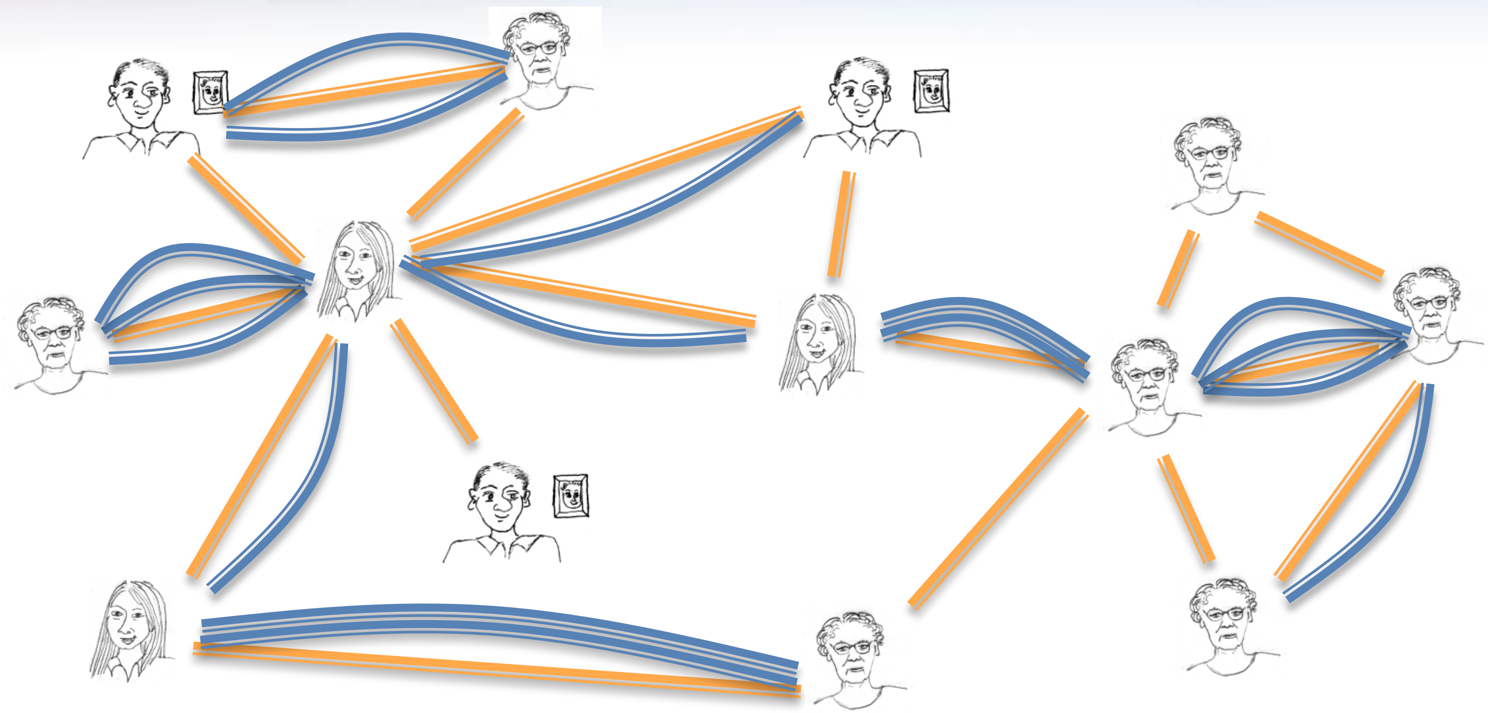
New Lenses on Behavior, Relations, Goals

- Data from widely used communication systems
  - *GPS, Wifi, cell tower, IM, searching, app usage*
- Insights about people
- Sensing phenomena in world



Instant Messaging as Laboratory

Buddy Conversation



[Leskovec & Horvitz, 2008]

Behavioral Studies at Planetary Scale

Study: One month of Messenger data

- 245 million users logged in
- 180 million users engaged in conversations

Communication graph (two-way)

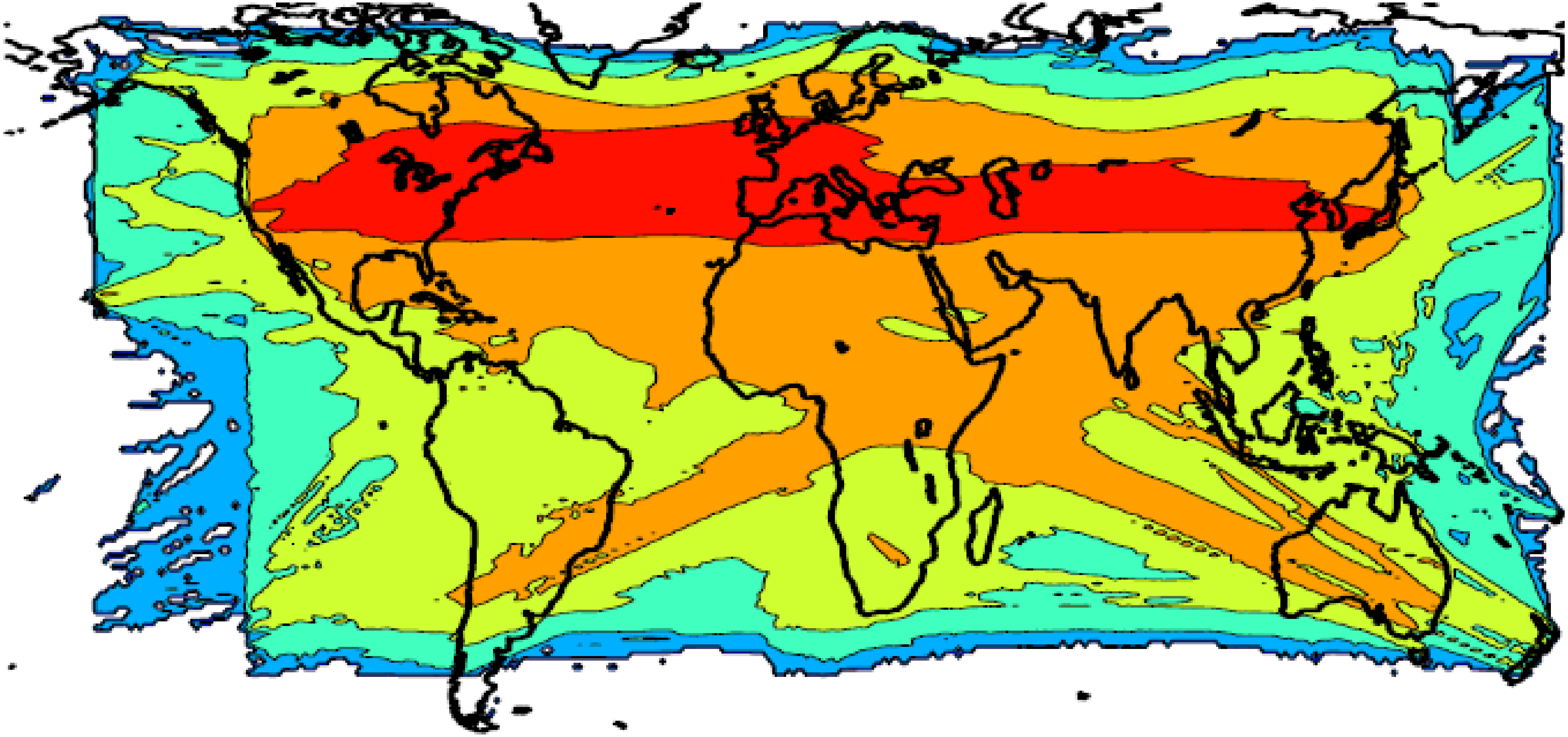
- > 30 billion conversations
- > 255 billion messages exchanged
- 1.3 billion edges

4.5 terabytes

All data anonymized. No message text

Leskovec & Horvitz, 2008

Visualizing World Communication Axis



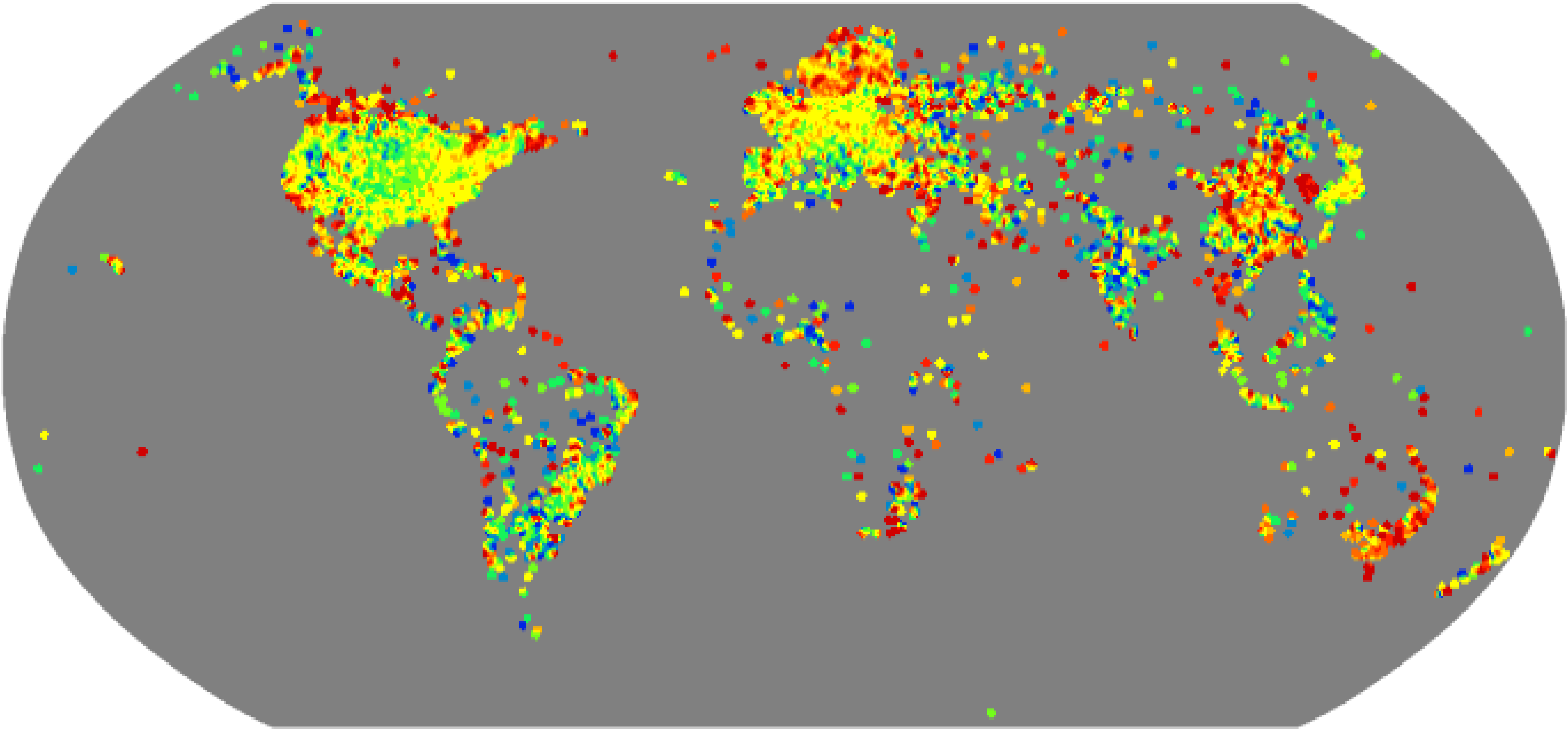
[Leskovec & Horvitz, 2008](#)

Who Talks to Whom?

Width: Numbers of conversations

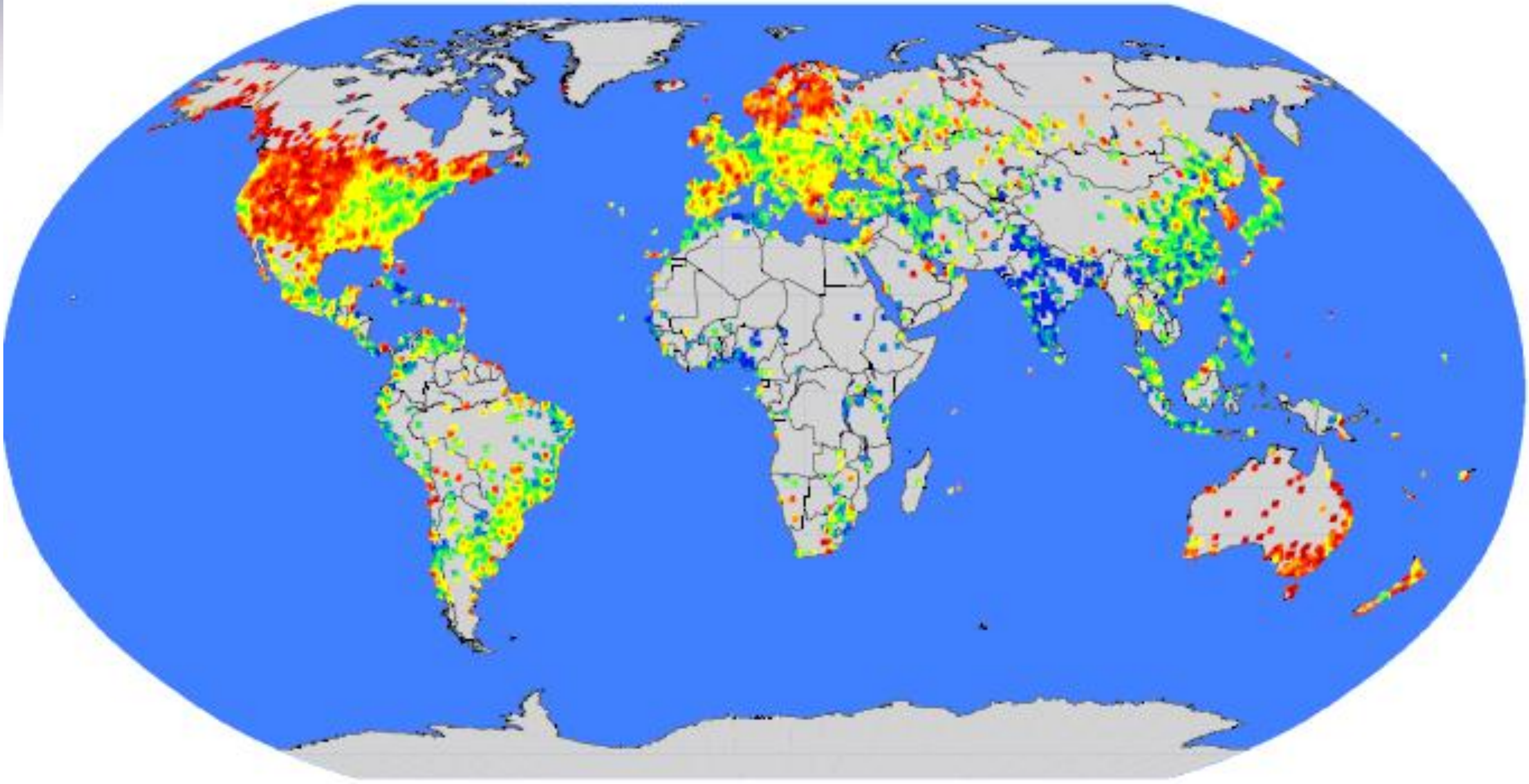
[Leskovec & Horvitz, 2008]

