

# Analyzing the spatial structure of the Internet topology

Sándor Laki

*ETOMIC TEAM*

*Eötvös Loránd University*

*Budapest, Hungary*

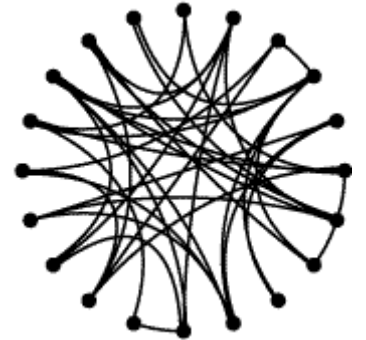
*[laki@etomic.org](mailto:laki@etomic.org)*



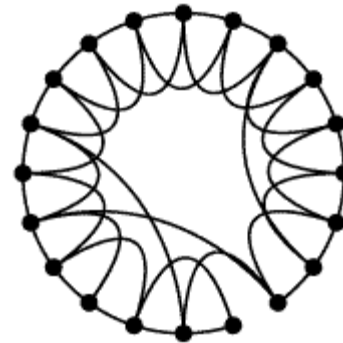
*ISMA 2013 AIMS-5 - Workshop on Active Internet Measurements,  
6-8 February, 2013, San Diego, CA, USA*

# Network research

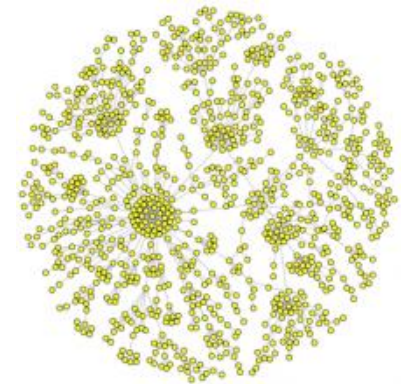
1959 ● Erdős and Rényi



1998 ● Watts and Strogatz



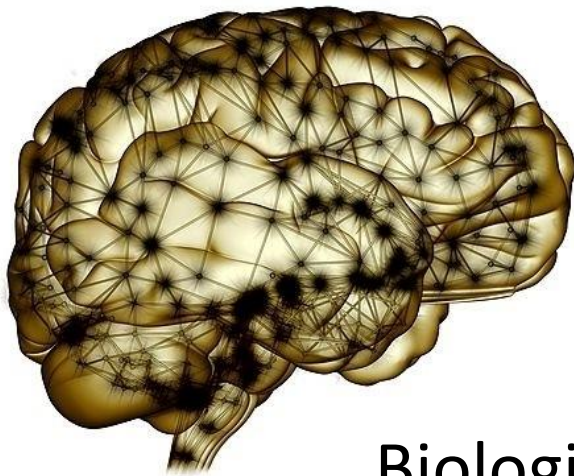
1999 ● Barabási and Albert





Transport

Social



Biological



Internet

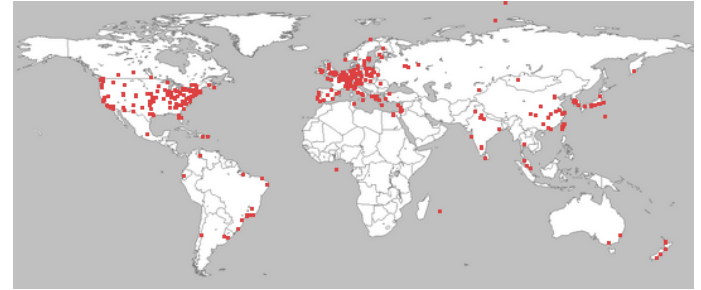
The distance is what really counts.

# What can we say about the spatial structure of the Internet?

P. Mátray, P. Haga, S. Laki, I. Csabai, G. Vattay  
***On the Spatial Properties of Internet Routes***  
*Elsevier Computer Networks, Volume 56, Issue 9 (2012)*

