

---

# Data Intensive Geospatial Analysis and Computation

---

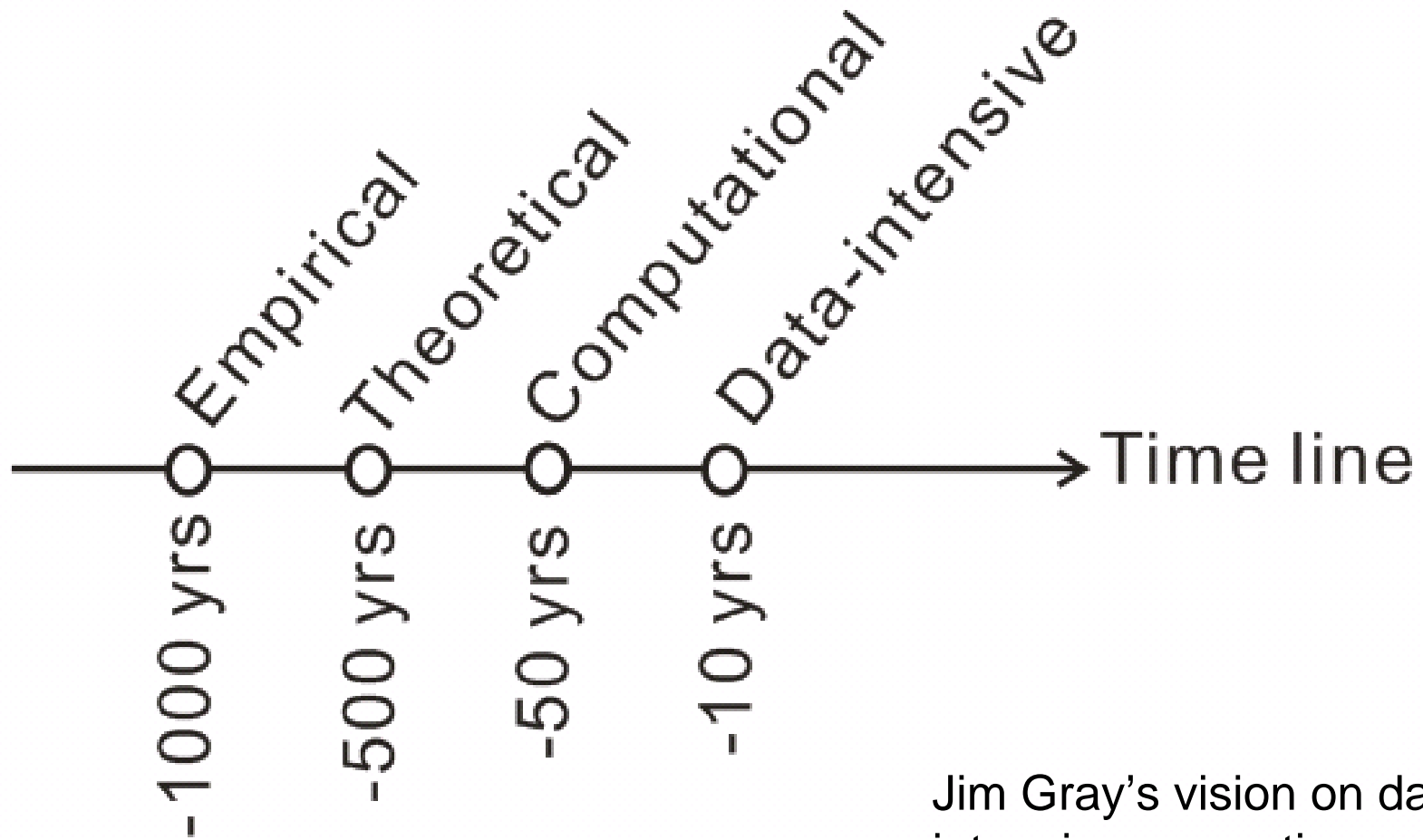
Bin Jiang

KTH Research School, University of Gävle,  
Sweden

<http://fromto.hig.se/~bjg/>

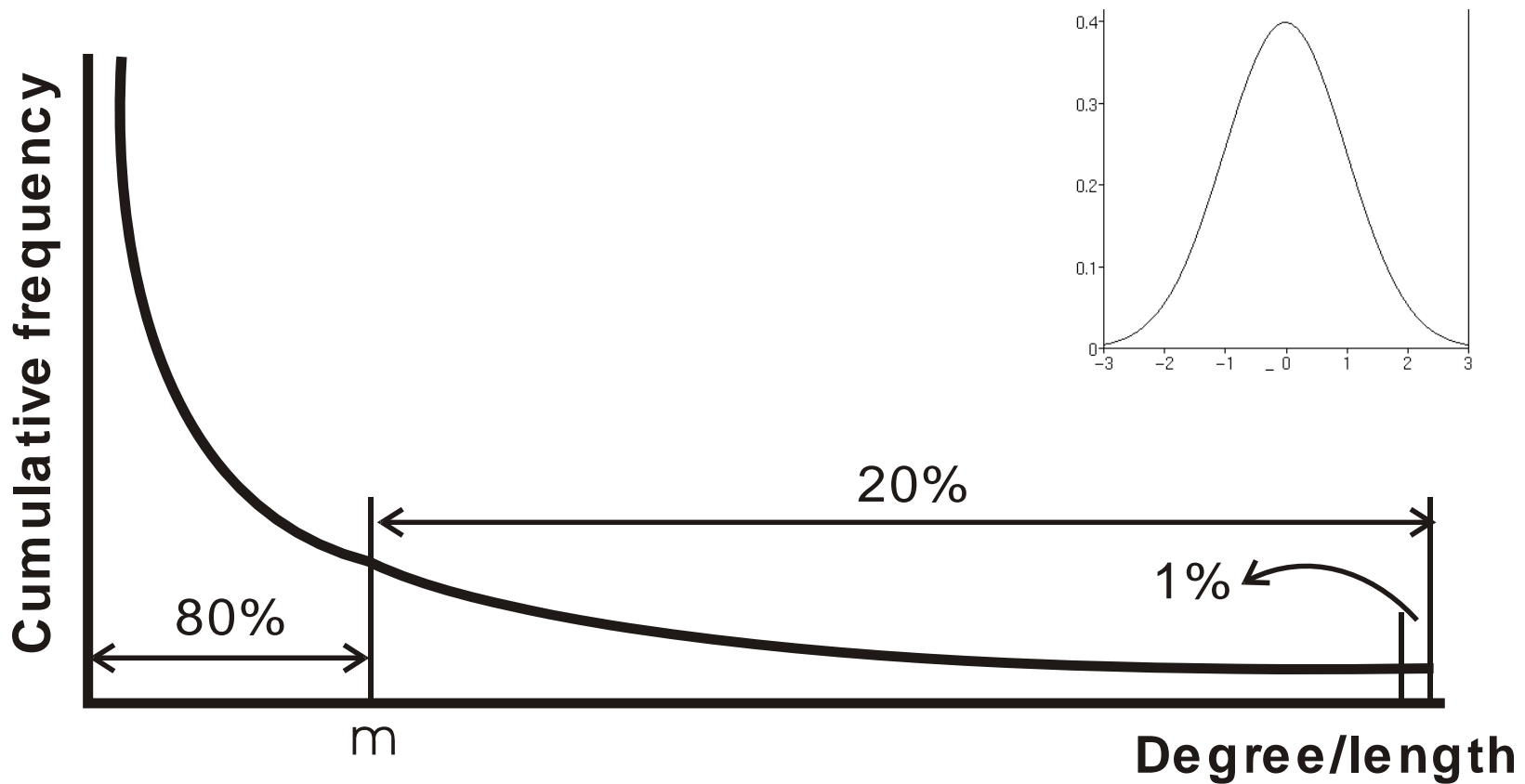
---

# Four scientific paradigms in history



Jim Gray's vision on data intensive computing

# A universal topological pattern – a longtail pattern

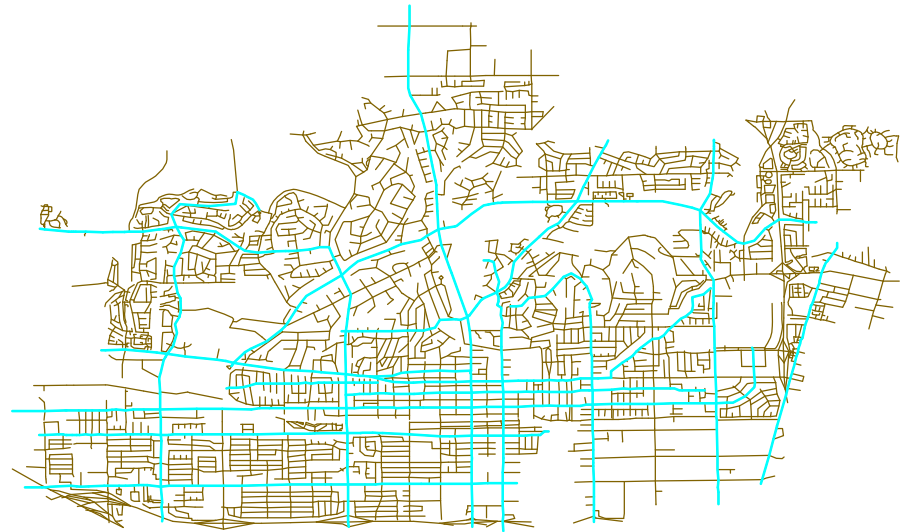


Small events

Large events

---

80% trivial versus 1% vital



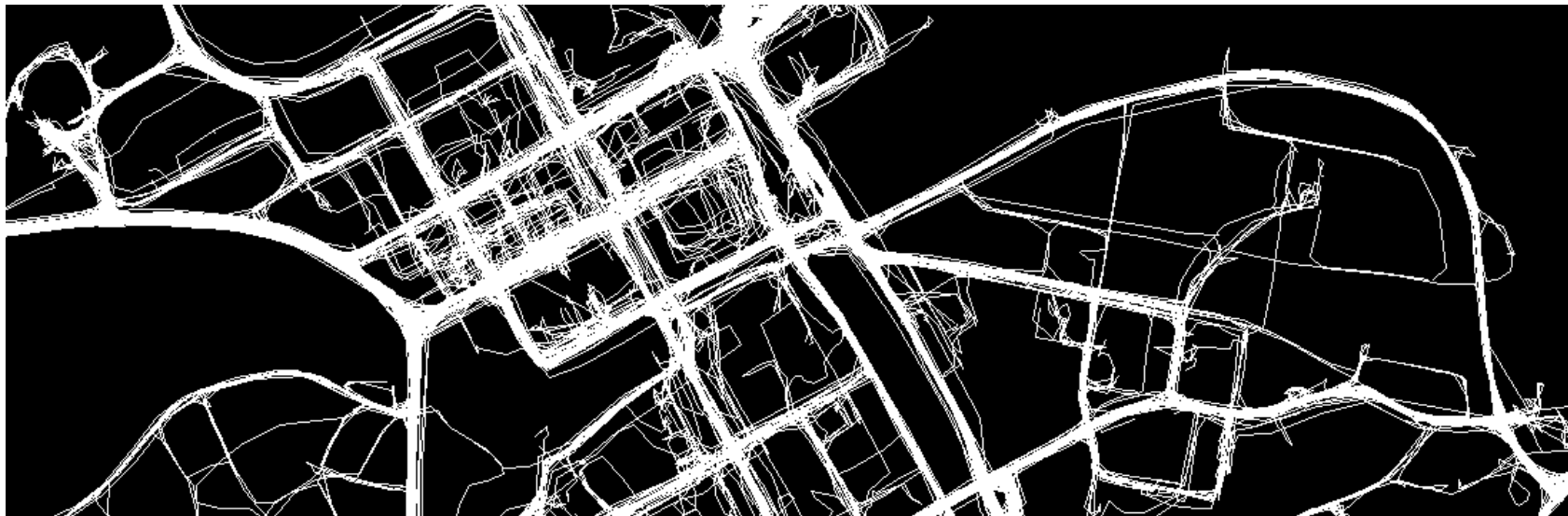
# Hägerstrand project



The Swedish Research Council Formas

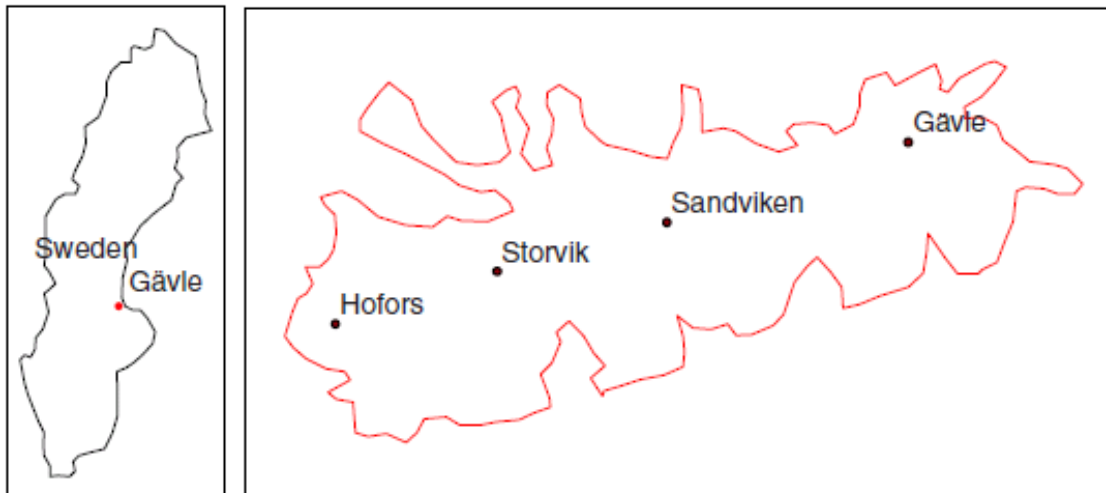