Communication Density



Each point represents number of users at location

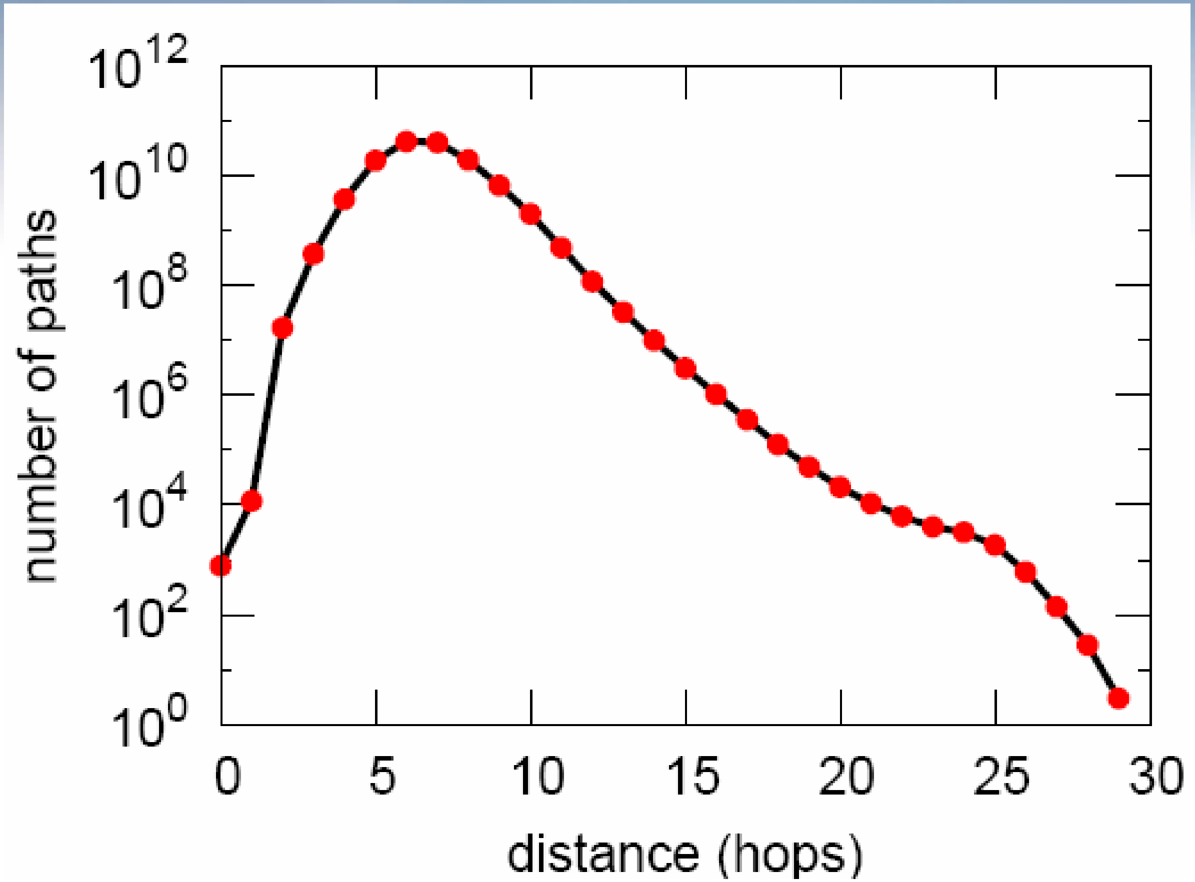
Per Capita Analysis



Users per capita

Leskovec & Horvitz, 2008

Is it a Small-World . . . After all?



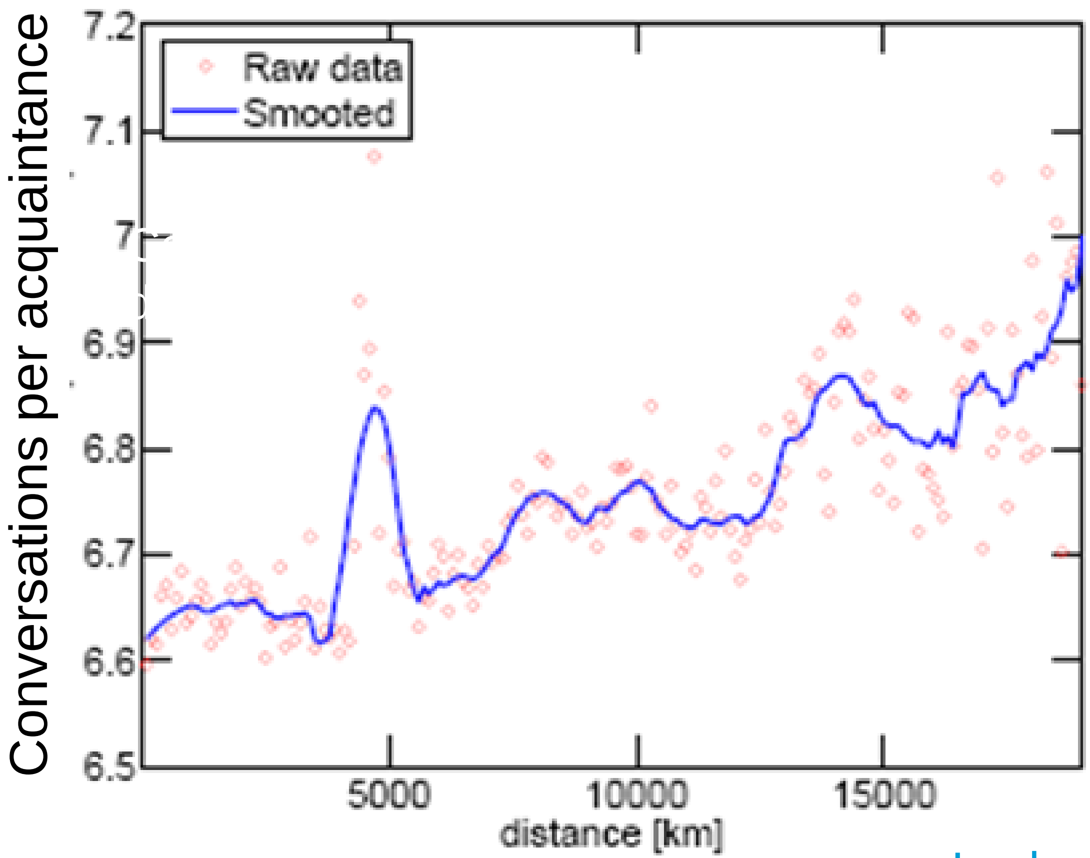
Avg. path length 6.6

90% of others can be reached in < 8 hops

[Leskovec & Horvitz, 2008]

Communication: Geo distance

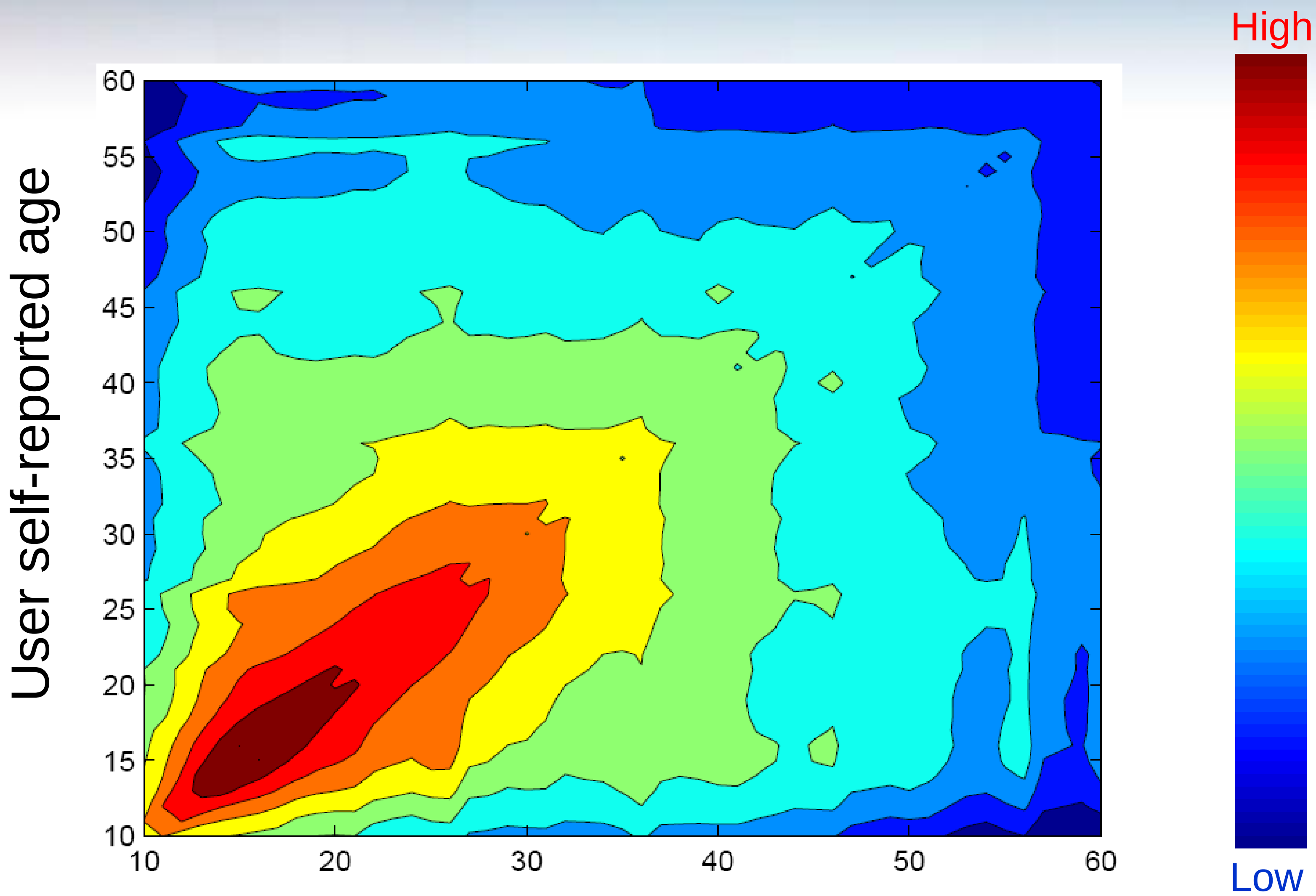
- Longer links used more



Leskovec & Horvitz, 2008

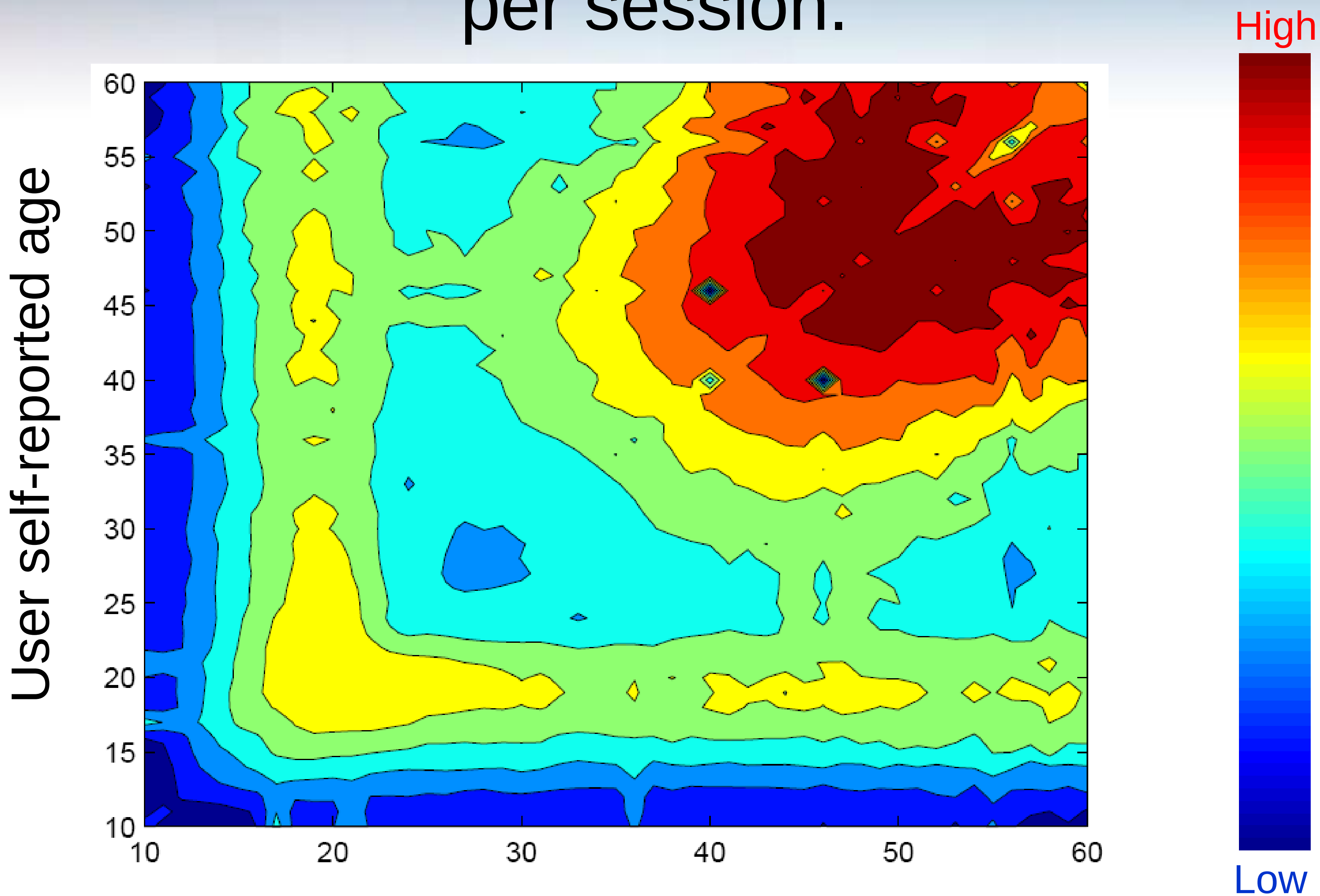
Age: Number of conversations

- Young people communicate with same age



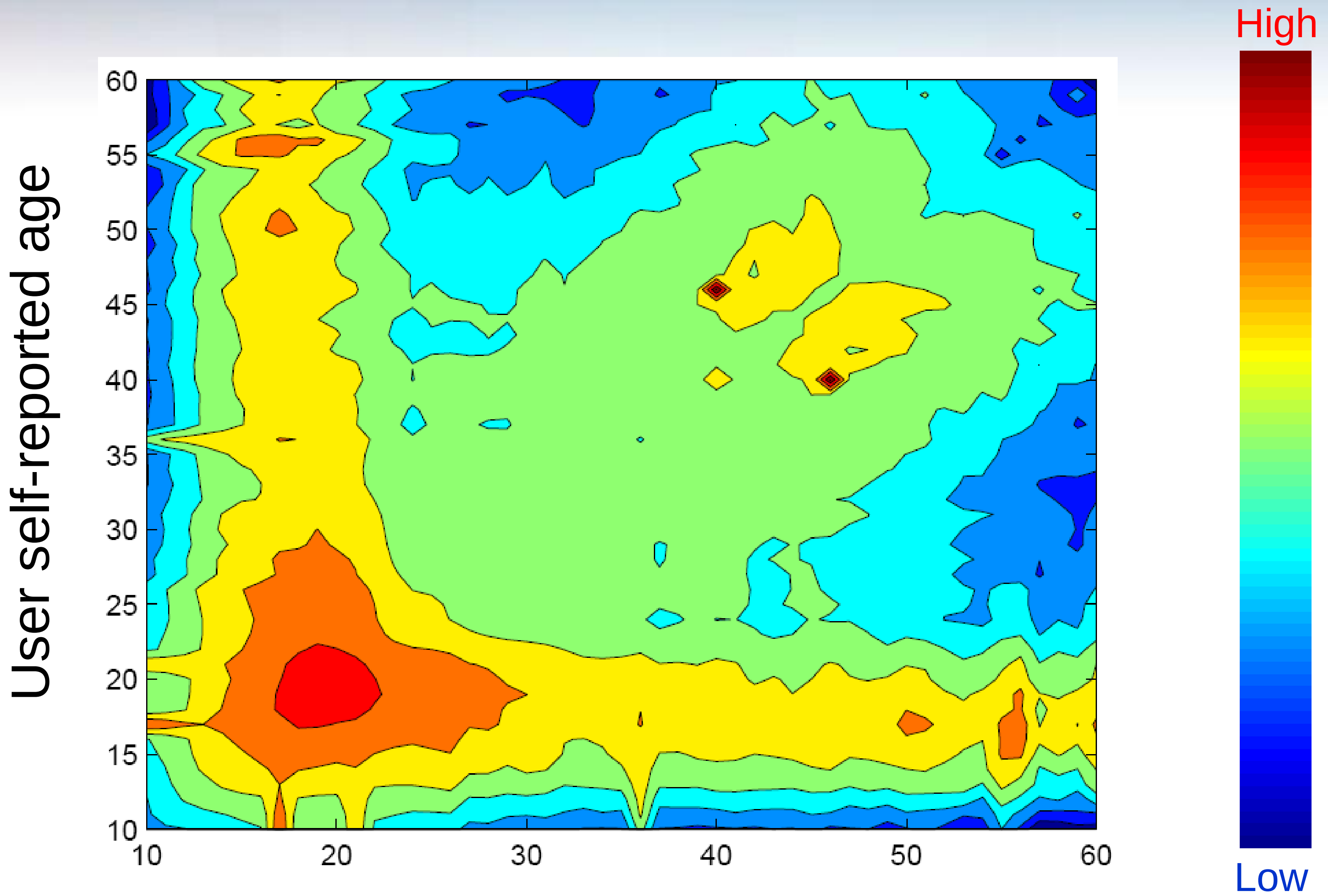
Age: Messages per conversation

- Older people exchange more messages per session.