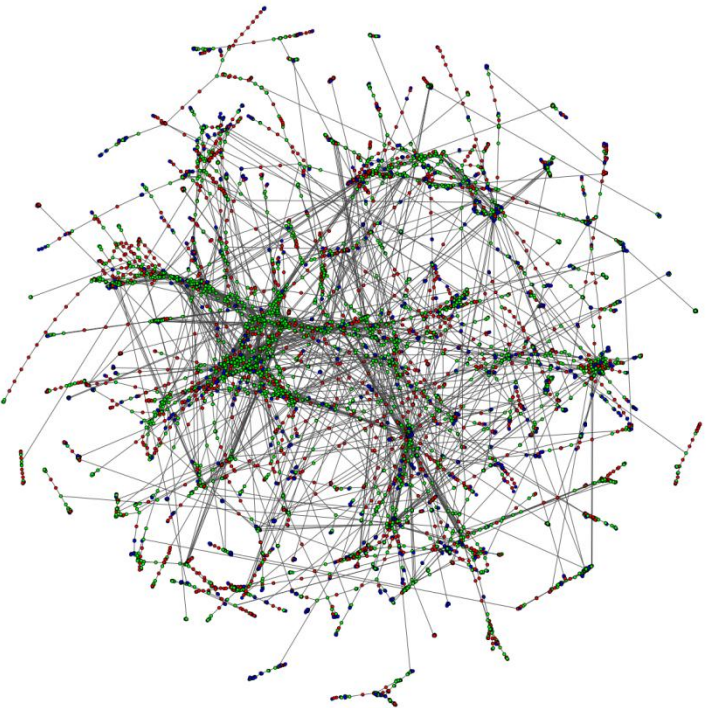
# Data collection

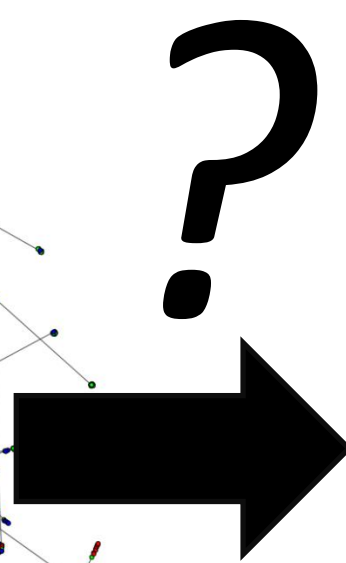
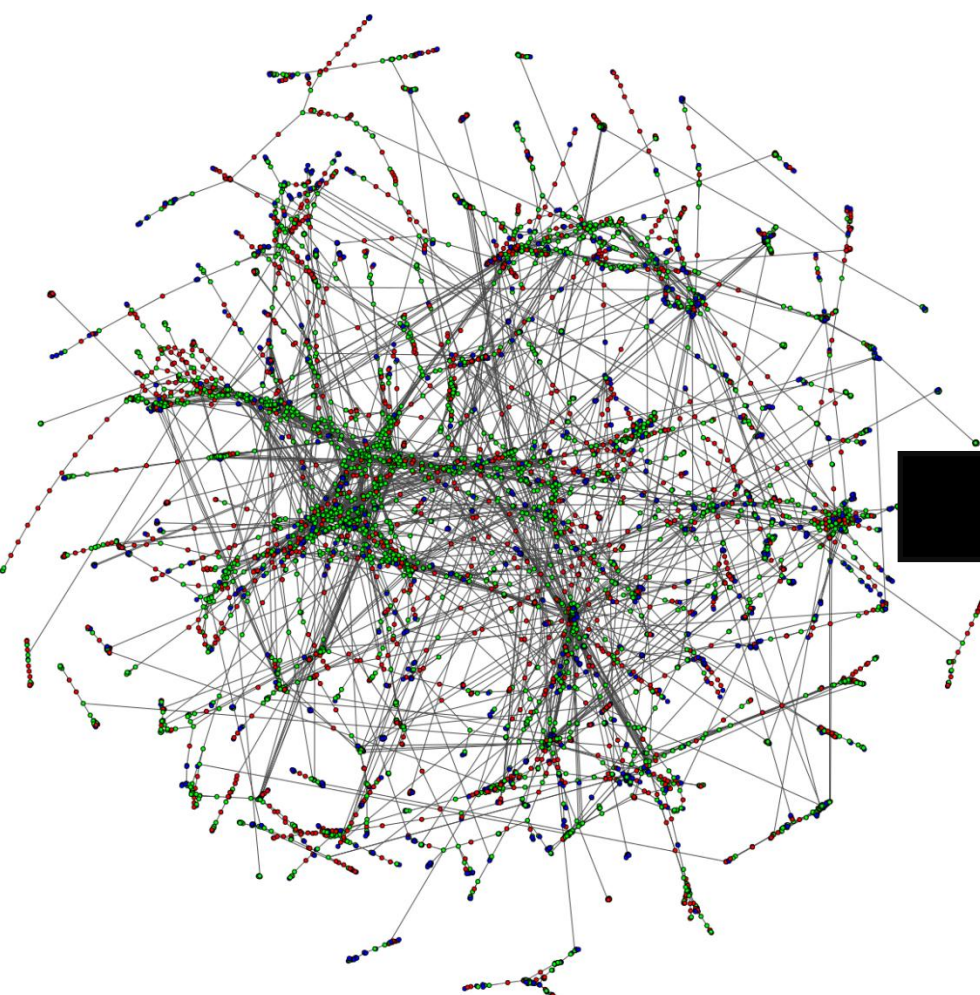
- 700 PlanetLab nodes