**Professor Hägerstrand (1916-2004)**



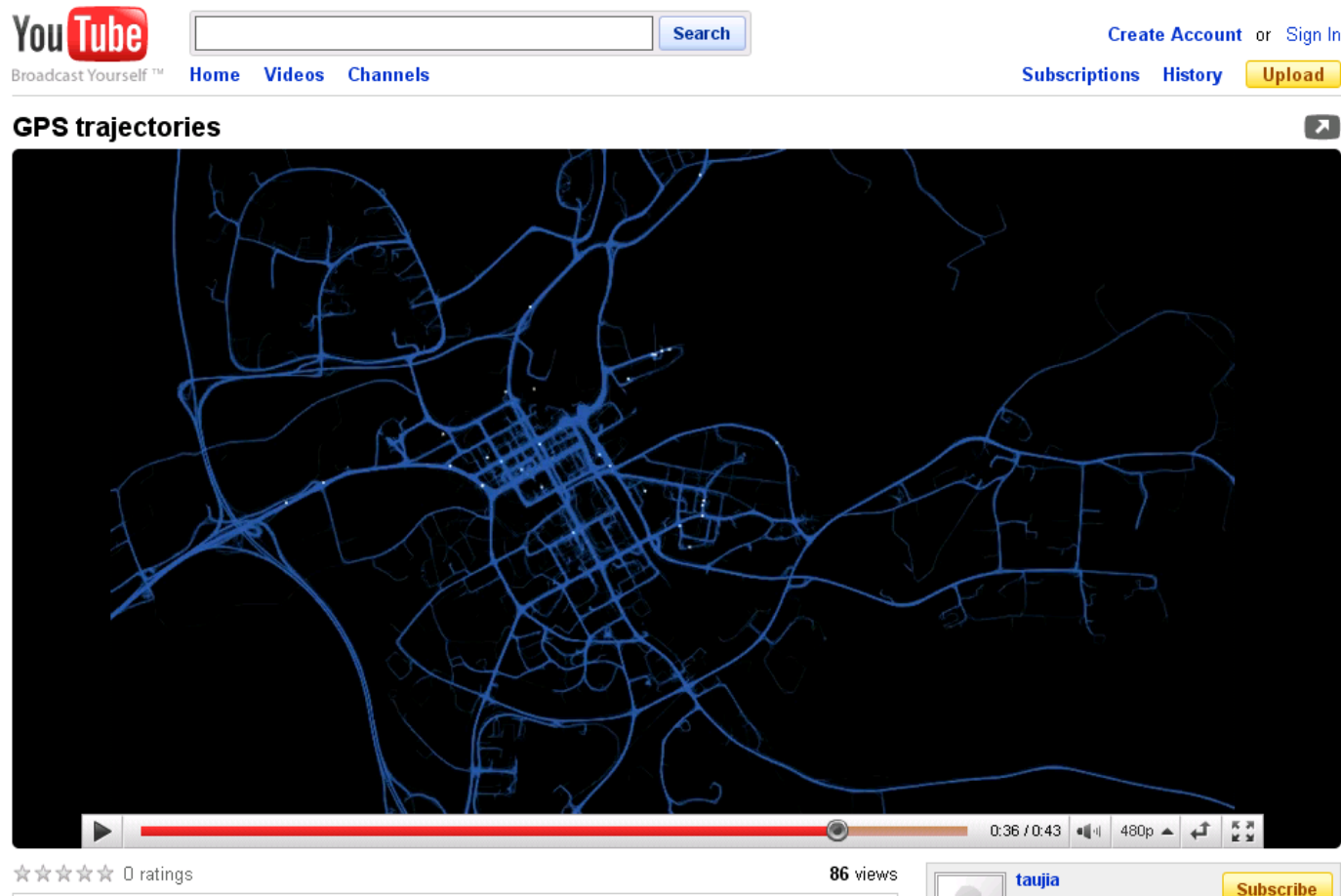
## Data collection and processing

- A six-month GPS data of 50 taxicabs' positions obtained automatically by GPS receivers,
- Anonymized customer data as to when and where customers are picked up and dropped off, and
- The underlying street network (4 056 natural streets from a total of 10 439 street segments)



# Human mobility – observed trajectories

- <http://www.youtube.com/watch?v=5SyXxbU3Feg>



The image shows a screenshot of a YouTube video player. At the top left is the YouTube logo with the tagline "Broadcast Yourself™". To