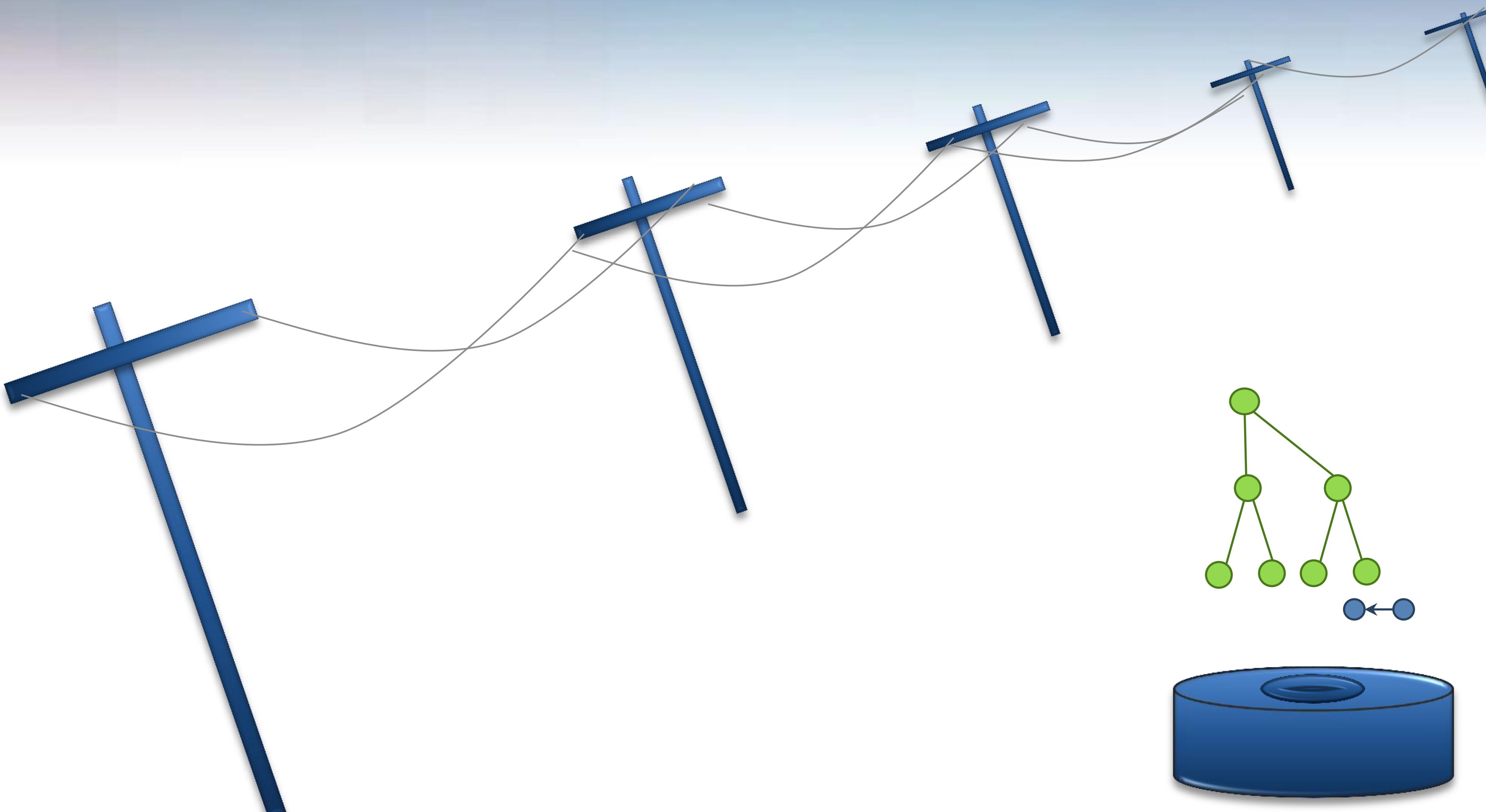


Age: Messages per unit time

- Young people converse more quickly



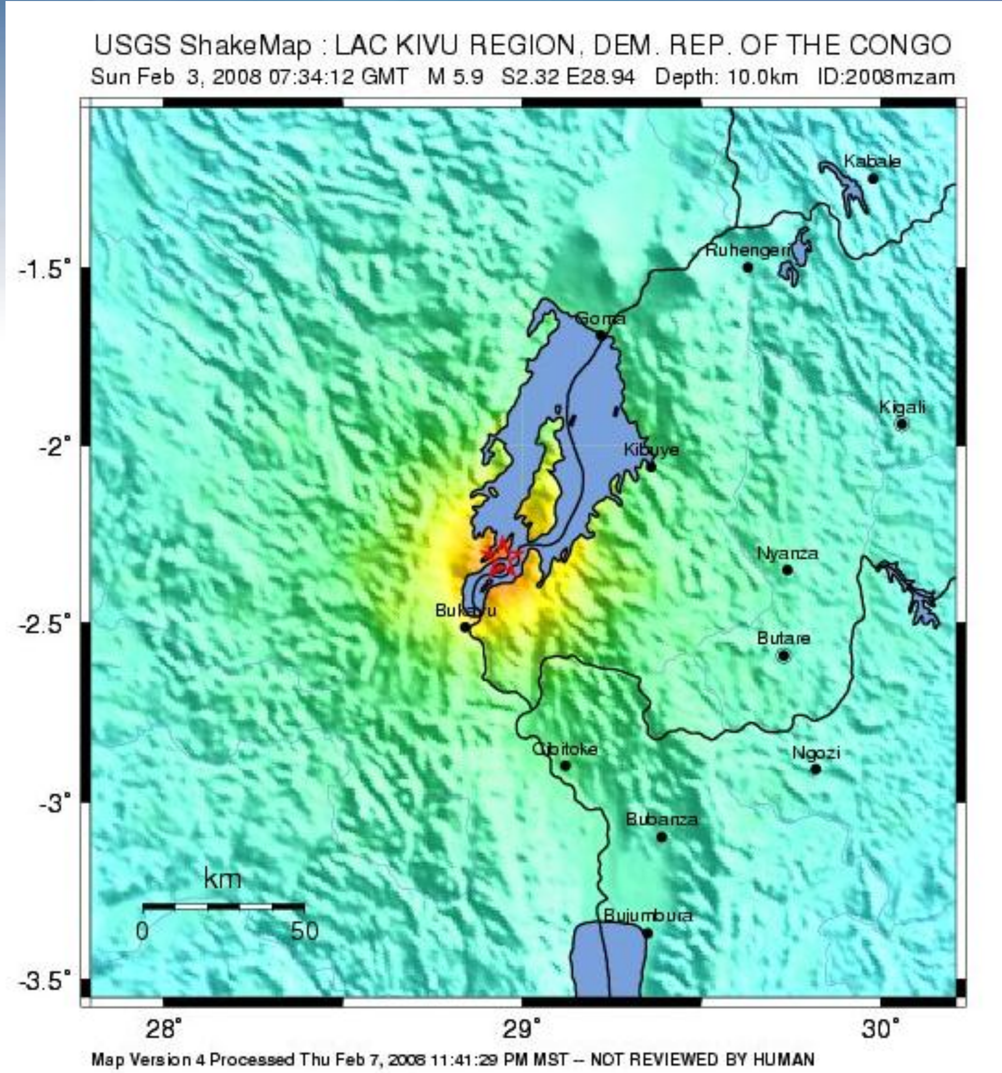
Sensing & Communications Infrastructure



Sensing and World Phenomena

Lac Kivu Quake

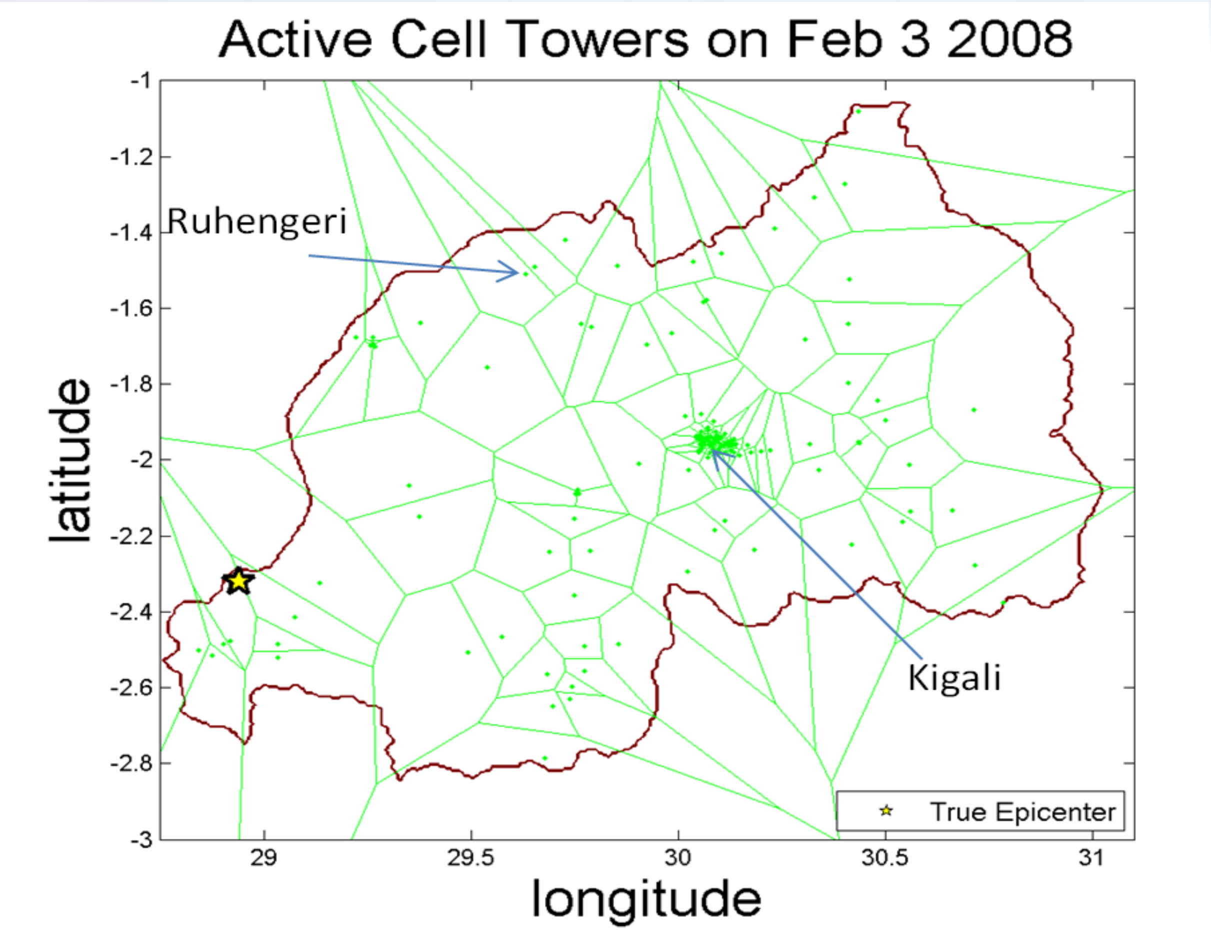
Feb. 3, 2008
5.9



PERCEIVED SHAKING	Not felt	Weak	Light	Moderate	Strong	Very strong	Severe	Violent	Extreme
POTENTIAL DAMAGE	none	none	none	Very light	Light	Moderate	Moderate/Heavy	Heavy	Very Heavy
PEAK ACC.(%g)	<.17	.17-1.4	1.4-3.9	3.9-9.2	9.2-18	18-34	34-65	65-124	>124
PEAK VEL.(cm/s)	<0.1	0.1-1.1	1.1-3.4	3.4-8.1	8.1-16	16-31	31-60	60-116	>116
INSTRUMENTAL INTENSITY	I	II-III	IV	V	VI	VII	VIII	IX	X+

Communications Data Set

- 3 years of all cell communications in Rwanda
  - 140 cell towers, 6 days: 10,527,799 calls