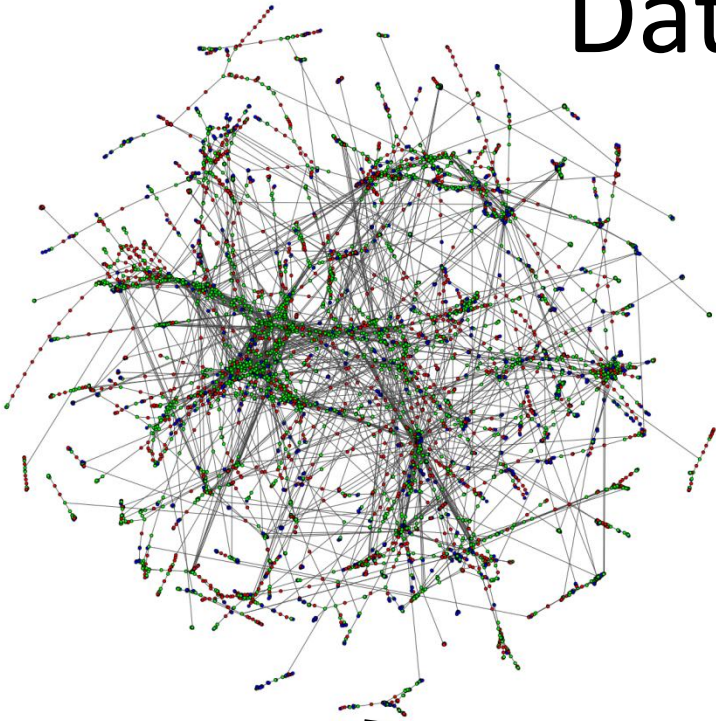
- 400,000 traceroutes

- 16,000 unique IP addresses





# Data collection



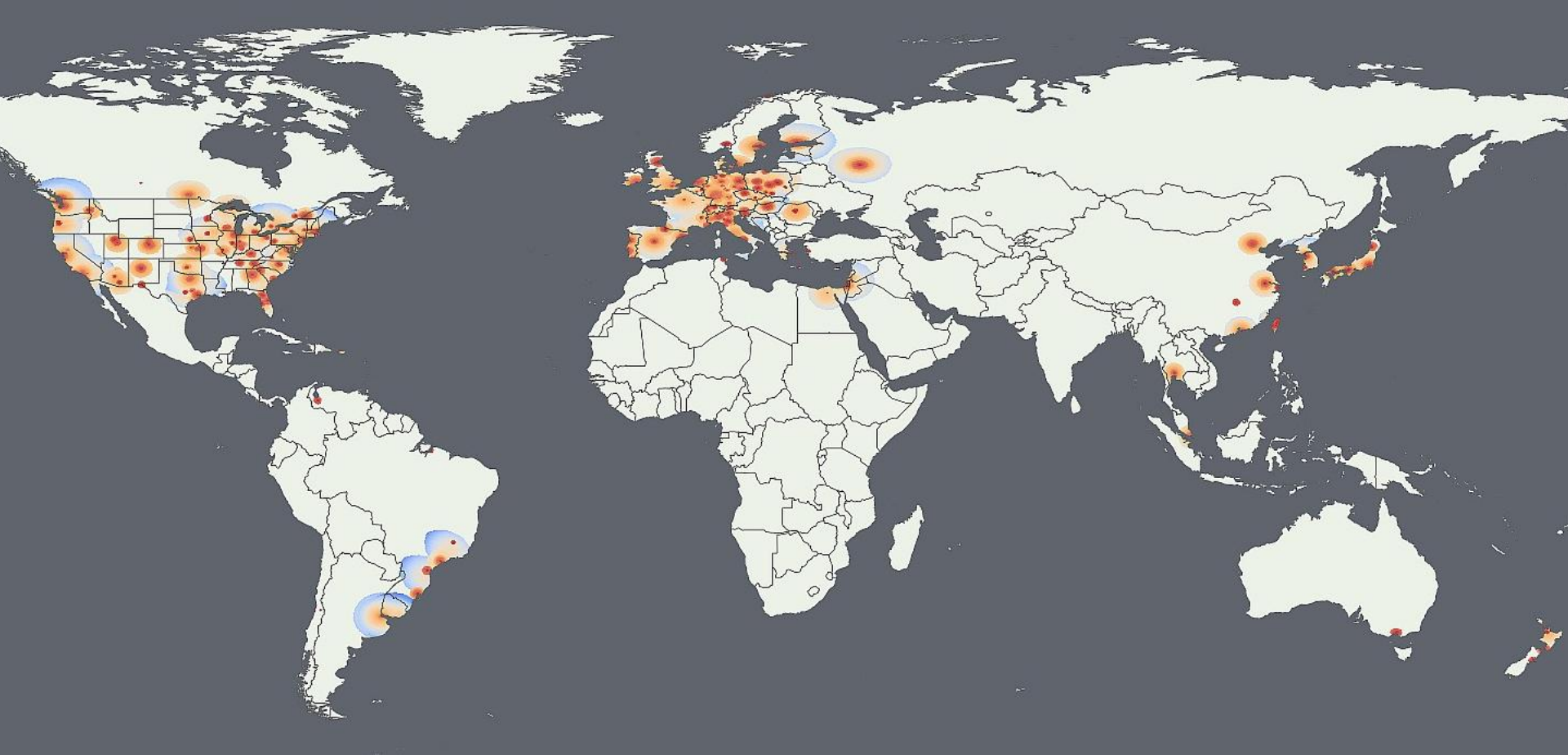
13,000 filtered addresses  
44,000 links