its right is a search bar with a "Search" button. Further right are links for "Create Account" or "Sign In", and buttons for "Subscriptions", "History", and "Upload". Below the navigation bar, the video title "GPS trajectories" is displayed. The main video area shows a map with a complex network of blue lines representing movement paths. At the bottom of the video player, there is a progress bar, a timestamp of "0:36 / 0:43", and icons for volume, resolution (480p), and full screen. Below the video player, there are five stars for ratings (0 ratings), "86 views", a channel name "tauja" with a profile picture, and a "Subscribe" button.

# Human mobility – scaling of trail length

- [http://www.youtube.com/watch?v=hG7DOyzseps&feature=player\\_embedded](http://www.youtube.com/watch?v=hG7DOyzseps&feature=player_embedded)

YouTube  
Broadcast Yourself™

Home Videos Channels

Create Account or Sign In

Subscriptions History Upload

Scaling of trail length



0:15 / 0:16 480p

☆☆☆☆ 0 ratings

48 views

taijia



# Simulation of random and purposive movement

- [http://www.youtube.com/watch?v=P\\_1htVK9D4E](http://www.youtube.com/watch?v=P_1htVK9D4E)

YouTube

Broadcast Yourself™


Home Videos Channels

Search

Create Account or Sign In

Subscriptions History Upload

Simulation of traffic



0:08 / 0:24

360p

★★★★☆ 0 ratings

8 views

tauja

Subscribe

---

## Data-intensive computing

- Under the fourth paradigm, data scale dealt with is terabytes ( $10^{12}$  bytes) or petabytes ( $10^{15}$  bytes)
- Google Maps or Google Earth has integrated terabytes of geospatial data for the world
- HOWEVER, these data are in the hands of
  - National mapping agencies
  - Private companies (e.g., Tele Atlas, NAVTEQ or DigitalGlobe)
- Steve Coast's vision leads to openstreetmap

# OpenStreetMap (OSM)

- A wiki-like collaboration to create a free editable map of the world.