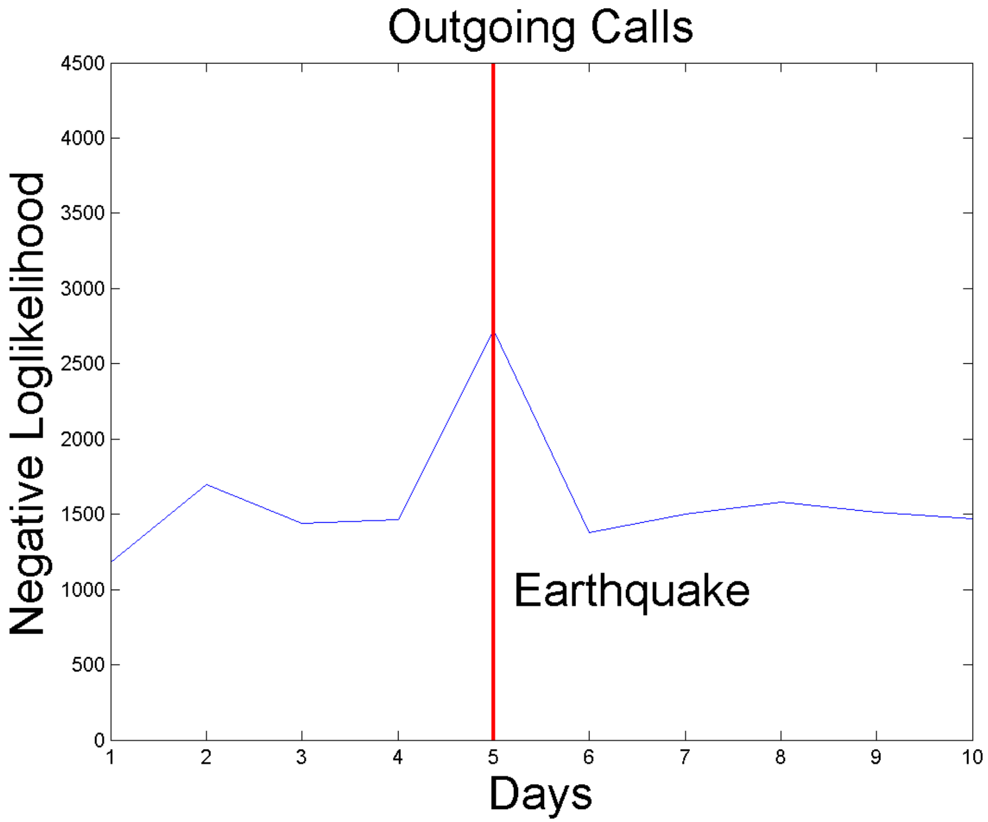


Kapoor, Eagle, Horvitz, 2010

Detecting the Earthquake

- Normal trend at each cell tower

$$p(a_i^t \mid No\ Event) \sim N(m_i, \sigma_i^2)$$



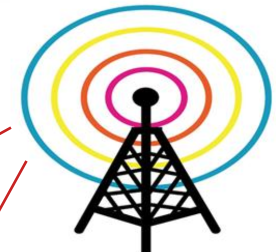
'story

Kapoor, Eagle, Horvitz, 2010

Inferring the Epicenter

- Modeling deviations from the trend

$$p(a_i \mid Event) \sim N(m_i(1+\Delta_i), \Sigma_i)$$



x_i, y_i

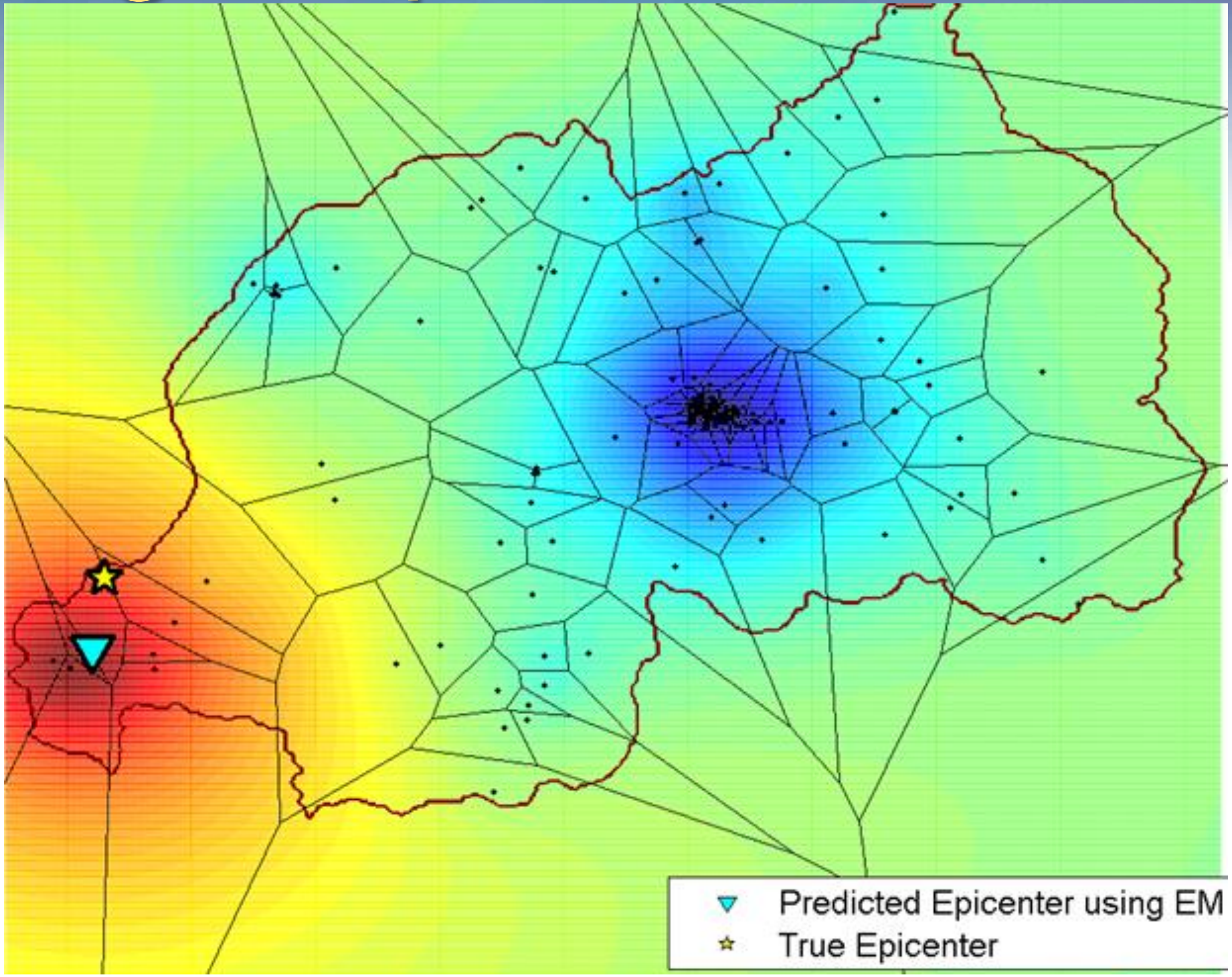
$$\Delta_i = \frac{\alpha}{\beta + \left[(e_x - x_i)^2 + (e_y - y_i)^2\right]^{\gamma}}$$

Unknown parameters: $\theta = \left(\alpha, \beta, \gamma, e_x, e_y\right)$

epicenter

Kapoor, Eagle, Horvitz, 2010

Inferring the Epicenter

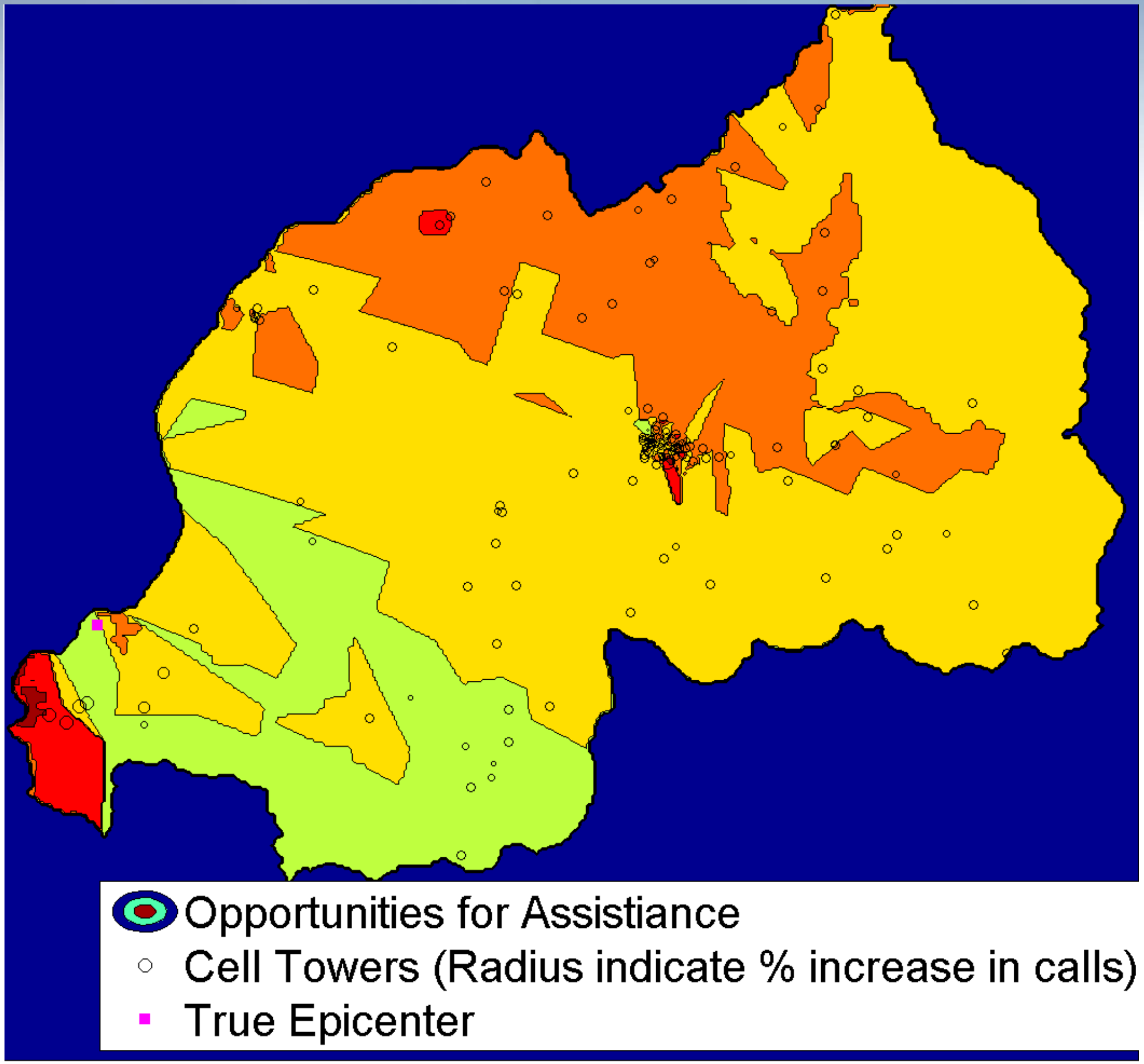


Error: 17.12 km

Kapoor, Eagle, Horvitz, 2010

Persistence: Opportunities to Assist?

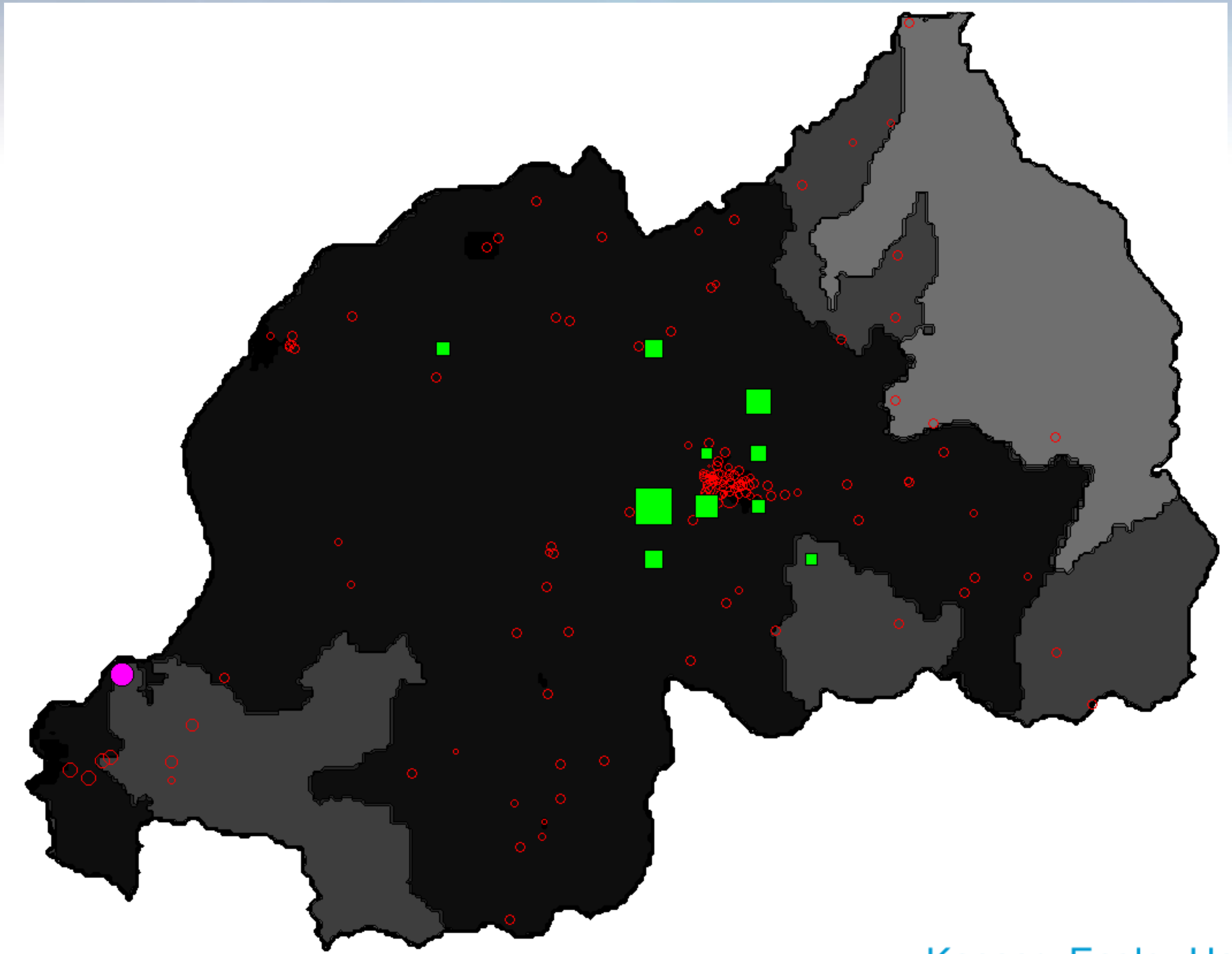
- Opportunities for Assistance (Day 2)



