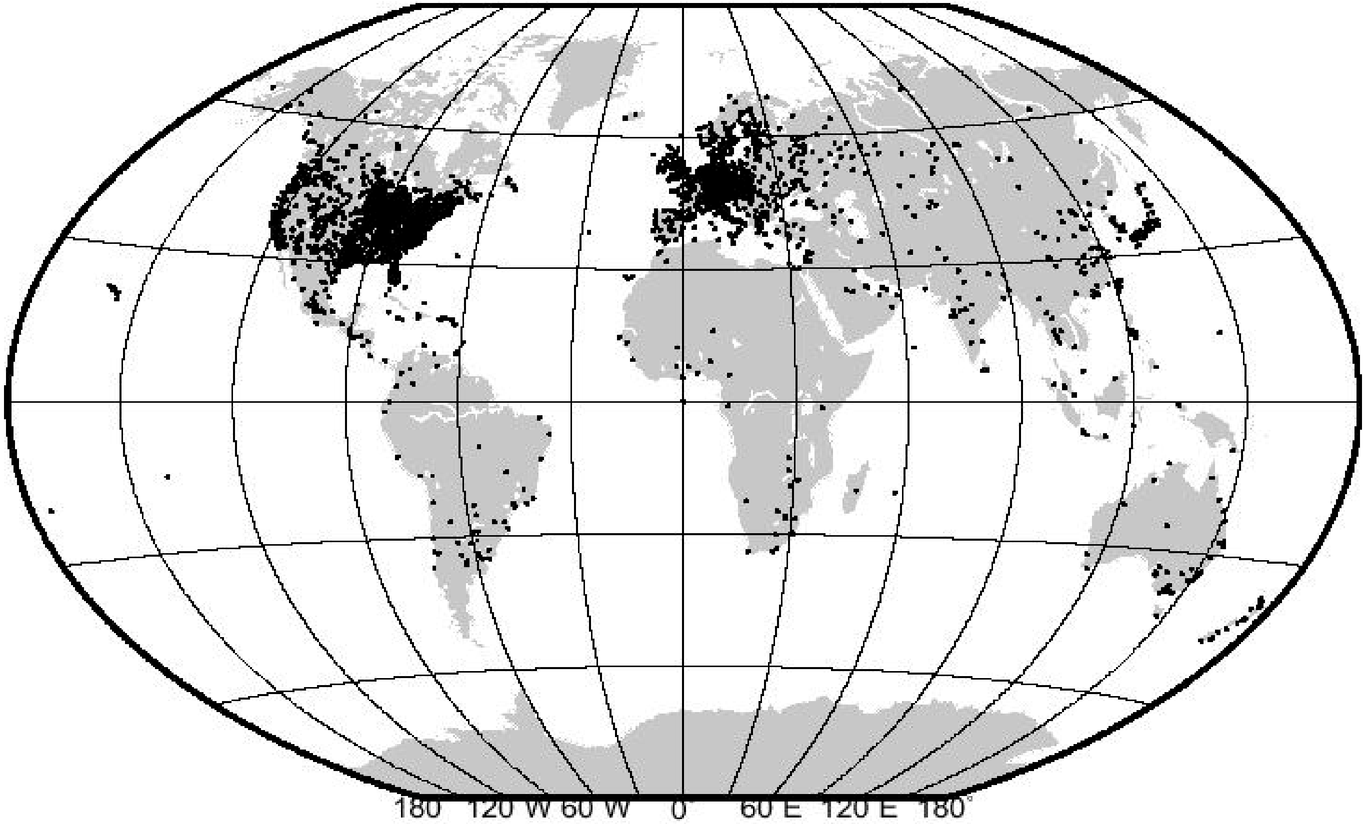
On the Geographic Location of Internet Resources

Anukool Lakhina

with **John Byers, Mark Crovella and Ibrahim Matta**
Department of Computer Science
Boston University

Computer Science

The question we ask:



Where on earth is the Internet?

Why does this matter?