Spotter<sup>5</sup>

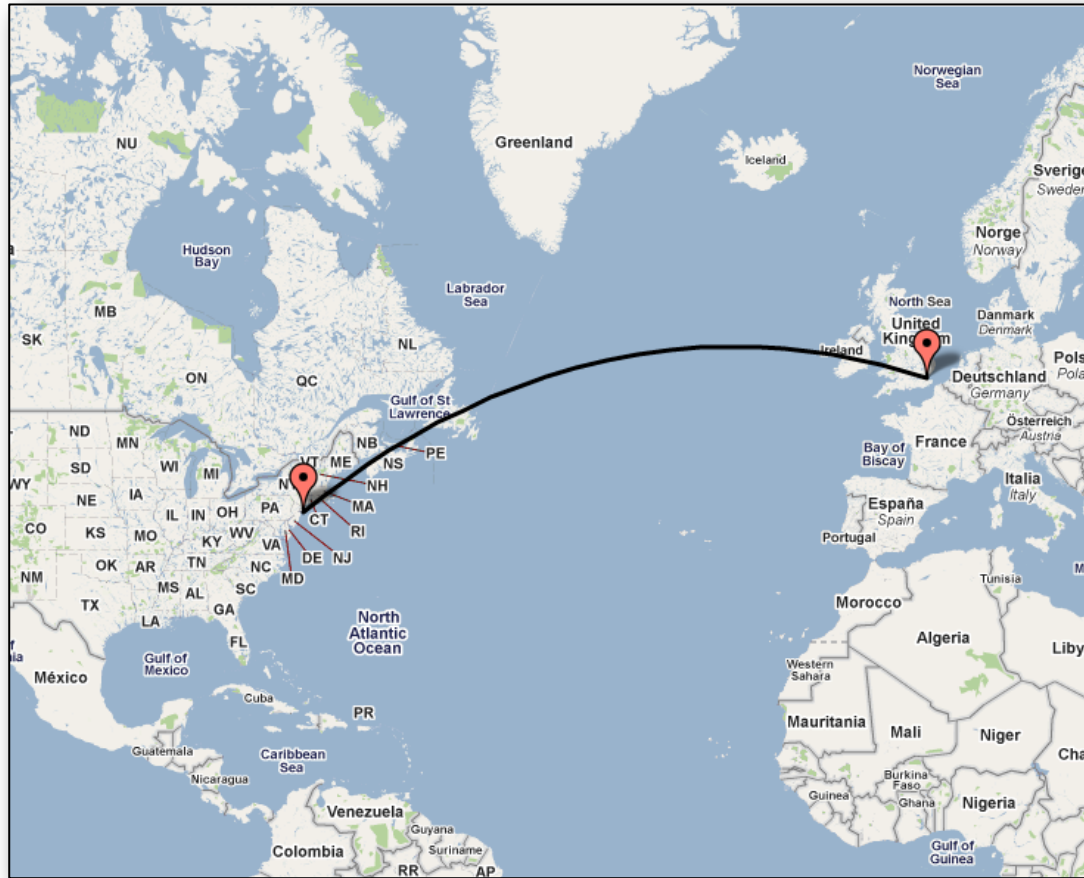
S. Laki et al.: **Spotter: A Model Based Active Geolocation Service**, *IEEE INFOCOM 2011, April 2011, Shanghai, China*



# Router Likelihood map



# Characterizing the link length

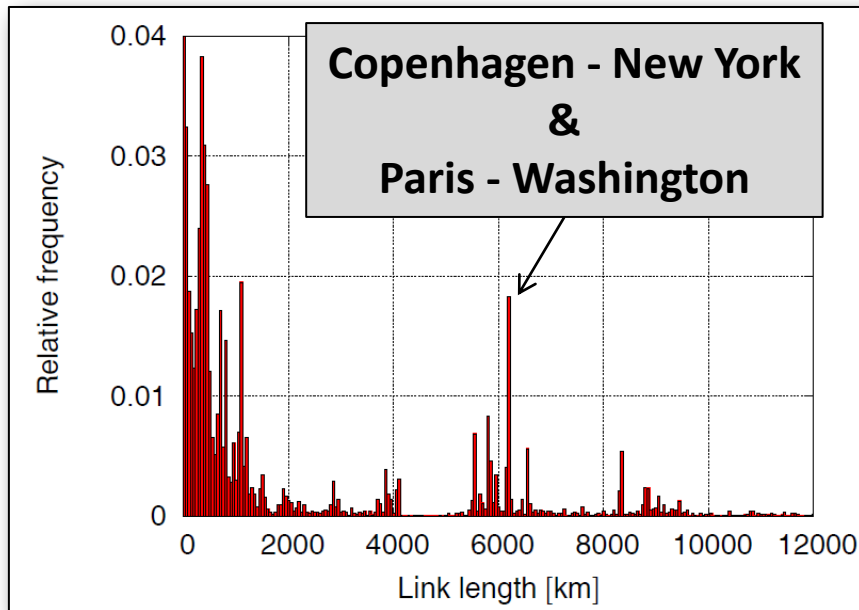


**Link length** is approximated by the spherical distance between the two routers

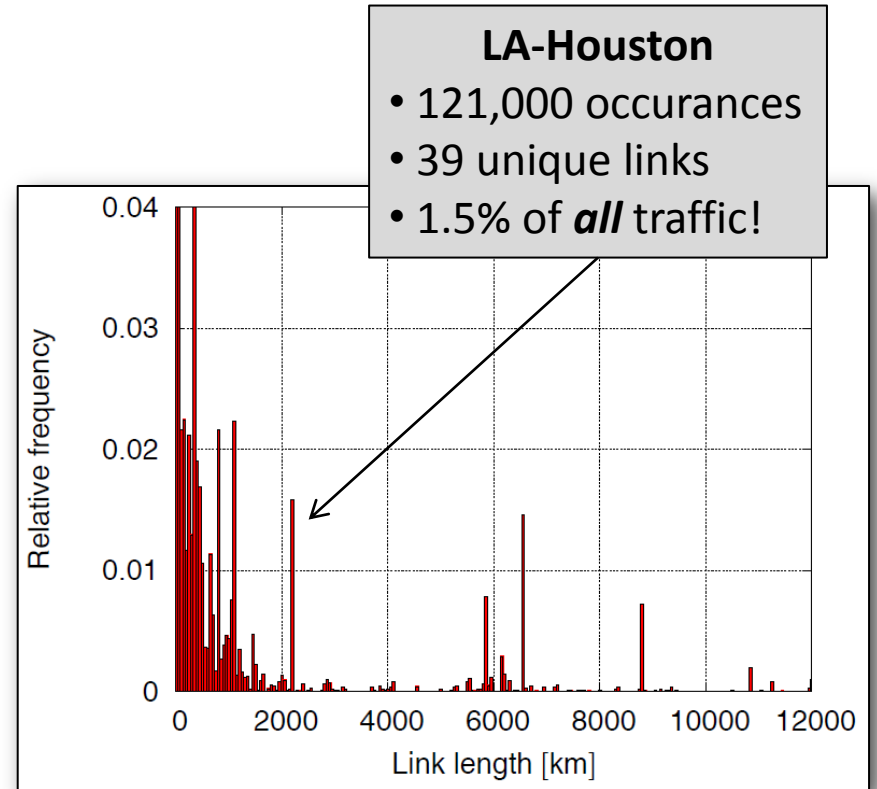
# Characterizing the network links

- Which links are important?
- Which cities are the most interconnected?
- Which link length is the most frequent?
- How to model link length distribution?
- What can be said about the spatial structure of the network?
- etc.