Kapoor, Eagle, Horvitz, 2010

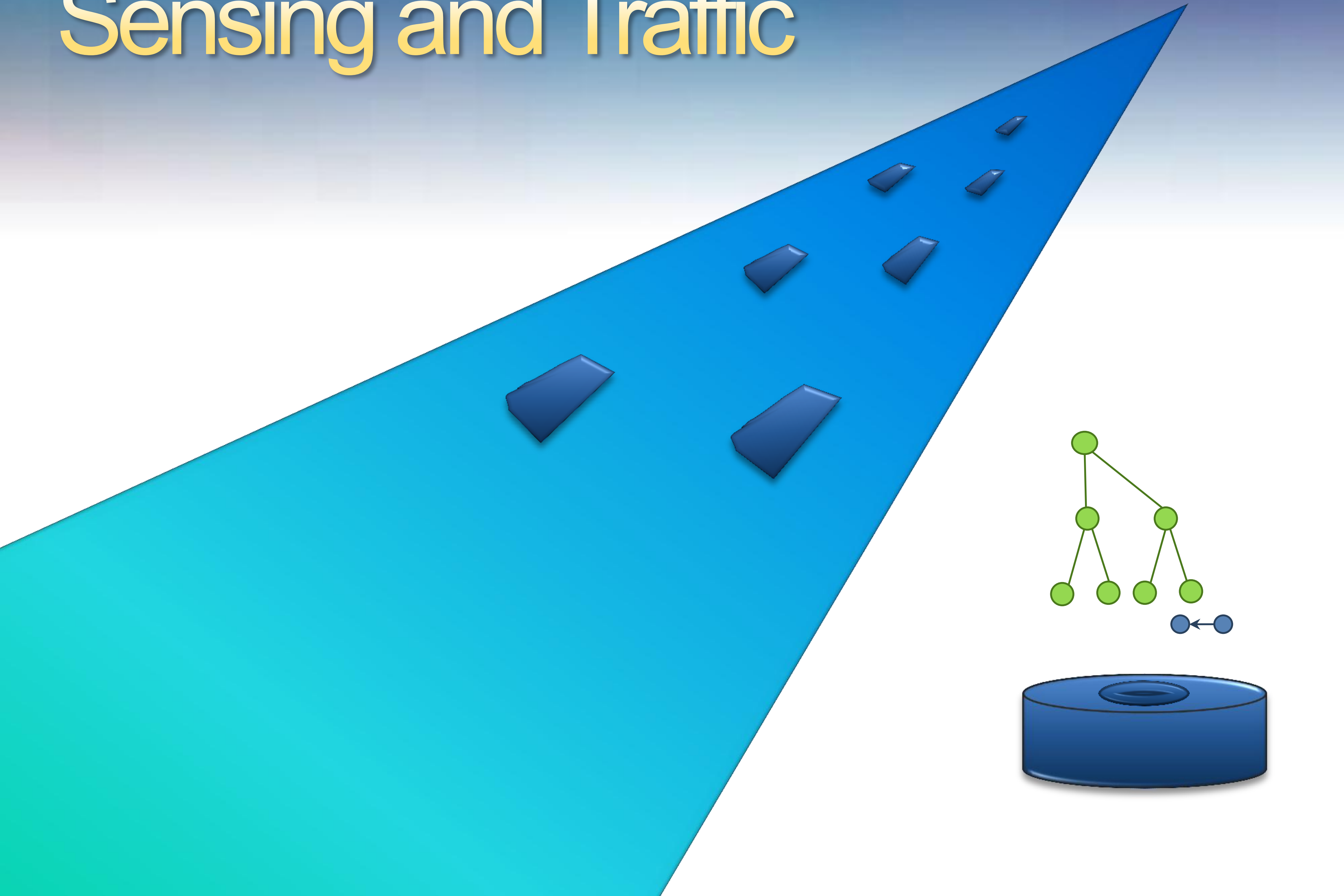
Value of Survey

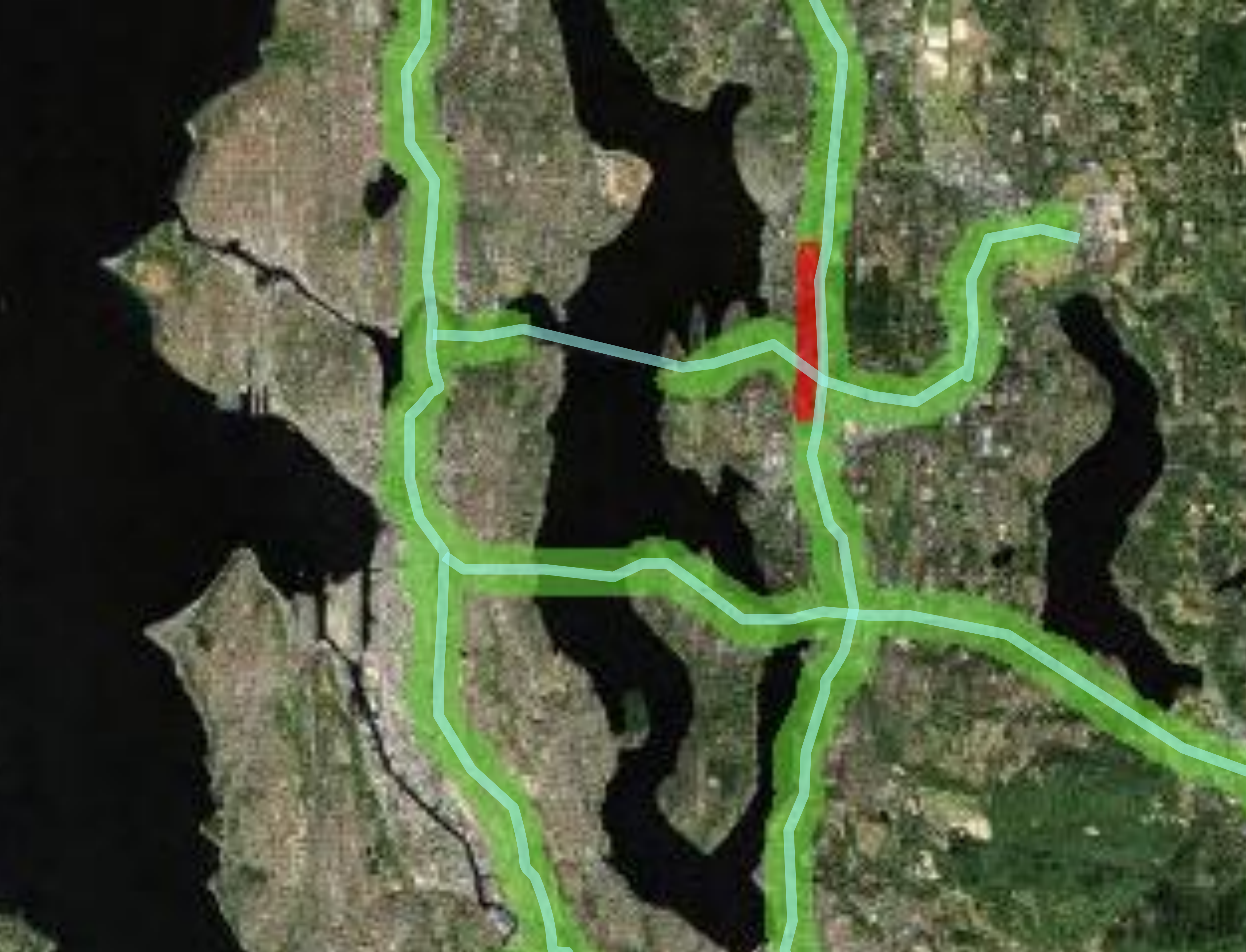
- Ideal Reconnaissance (Day 2)



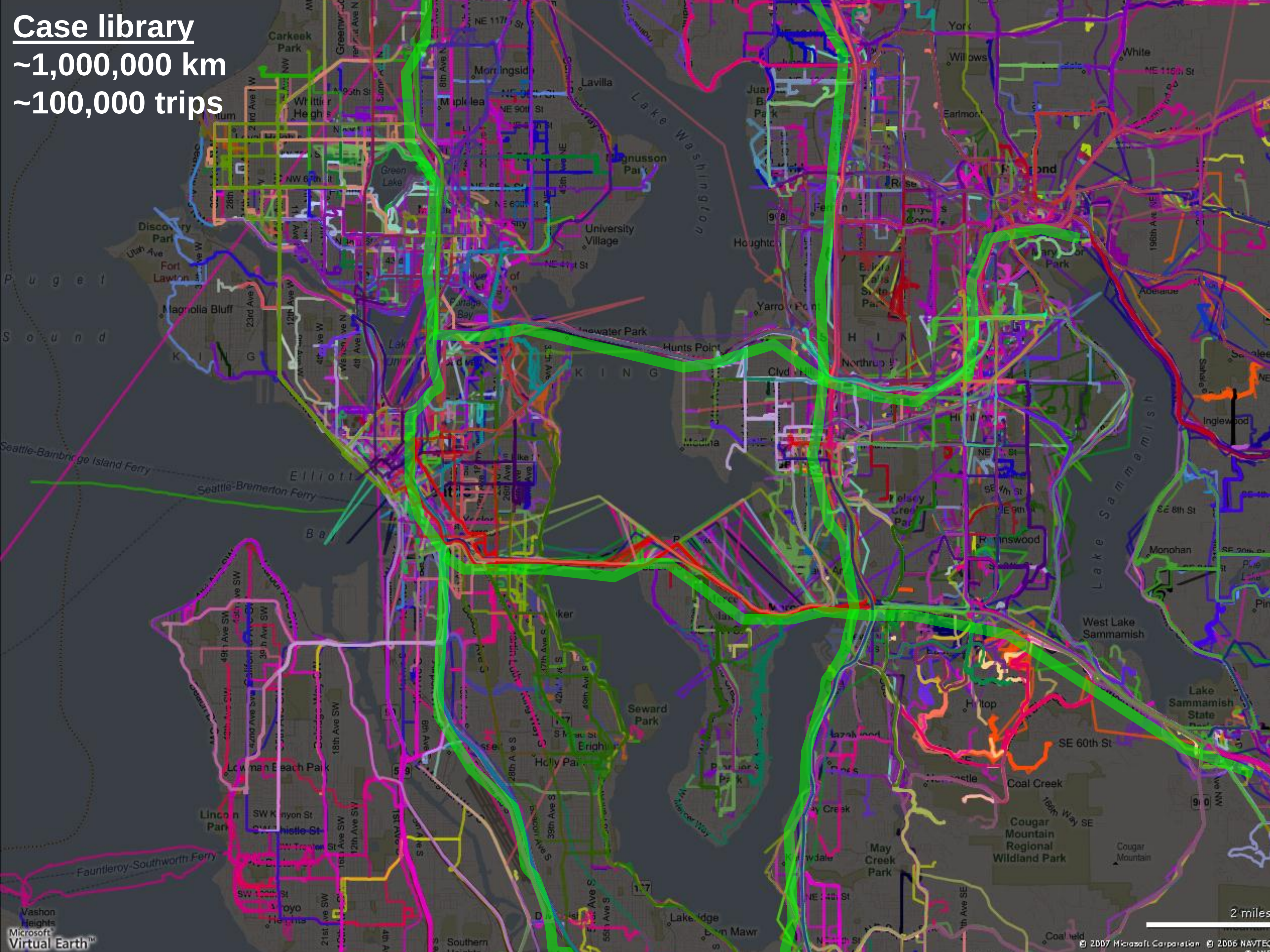
Kapoor, Eagle, Horvitz, 2010

Sensing and Traffic





Case library
~1,000,000 km
~100,000 trips



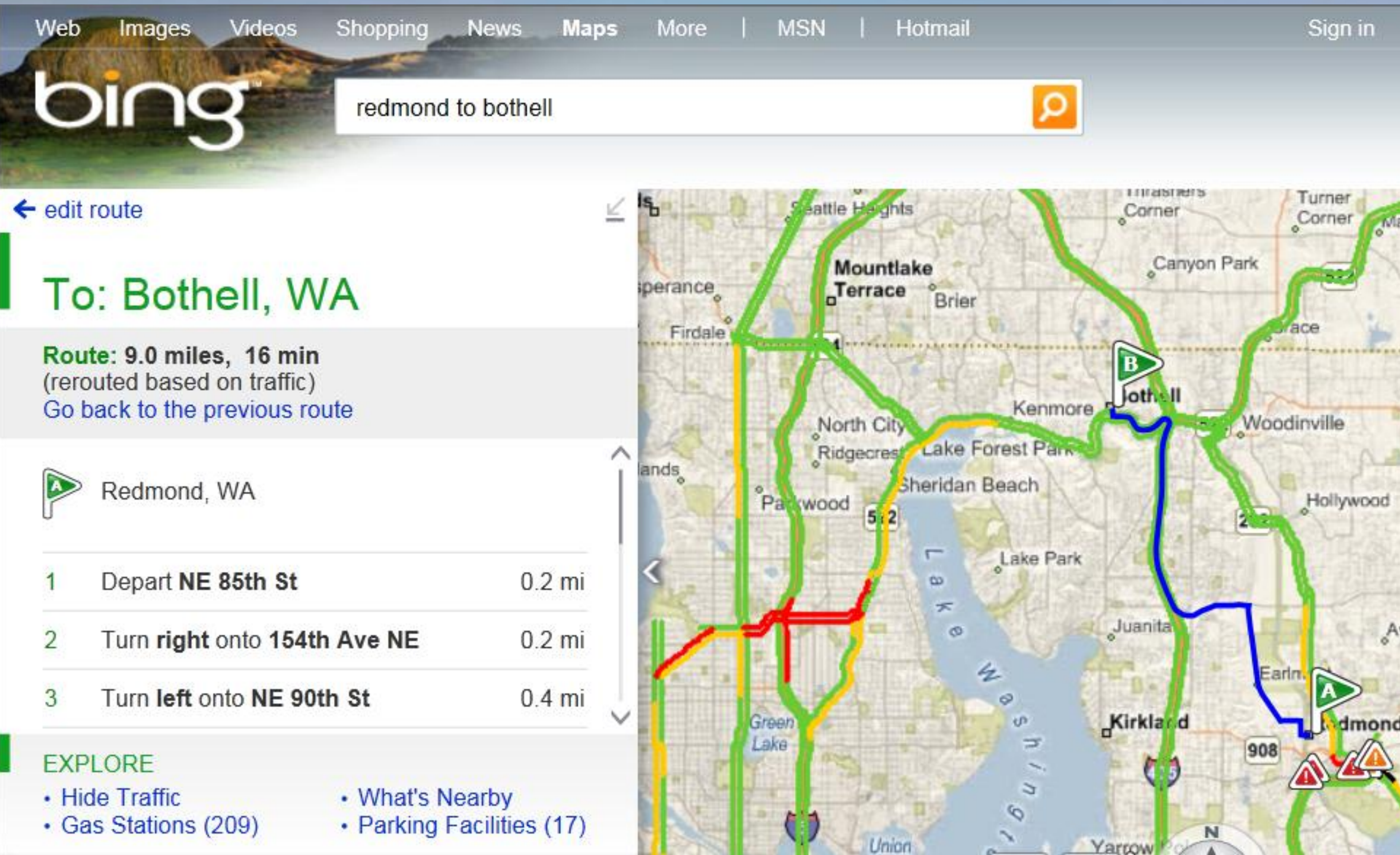
ClearFlow

maps.bing.com * m.bing.com

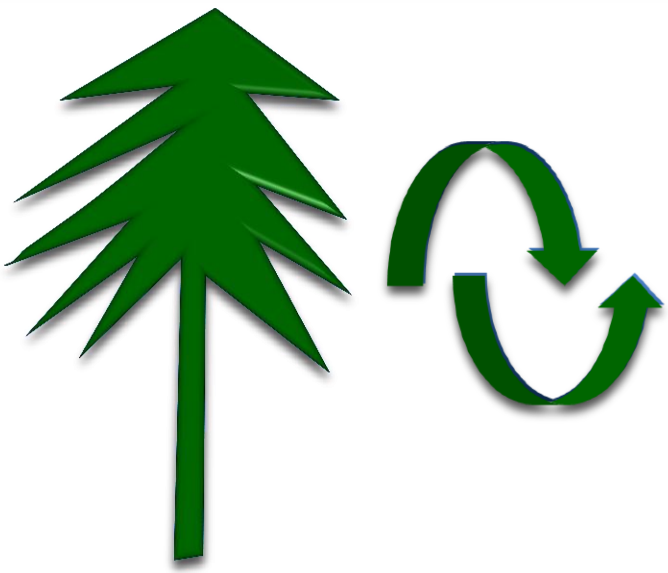
- 72 cities across North America
- Flows assigned to ~60 million streets *every few minutes*