- **Generating representative network topologies**
  - Many simulation based results in networking depend critically on network topology
  - Topology generation is still fairly ad hoc: generators discount geometry and focus solely on connectivity
- **Worthwhile scientific goals**
  - Basic investigations pay off in unexpected ways

Our Approach

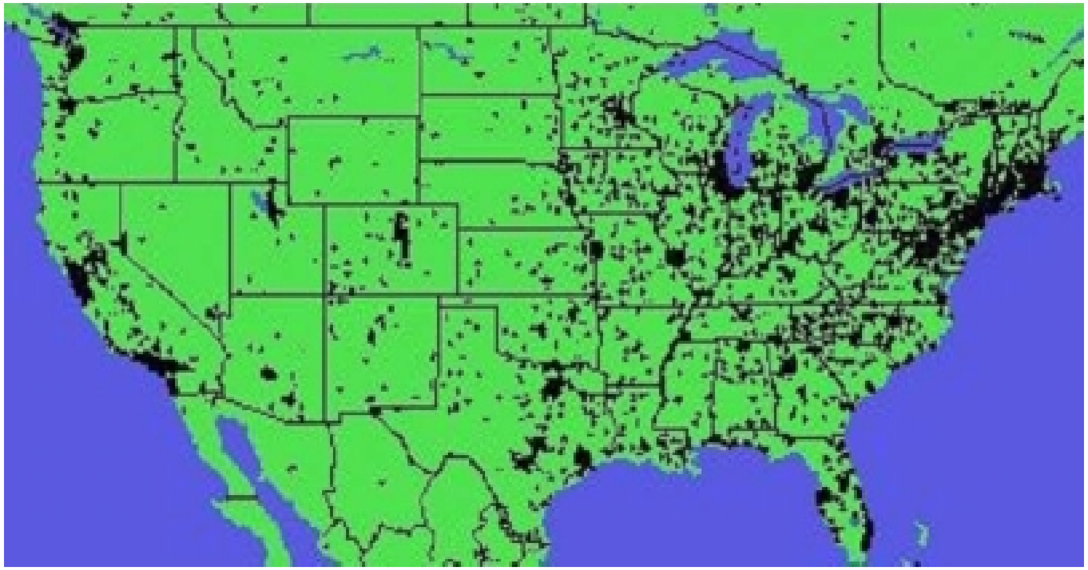
- **Treat Internet as undirected graph embedded on Earth's surface**
  - Nodes are routers or interfaces (not end-hosts)
  - Edges are adjacent IPs as seen by `traceroute` (ignore lower level details)
- **We then examine all combinations of**
  - 2 large router datasets (*Mercator and Skitter*),
  - 2 different geographic mapping tools (Ixia's *IxMapper* and Akamai's *EdgeScape*); both rely on hostname based mapping techniques and external information sources.
  - 3 geographic regions (*USA, Europe, Japan*)

Two sets of questions

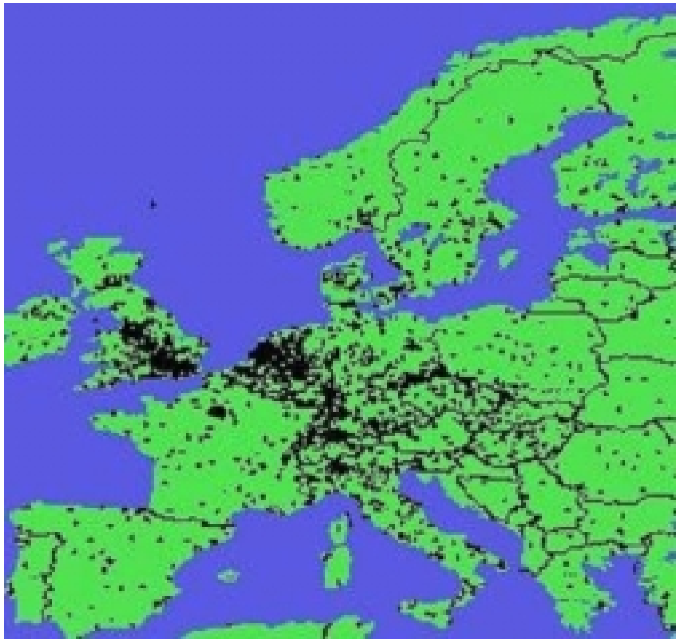
1. Where are the routers?
   What is the relationship between population and location of routers?