# Which links are frequent or important?



each link is represented once

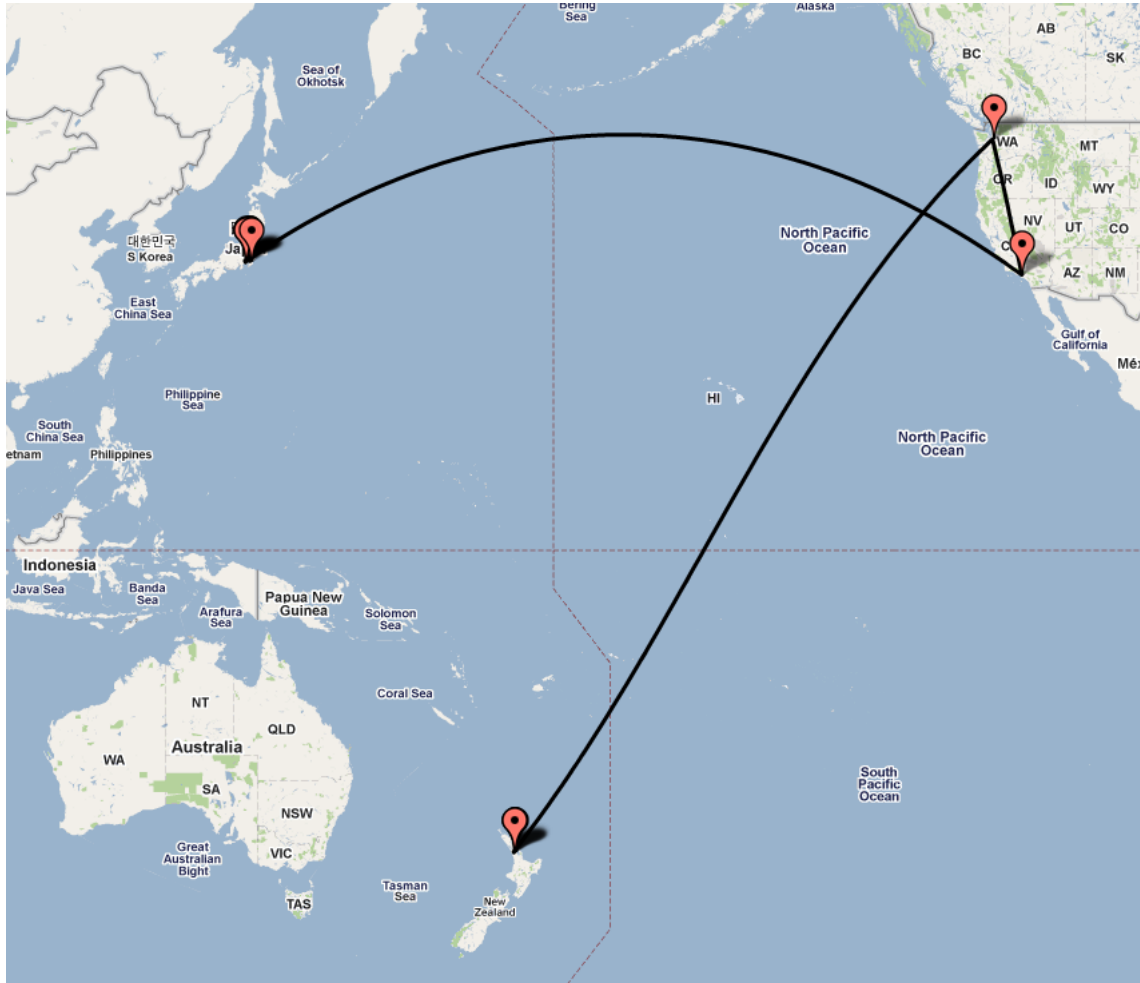


links are weighted up with their prevalence in the traceroute data set

# Characterizing network paths

- Circuitousness
- Direction dependence of lateral deviations
- Hop distance analysis
- Symmetry of Internet routes