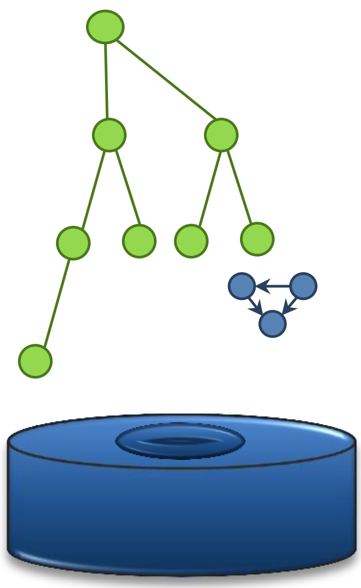
ClearFlow

maps.bing.com * m.bing.com



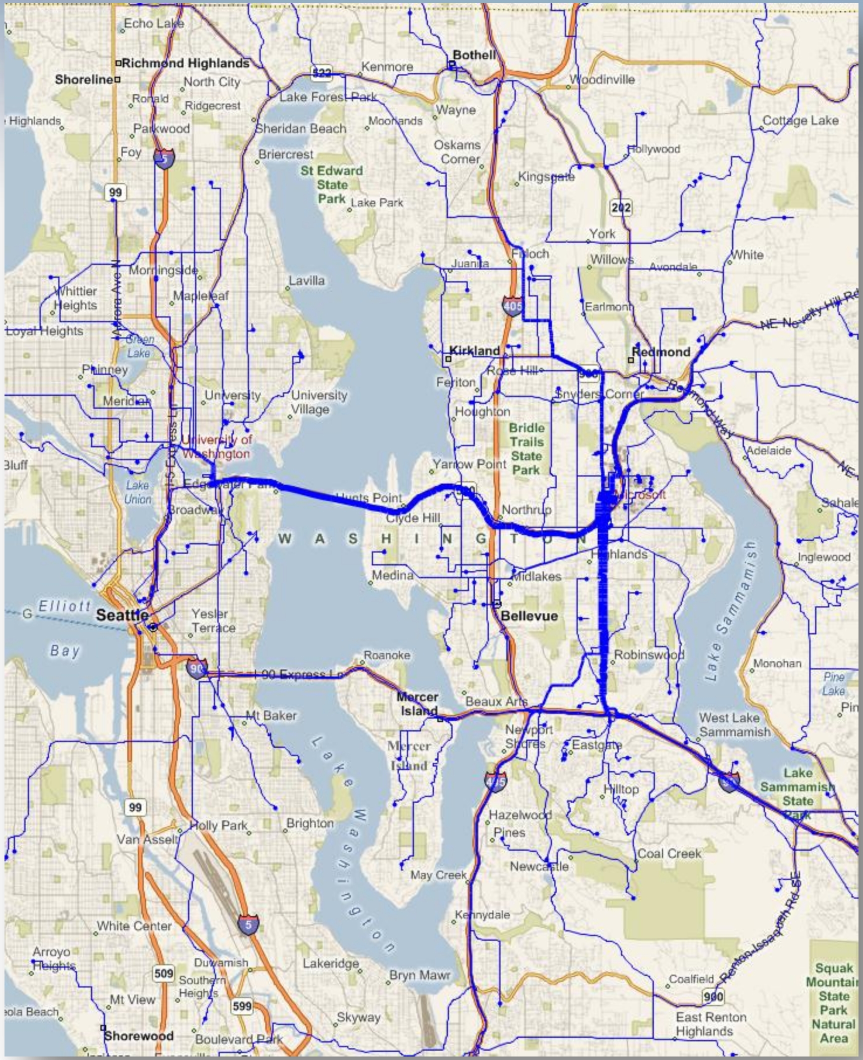
Sustainability



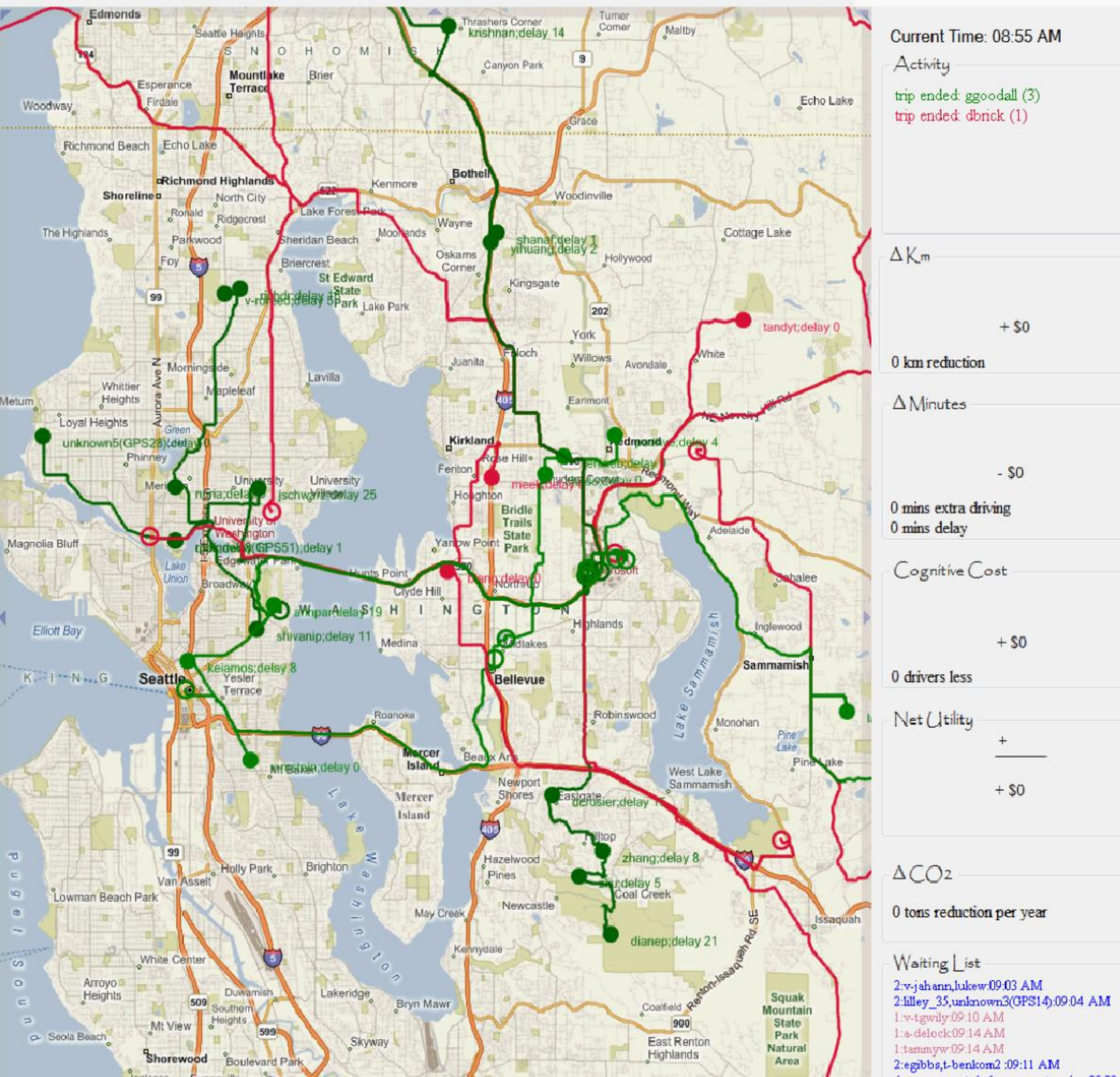


Inferring Commute Trips

- Data: AM/PM commutes to & from Microsoft



[Kamar & Horvitz, 2009](#)



Commute request

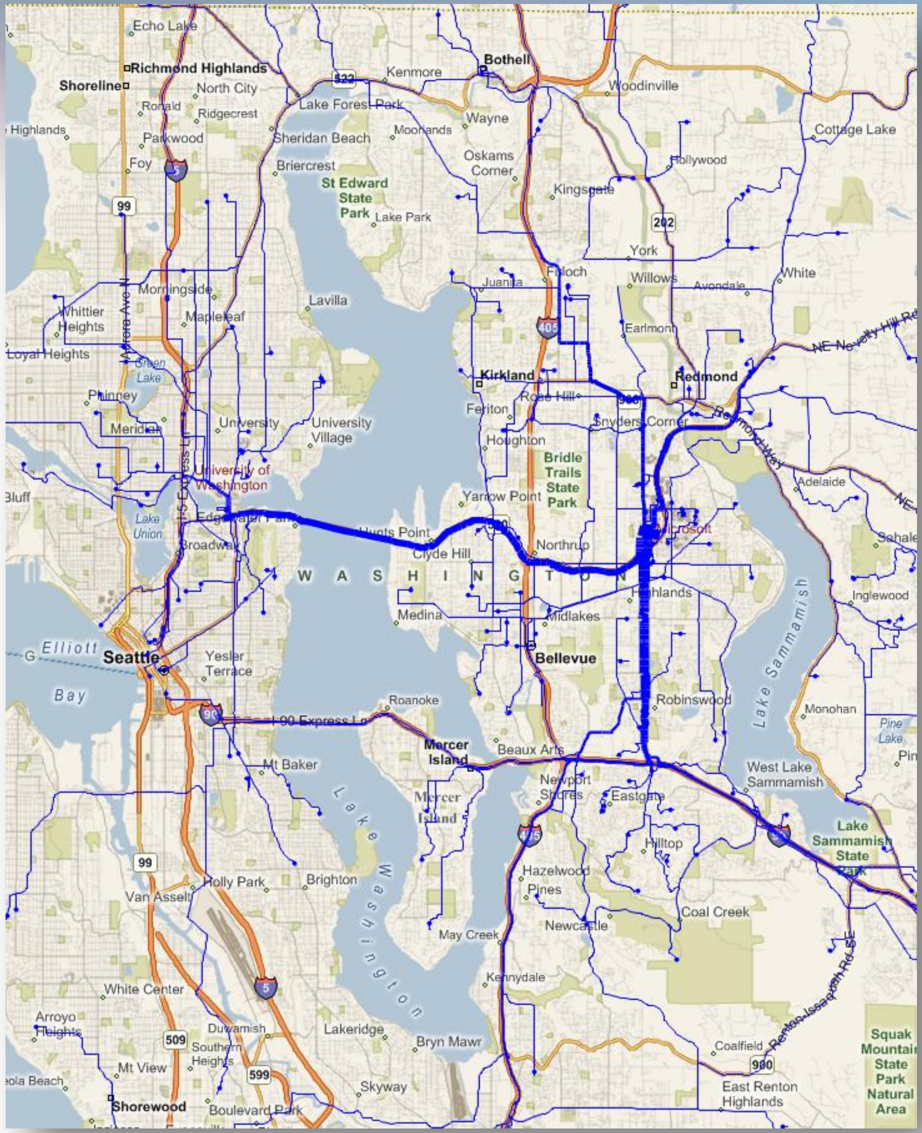
Rideshare queued

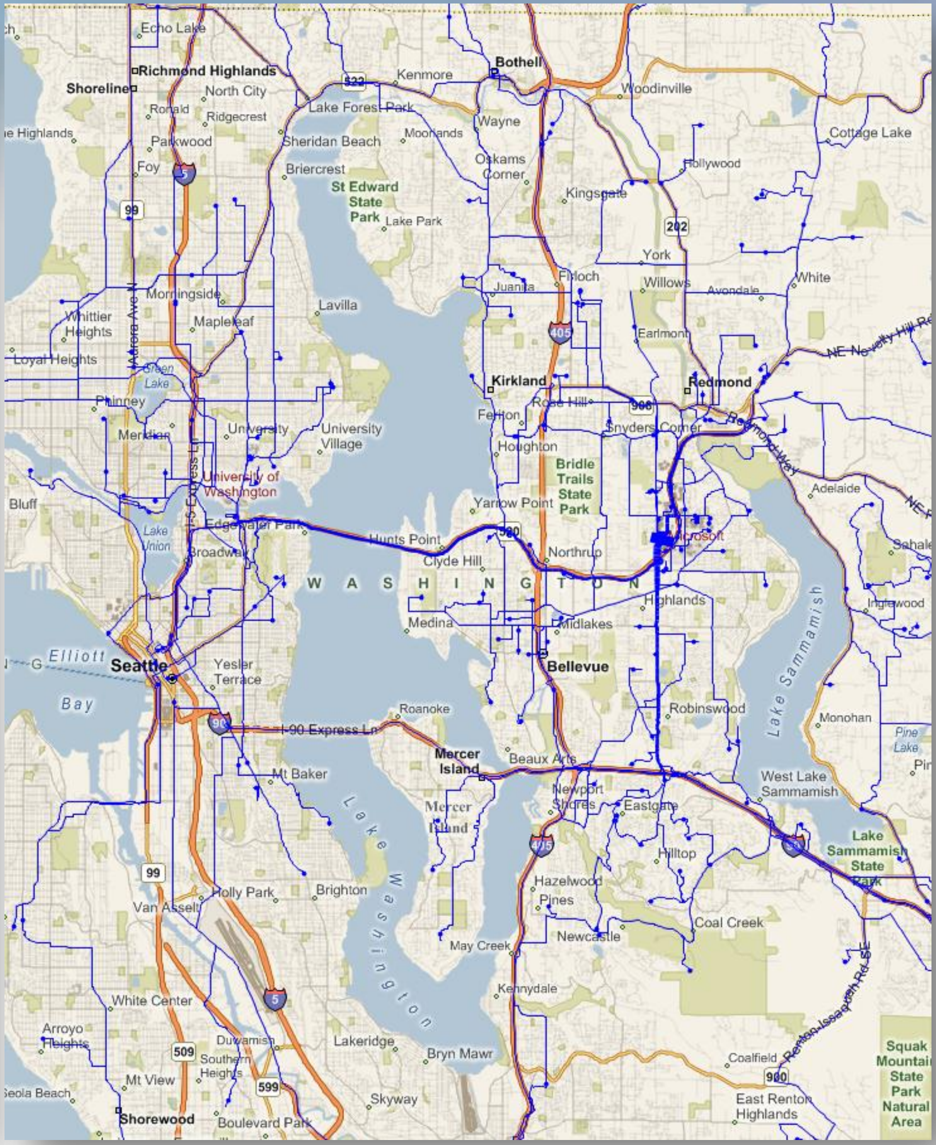
Rideshare starts

Single rider starts

Results

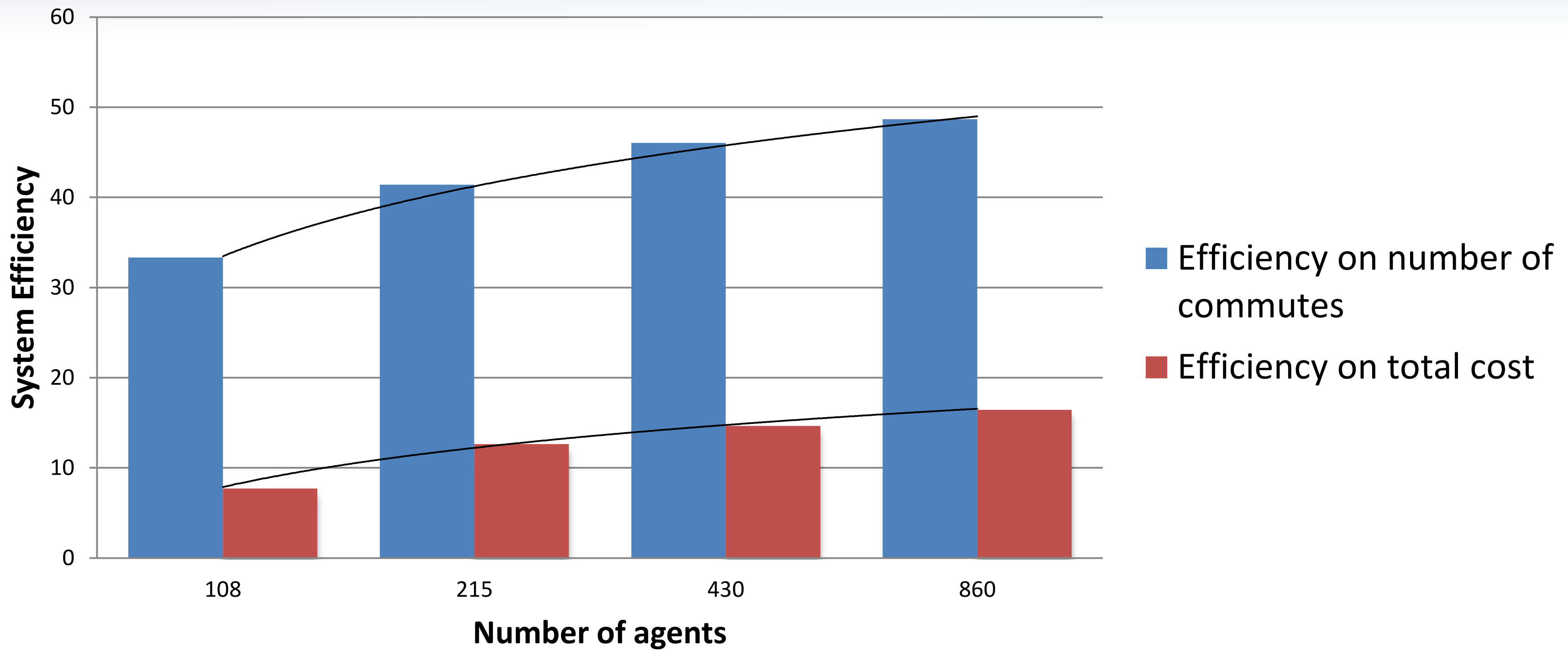
Normal commute → **Computed rideshares**