2. How long are the links?
   What effect does distance play on link creation?

Where are the routers?



USA

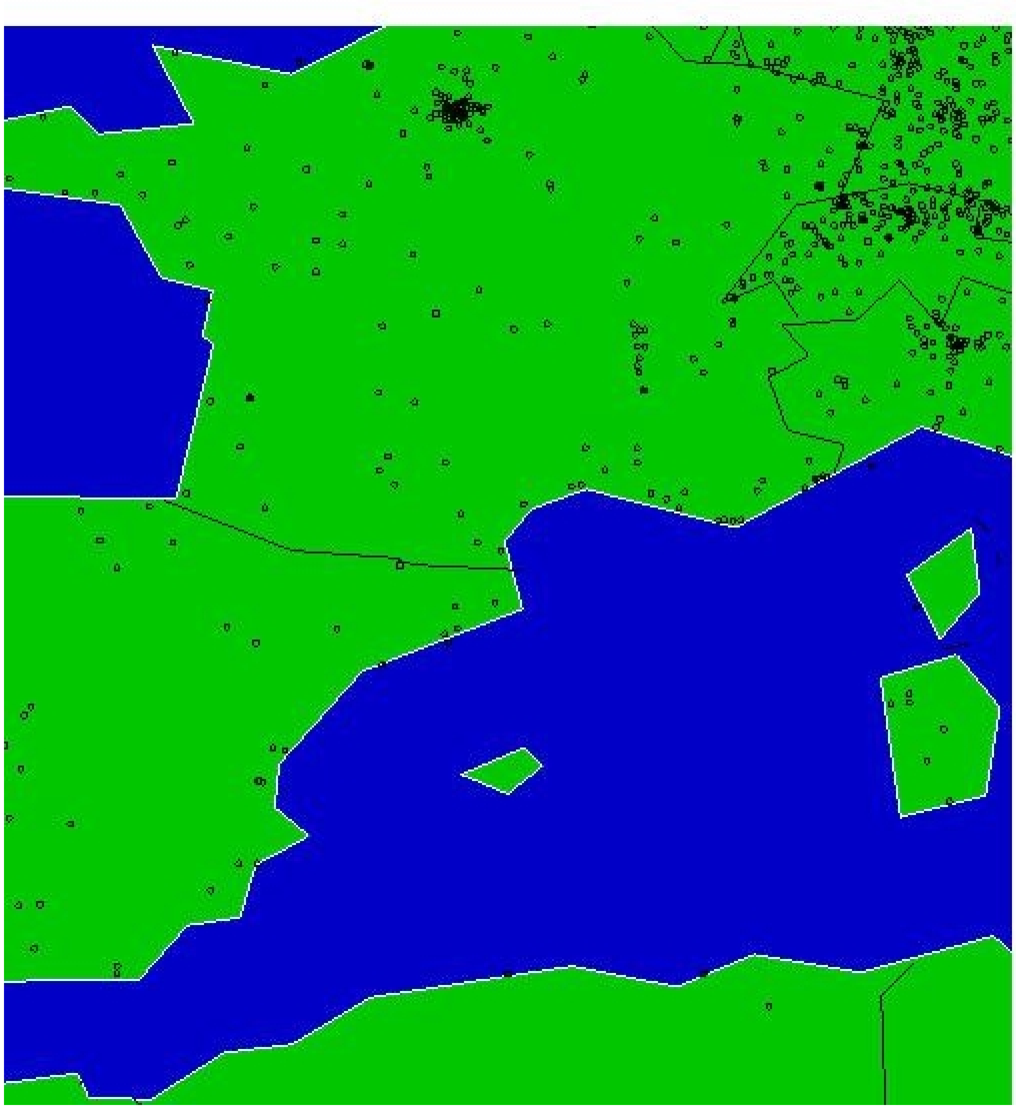


Europe

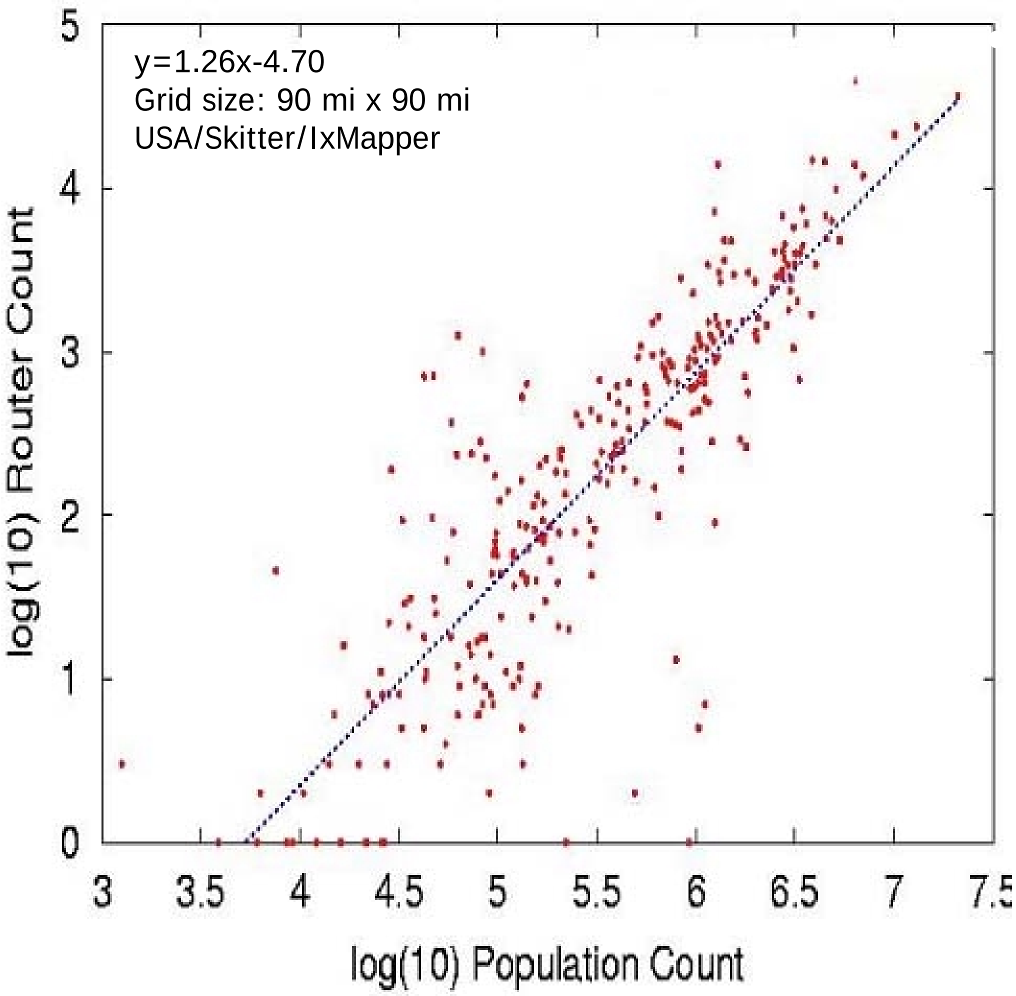
Spatial Distribution of Routers is Fractal, not Uniform.

(Yook et al, PNAS 2002)

South of France and neighboring areas



Routers and People



Router density shows a superlinear relationship to population density:

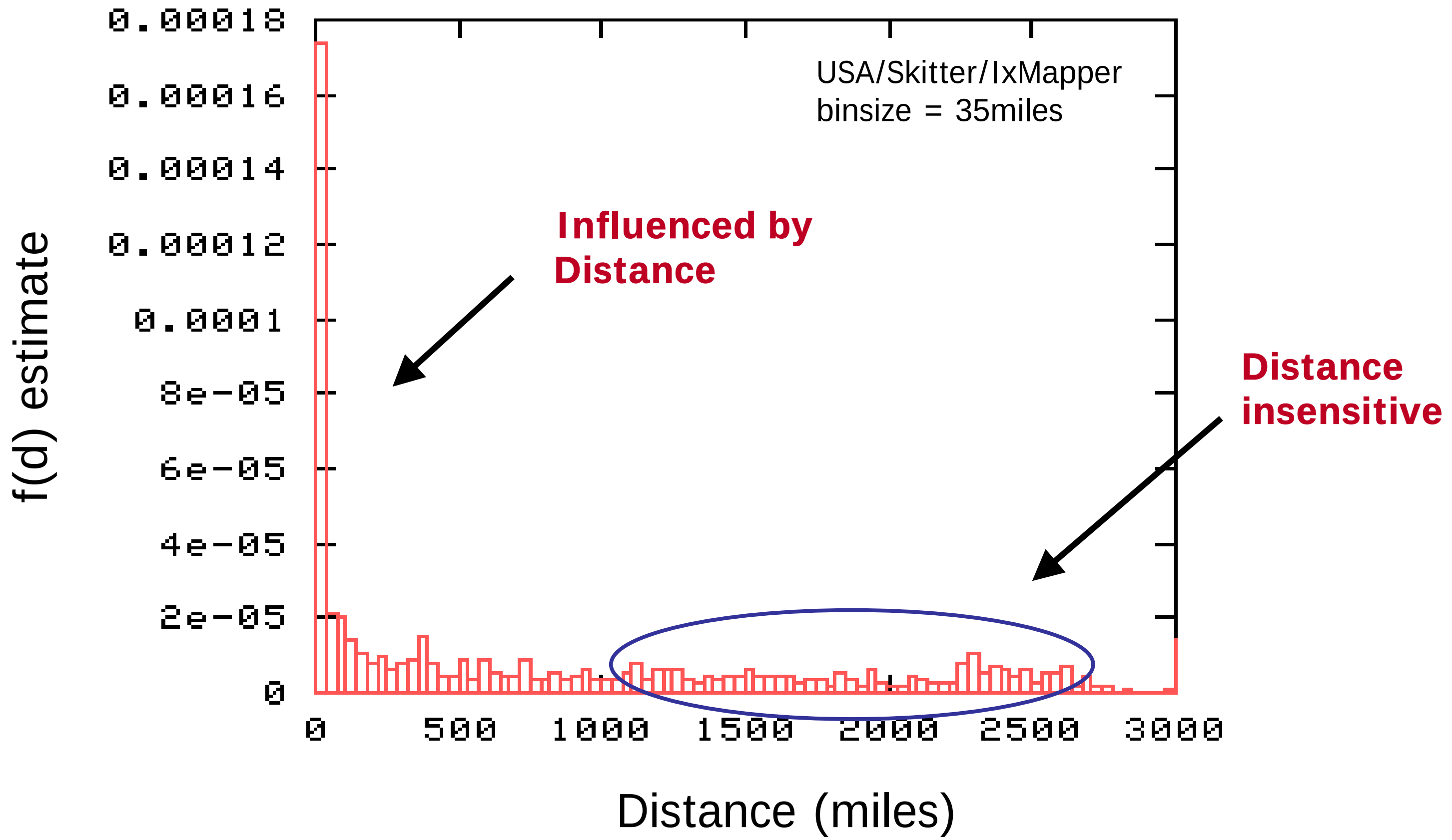
$$R \approx k P^{\alpha}$$

- k varies with economic development of region
- $\alpha > 1$: More routers per person in densely populated areas

Links: Does Distance Matter?