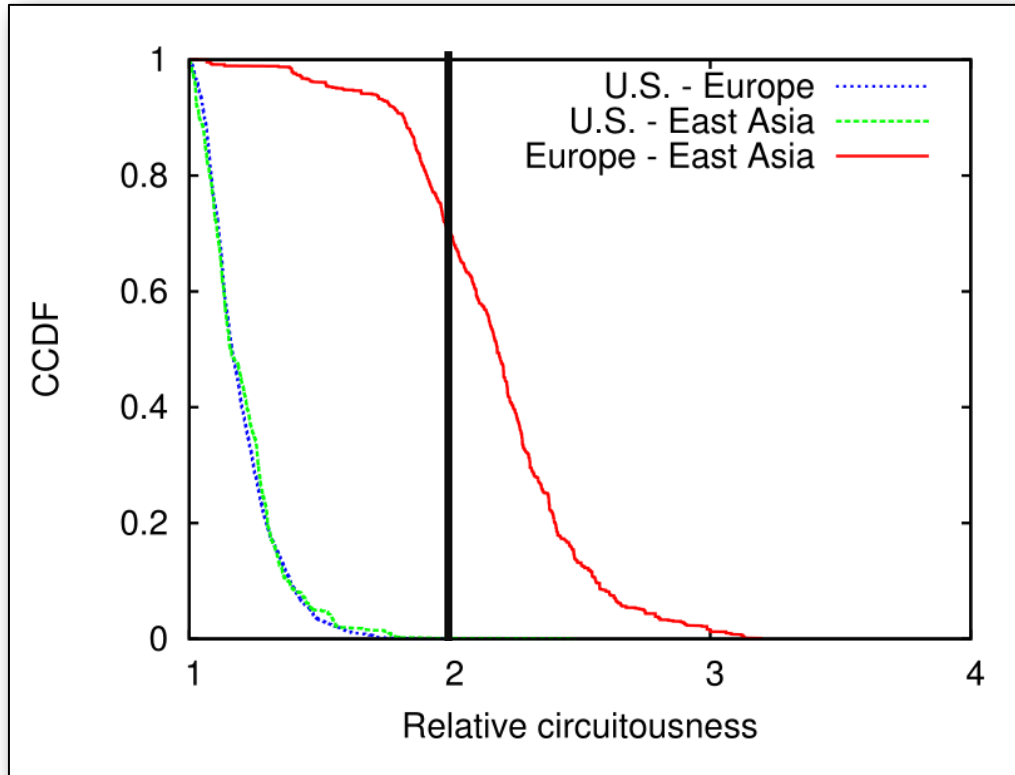
# Aggregated path length



$$L(P) = \sum_{i=0}^N S(p_i, p_{i+1}),$$

The sum of the length of the consecutive links.

# Circuitousness

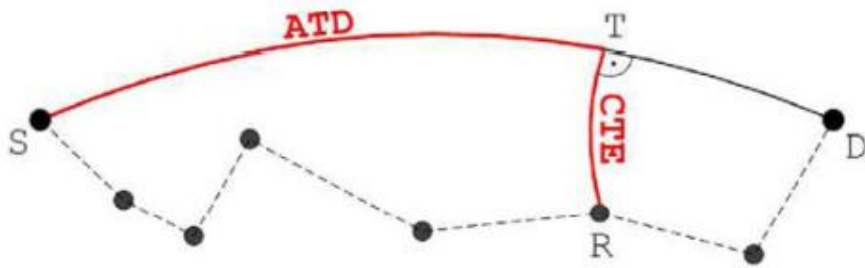


$L(P)$   
The spherical distance  
of the two endpoints

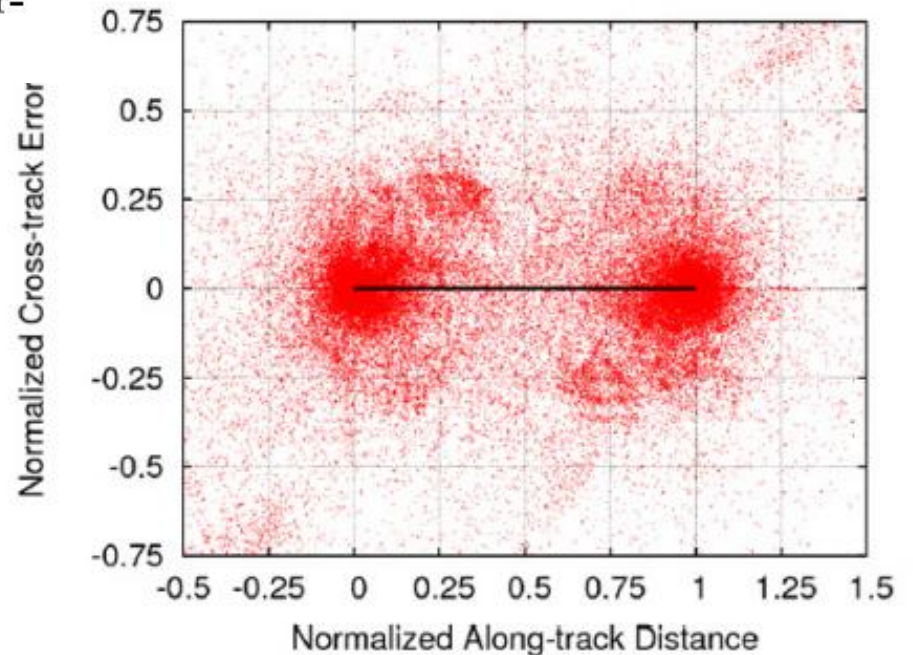
Geographic, geopolitical and economical factors also affect routing



# Direction dependence of lateral deviations



(a) Along-track Distance and Cross-track Error.