Computation Models and Insights

"What If?" Studies



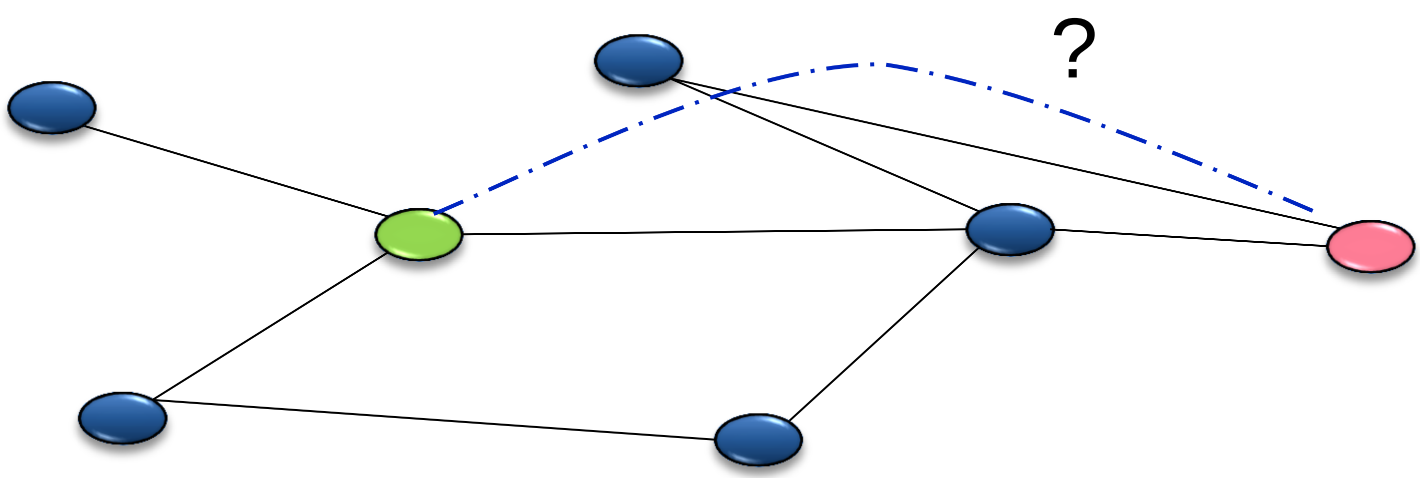
Number of participants →

Kamar & Horvitz, 2009

Acceptance and Trust

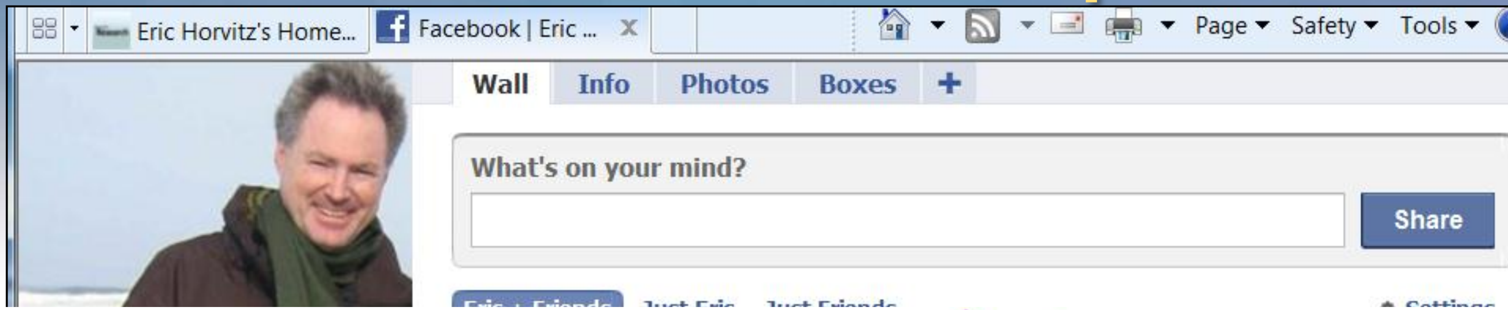
Challenge: Understanding acceptance, perceptions, social considerations

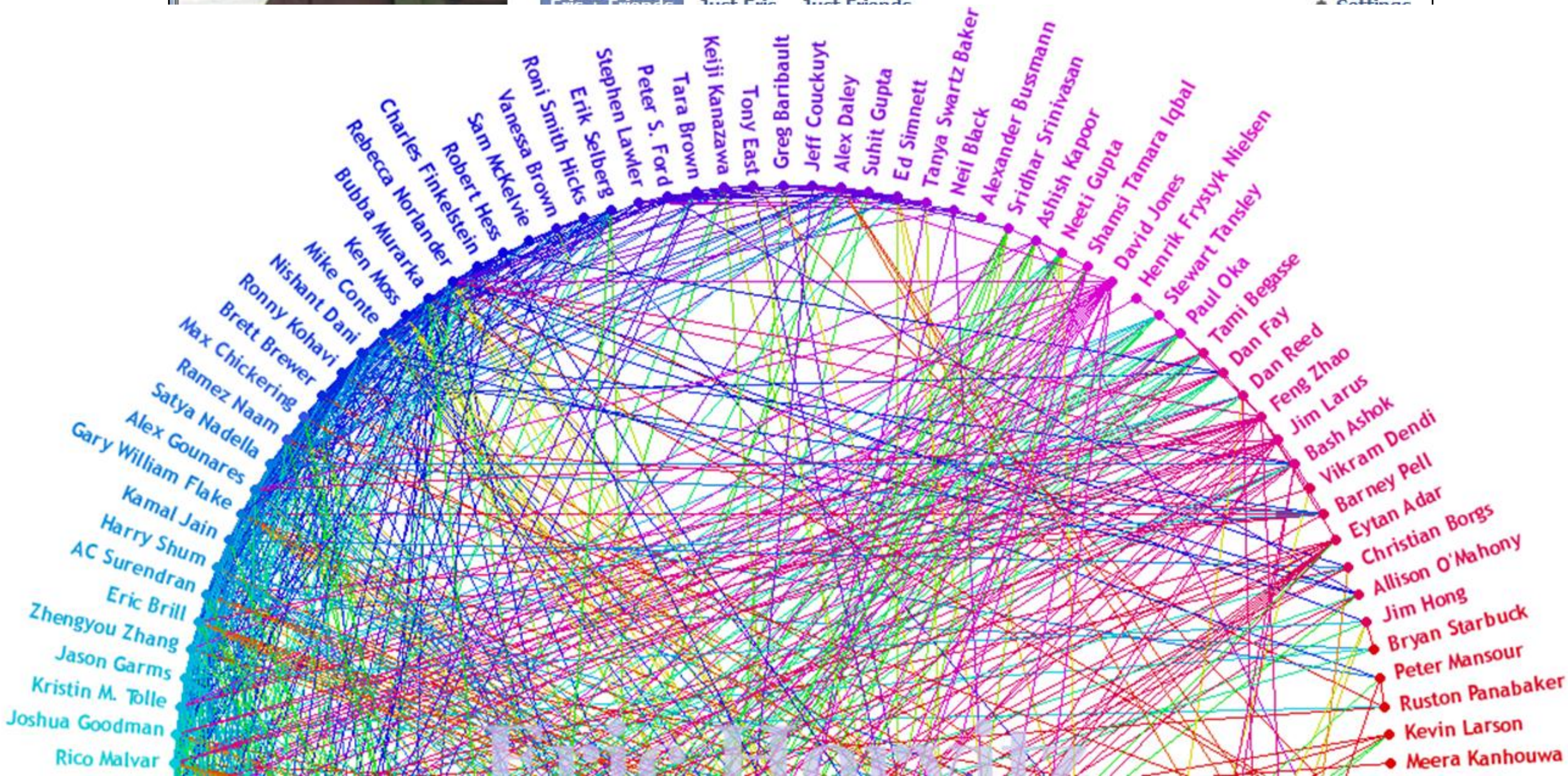
- Address concerns, leverage opportunities
- Trusted organizations
- Referral, reputation
  - e.g., existing online social networks (e.g., link distance bounds)



Kamar & Horvitz, 2009

Distances and Relationships





Privacy and Sensing

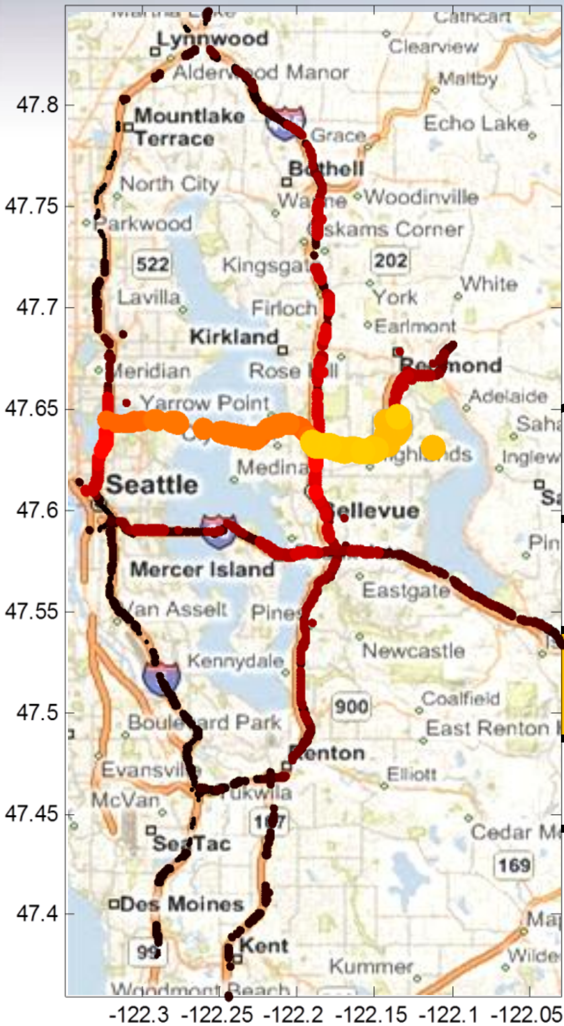
How can we harness data about people while acting in accordance with preferences about privacy?

Privacy Challenges …and Opportunities

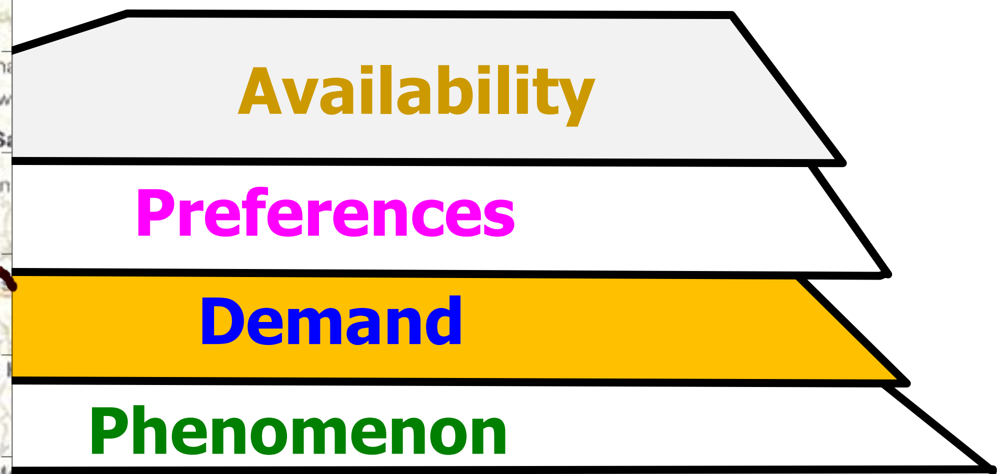
- Sensing guided by preferences & contracts
  - Cost-benefit analysis, *k*-anonymity, differential privacy, sampling with risk
- Protected sensing & personalization
  - Local sensing and execution, use population data when available

Principles of *Community Sensing*

We can share a little bit & have a great system for all.



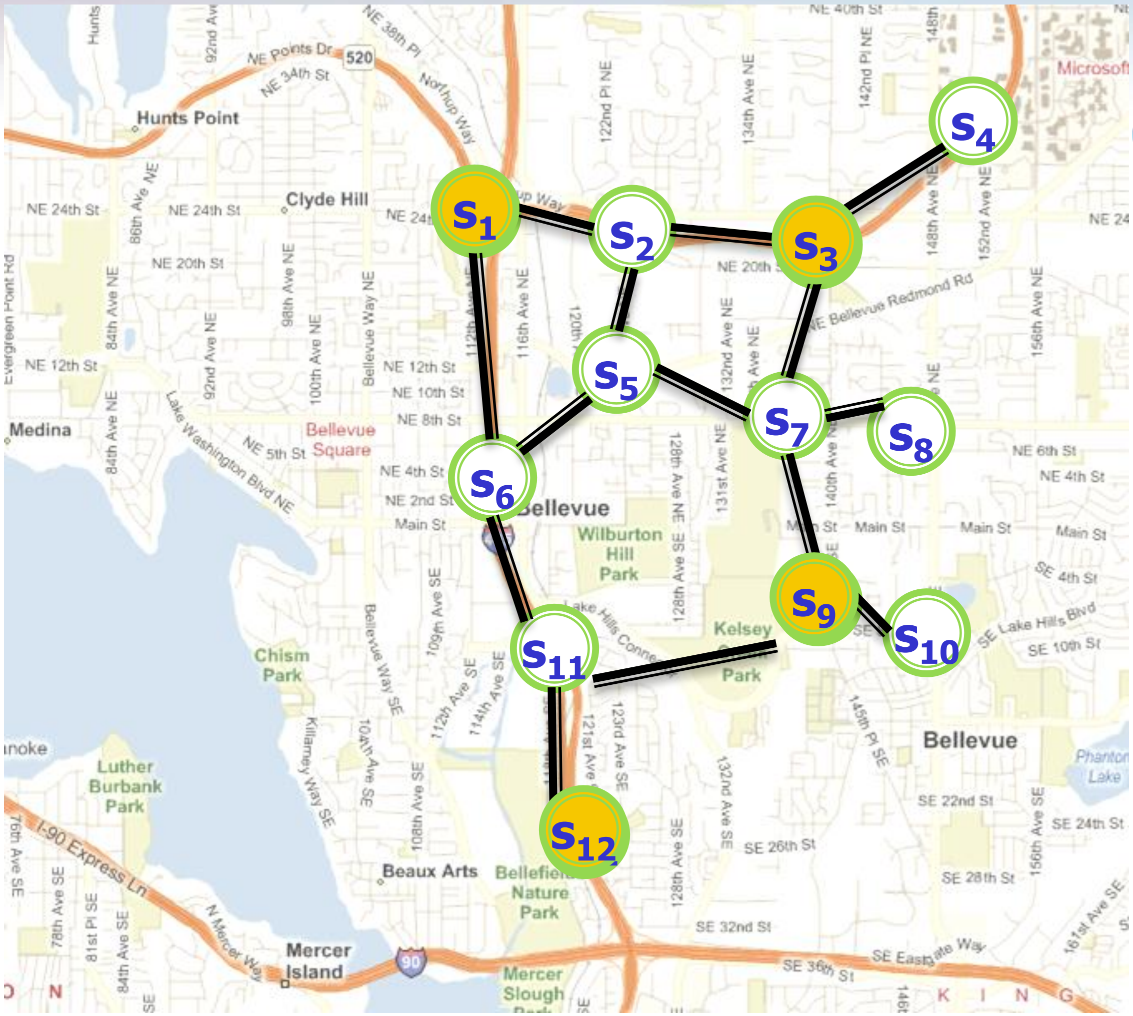
- Road demand + flow uncertainty used to compute *value of sensing*



Krause, Horvitz, Kansal, F. Zhao, 2008

Selective Sensing for Traffic

Compute value of information for selective sensing.