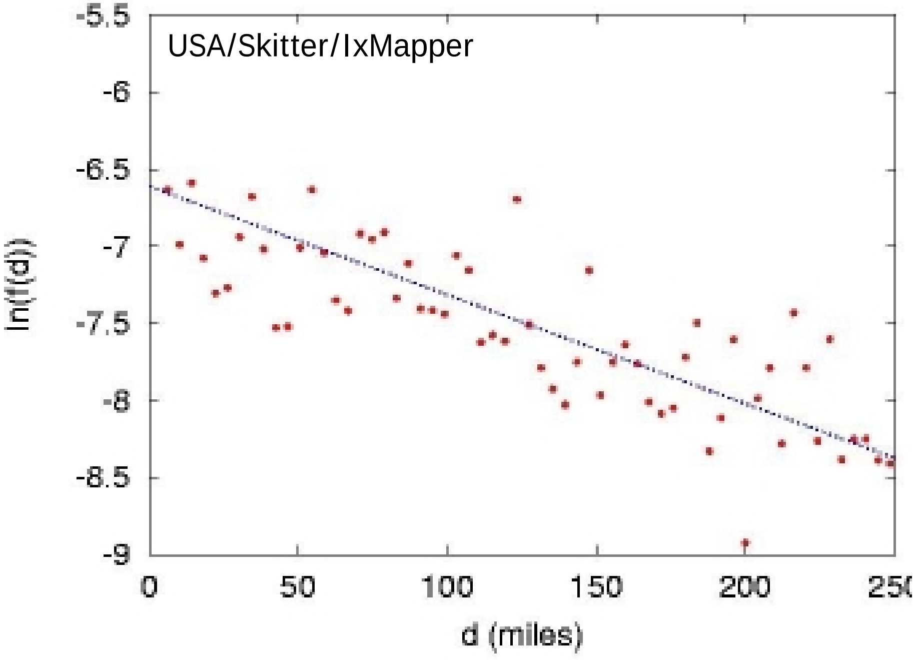
Distance Preference Function, f(d)

- Probability two nodes are connected given they are d miles apart, $f(d) = P(C \mid d) = \frac{\text{\# links with length d}}{\text{\# node pairs with length d}}$

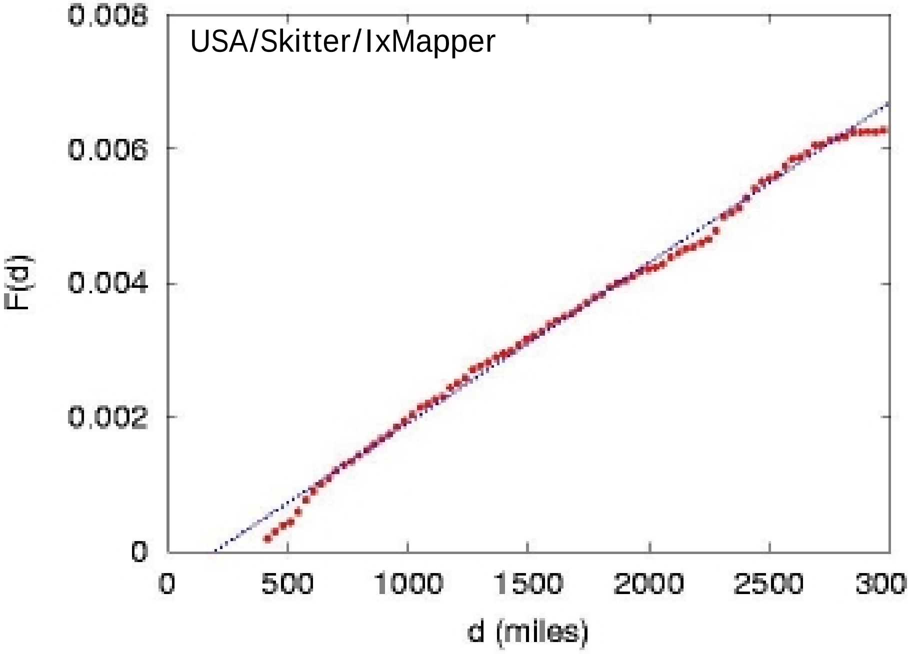


Two Classes of Links

Short Links: Distance Matters
(~ 80-90%)



Long Links: Distance Insensitive
(~10-20%)



- Linear fit on log-linear axis implies exponential preference function
- Driven by connectivity costs?

- Linear fit for cumulative f(d) implies uniform preference function
- Play a structural role?

Final Remarks

Towards geographically-inspired network topology models

- Irregular location of routers should not be discounted in models
- From link distances to link latencies…
- A wider set of bases for topology construction

More results can be found in our technical report,
http://www.cs.bu.edu/techreports/

Thanks!

- **CAIDA:**
  - David Moore, Andre Broido and k claffy
- **USC/ICSI:**
  - Ramesh Govindan and Hongsuda Tangmunarunkit
- **Akamai:**
  - Avi Freedman and Bruce Maggs
- **University of Notre Dame:**
  - Haewoong Jeong and Lazslo Barábasi