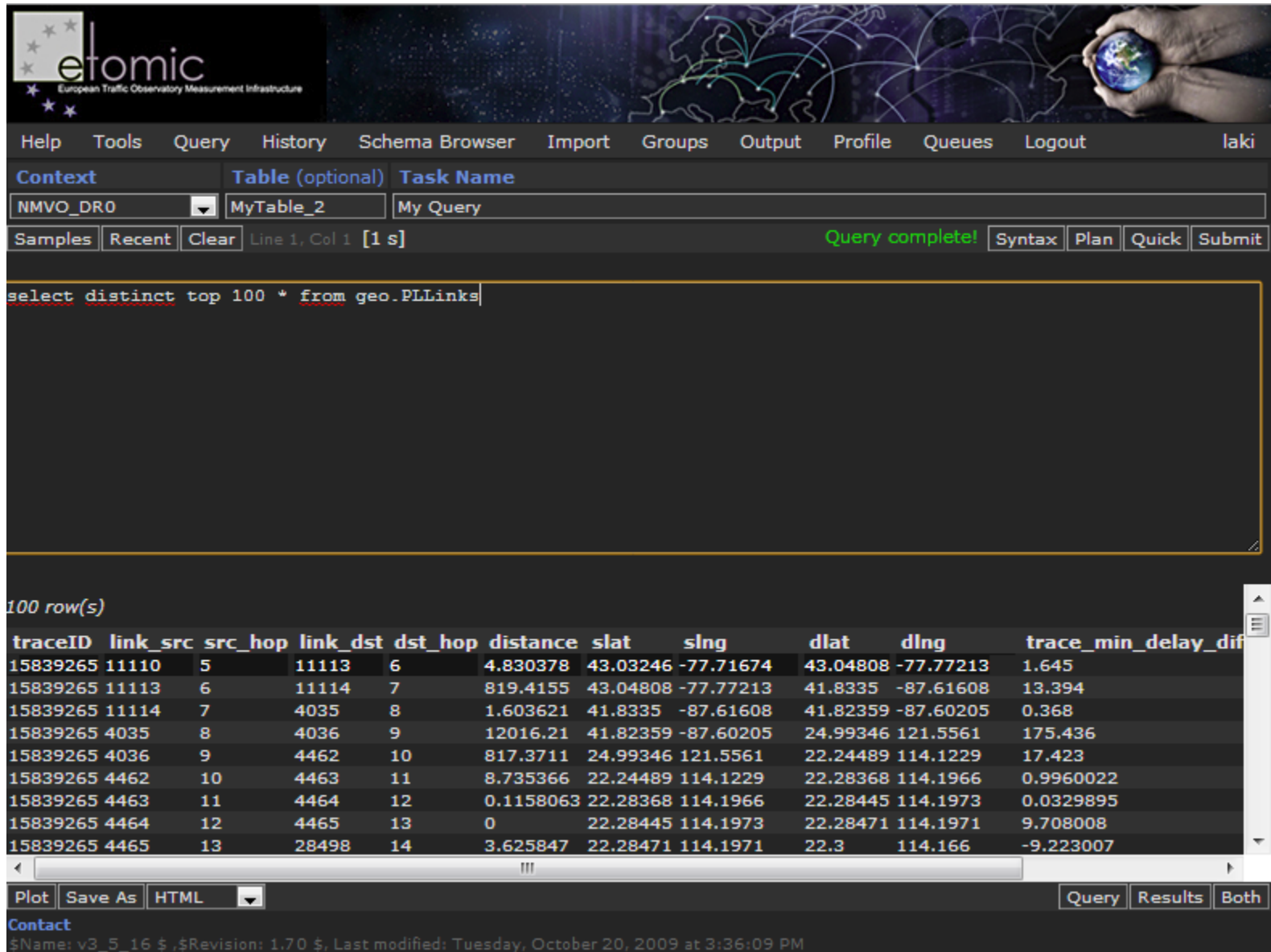
(b) Normalized spatial distribution of routers.



# Data

- Our data is freely available
- Stored in **Network Measurement Virtual Observatory (NMVO)**
  - Etomic's data sharing platform
  - Easy-to-use web Interface
    - to access different data in different databases
  - Standard **SQL queries**
  - Data exportation into CSV and XML
  - <http://nm.vo.elte.hu>
- Tutorial on getting the data:
  - <http://spotter.etomic.org/routerdata>

# Data: <http://nm.vo.elte.hu>



The screenshot shows the etomic web interface. At the top left is the etomic logo with the text "European Traffic Observatory Measurement Infrastructure". The navigation menu includes: Help, Tools, Query, History, Schema Browser, Import, Groups, Output, Profile, Queues, Logout, and laki. The main interface has a "Context" section with "Table (optional)" set to "MyTable\_2" and "Task Name" set to "My Query". Below this are buttons for "Samples", "Recent", and "Clear", along with a status indicator "Line 1, Col 1 [1 s]" and a green message "Query complete!". There are also buttons for "Syntax", "Plan", "Quick", and "Submit". The query editor contains the SQL query: `select distinct top 100 * from geo.PLLinks`. Below the query editor, it indicates "100 row(s)" and displays a table of results. The table has 12 columns: traceID, link\_src, src\_hop, link\_dst, dst\_hop, distance, slat, slng, dlat, dlng, and trace\_min\_delay\_diff. The results are as follows:

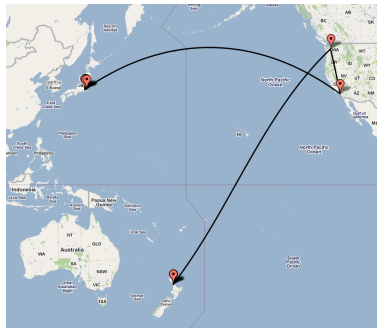
traceID	link_src	src_hop	link_dst	dst_hop	distance	slat	slng	dlat	dlng	trace_min_delay_diff
15839265	11110	5	11113	6	4.830378	43.03246	-77.71674	43.04808	-77.77213	1.645
15839265	11113	6	11114	7	819.4155	43.04808	-77.77213	41.8335	-87.61608	13.394
15839265	11114	7	4035	8	1.603621	41.8335	-87.61608	41.82359	-87.60205	0.368
15839265	4035	8	4036	9	12016.21	41.82359	-87.60205	24.99346	121.5561	175.436
15839265	4036	9	4462	10	817.3711	24.99346	121.5561	22.24489	114.1229	17.423
15839265	4462	10	4463	11	8.735366	22.24489	114.1229	22.28368	114.1966	0.9960022
15839265	4463	11	4464	12	0.1158063	22.28368	114.1966	22.28445	114.1973	0.0329895
15839265	4464	12	4465	13	0	22.28445	114.1973	22.28471	114.1971	9.708008
15839265	4465	13	28498	14	3.625847	22.28471	114.1971	22.3	114.166	-9.223007