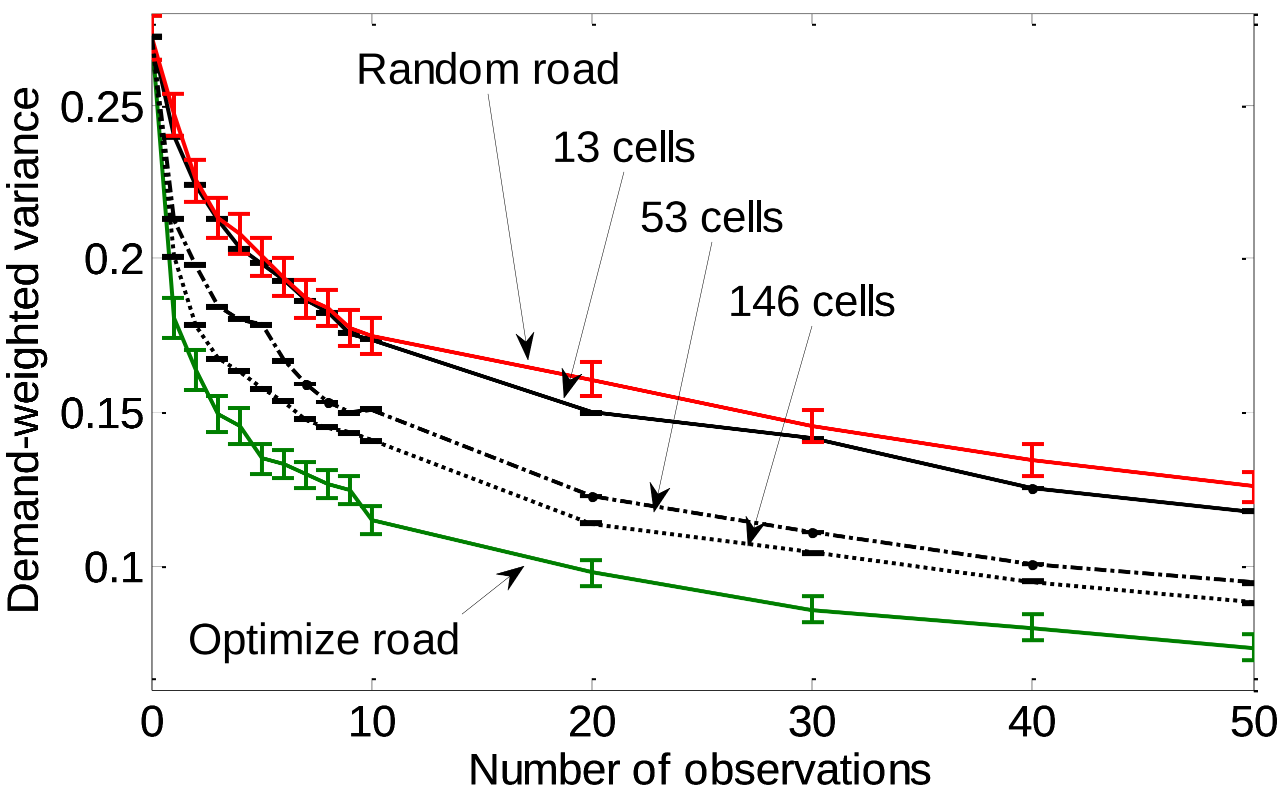
- Predict unmonitored flows from sensed flows

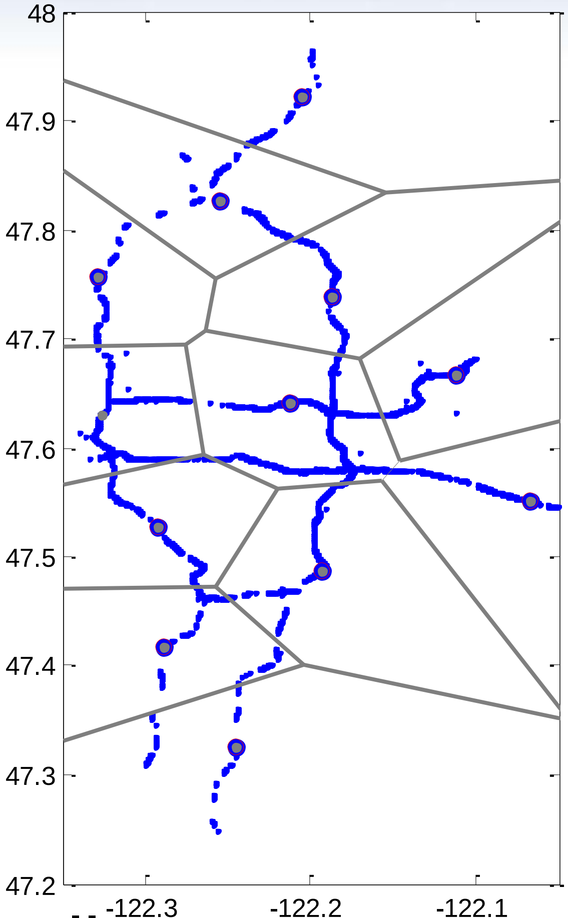
  e.g., $p(S_5 \mid S_1, S_9)$

[Krause, Horvitz, Kansal, F. Zhao, 2008](#)

Power of Selective Sensing

Sensing under *privacy budget* can perform well.





- Low selection noise even with 53 cells

Tradeoffs: Balancing Utility & Sensitivity

- Balance and Preferences
  - Value of information
  - Sensitivity of private information
- Identifying a sweet spot?

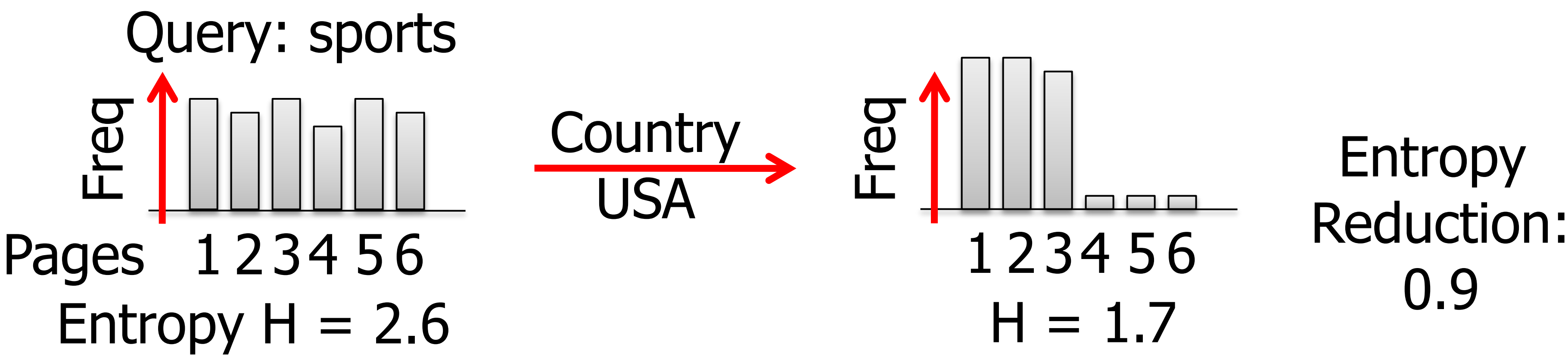
Maximize value at minimum sensitivity

Krause & Horvitz, 2010

Example: Web search

Value of personal data in enhanced service:

- **Demographic data**
- **Search history** (same query, searches per day?)
- **Topic interests** (ever visited business, kids, etc. website)
- **User behavior** (location, ever changed zip)?



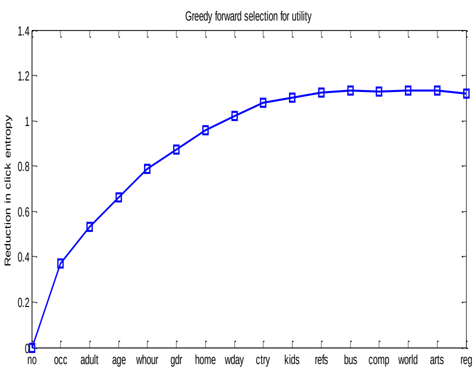
Krause & Horvitz, 2010

Trading off value & cost

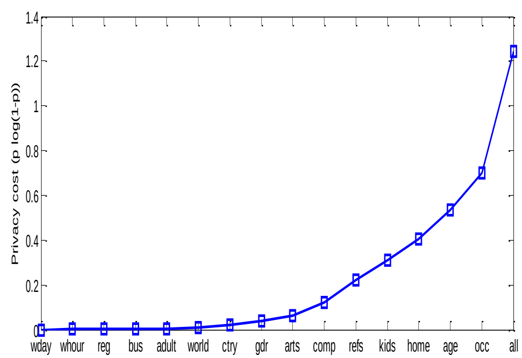
Value:
Diminishing returns
(submodular)

Cost:
Accelerating
(supermodular)

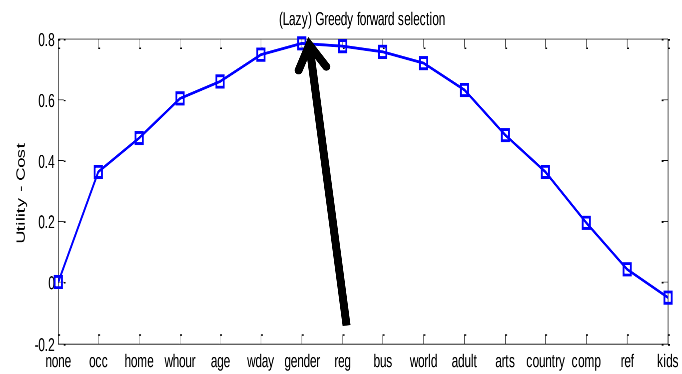
Optimization



More observations



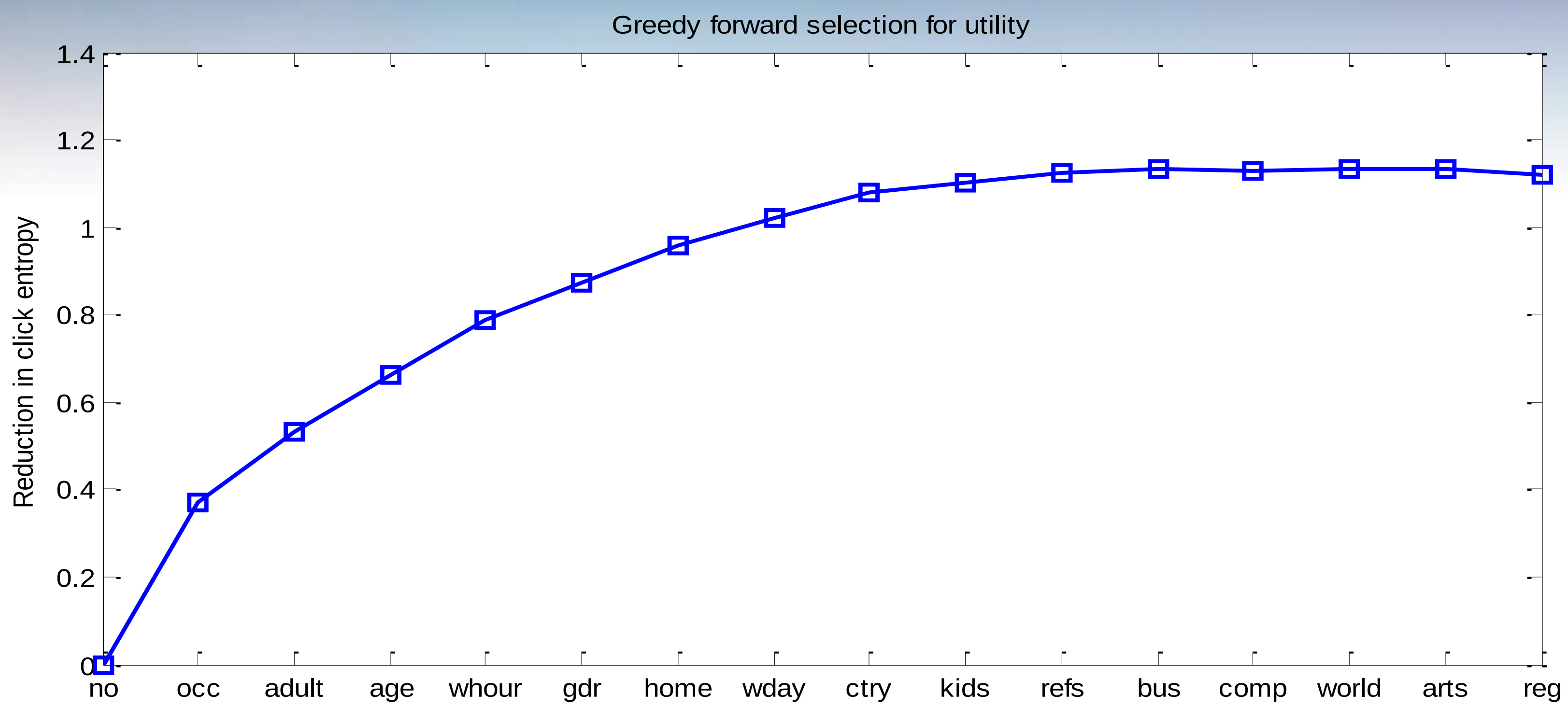
More observations



More observations

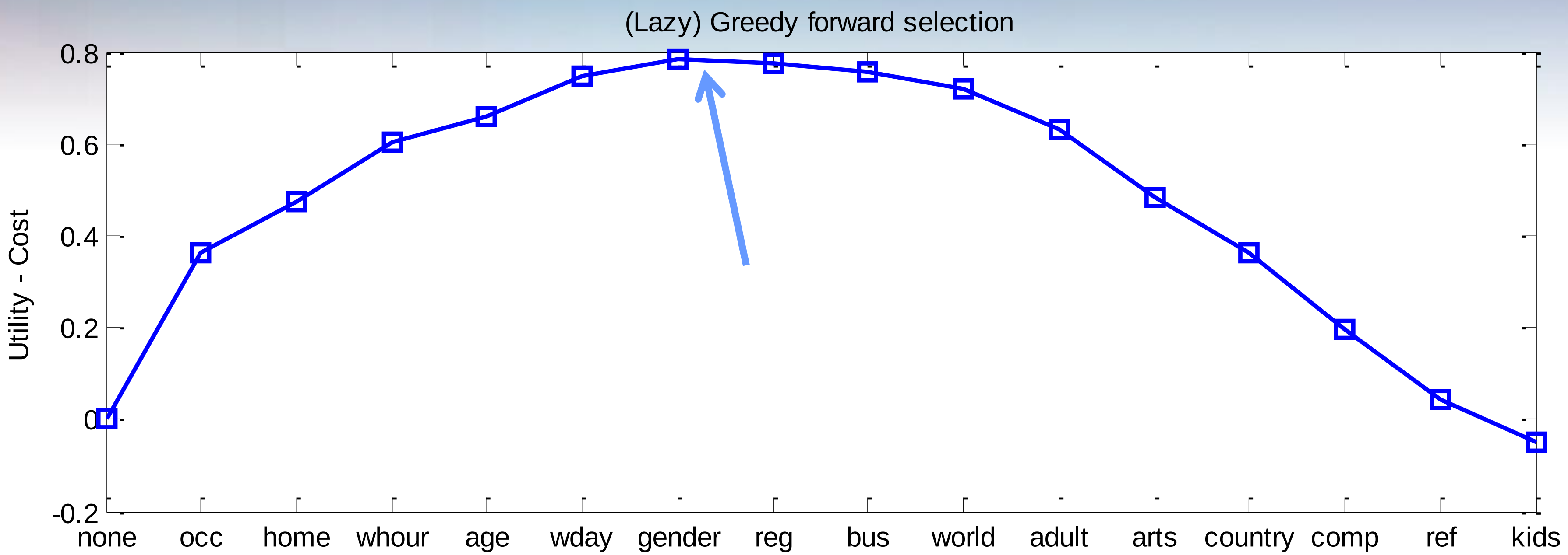
Krause & Horvitz, 2010

Diminishing Returns on Utility



Krause & Horvitz, 2010

Finding the Sweet Spot



Krause & Horvitz, 2010