At the bottom of the interface, there are buttons for "Plot", "Save As", and "HTML", along with a "Query Results Both" section. A "Contact" section at the very bottom provides the following information: `$Name: v3_5_16 $, $Revision: 1.70 $, Last modified: Tuesday, October 20, 2009 at 3:36:09 PM`

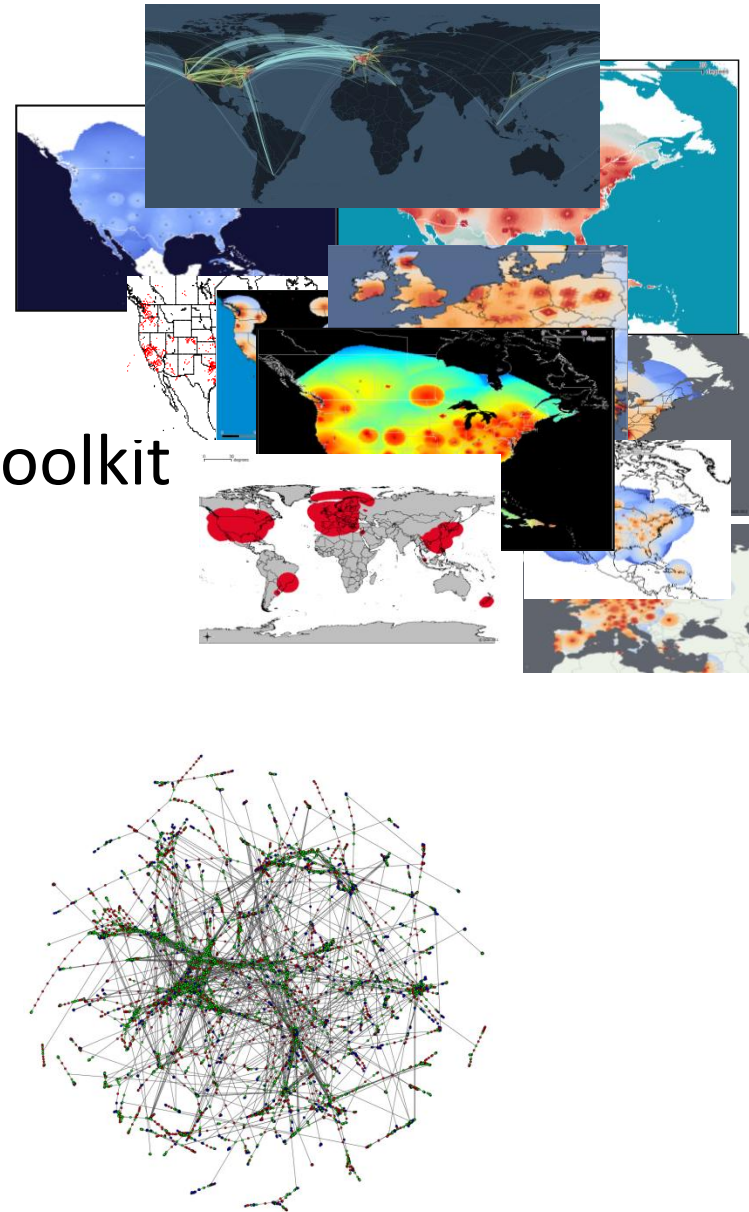
# Data Visualization

- Quantum GIS
  - Open source GIS software
  - with our own preprocessing toolkit

- Google Map



- iGraph for R



# What do we expect from you?



- More data
  - More complete router level topologies
  - Or other measurements

- Collaboration

- Please feel free to use our data
  - <http://spotter.etomic.org/routerdata>
- Measurement agents other than PlanetLab nodes
  - E.g. in commercial networks
- Federating external data sources in NMVO
  - Unified querying interface



- Feedback on our tools

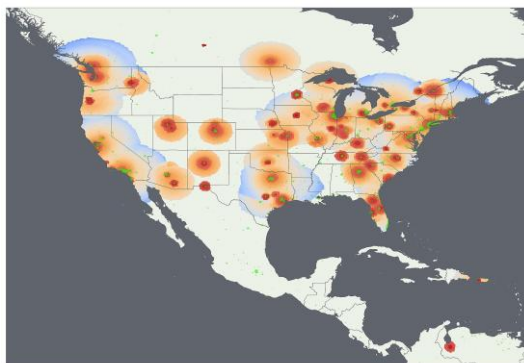
- NMVO – A data sharing platform with unified SQL-based querying interface
  - <http://nm.vo.elte.hu>
- Spotter – An active IP Geolocation service
  - <http://spotter.etomic.org>
- SONoMA – A Network Measurement Platform
  - <http://sonoma.etomic.org>

**I WANT YOU**



**FOR FEEDBACK!**

# Thank you



Contact:  
Sándor Laki  
[laki@etomic.org](mailto:laki@etomic.org)