# Google Maps (left) vs OSM (right)

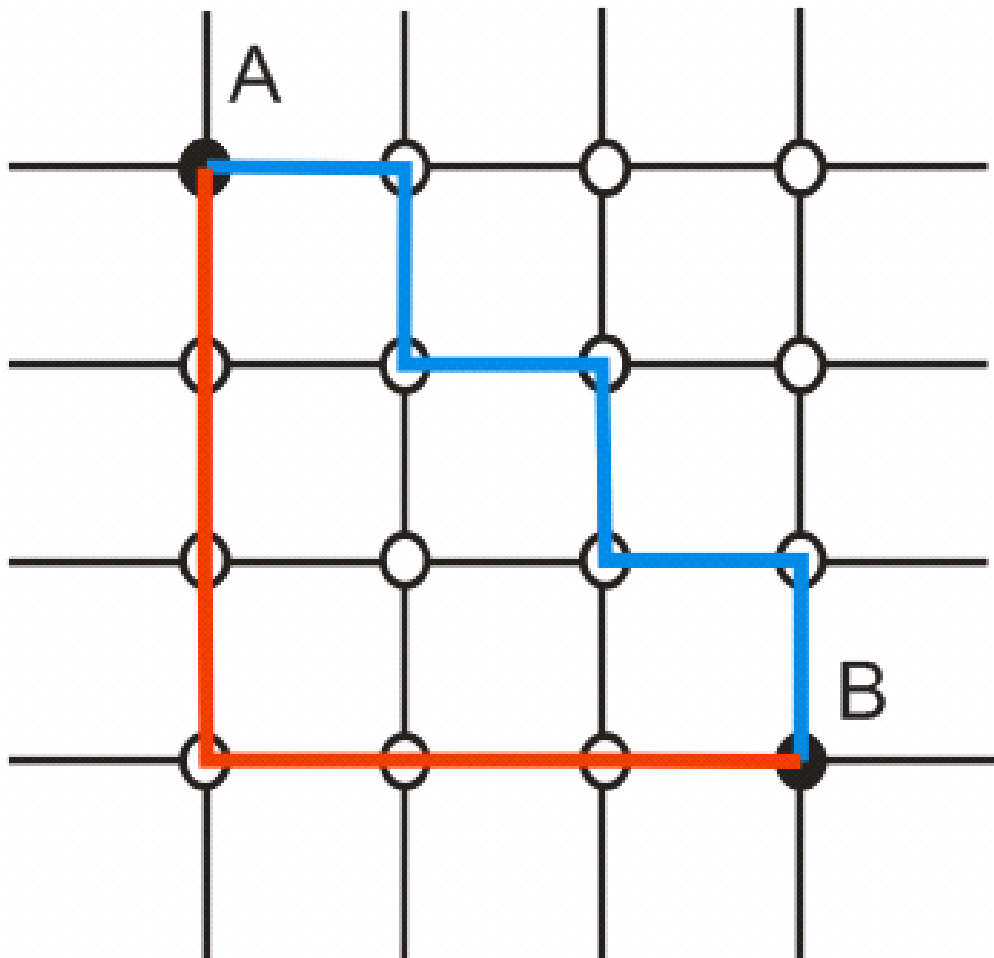


---

## Our goal with FromToMap is...

- To build up an intelligent routing service that are
  - based on a novel method for deriving a **fewest-turn-and-shortest route** between two locations F and T.
  - using **OpenStreetMap (OSM)**,
  - for both **route planning** and **personal navigation**.

## Routes with different turns



Shortest  $\neq$  Fastest

---

## Computation of fewest-turn routes

- A huge graph of 10 millions nodes and 17 millions links for the entire Europe road networks
- about 30 GB memory occupied;
- Even more larger graph of about 50 millions nodes and 70 millions links for the entire world road network
- about 200 GB memory needed.

# Isochrones for the entire Europe network

**FromTo** Stockholm    
MAP A semantic-based map engine using openstreetmap (BETA).

**Get Directions** (Europe) **Isochrones** (Europe)  
(Right click map to pick up location I)

Walking

**Point of Interest (POI) Information:**

Select POI by: All

- 1 Restaurant: Shanti
- 2 Restaurant: Nem Nem Quan
- 3 Cinema: Filmstaden Söder
- 4 Restaurant: Kebaben
- 5 ATM: Swedbank
- 6 Swimming: Forsgrenska
- 7 Restaurant: Magnus krog
- 8 Fast Food Restaurant: Sushi
- 9 Cafe

Data CC-BY-SA by [OpenStreetMap](#)  
Scale = 1 : 54K  
17.99617, 59.32193



# FromToMap demo ([www.fromtomap.org](http://www.fromtomap.org))

- [http://www.youtube.com/watch?v=9rSQ\\_zoKoEA](http://www.youtube.com/watch?v=9rSQ_zoKoEA)

The screenshot shows a YouTube video player displaying a demo of the FromToMap website. The video content shows a web browser window with the URL <http://www.fromtomap.org/>. The website interface includes a search bar with "Stockholm" entered, a "Search Map" button, and a "Get Directions" section. The "Get Directions" section shows a route from Myntgatan to Viagatan, with a "Driving" mode selected. Below this, there is a list of driving directions to Viagatan, starting from Myntgatan. The directions are as follows:

Step	Instruction	Distance
1	Head SouthEast on Myntgatan	0.2km
2	Turn left at Rådhusgränd after 3 blocks	0.8km
3	Turn right at Rodbedtorget after 4 blocks	0.4km
4	Turn left at Herkulesgatan after 3 blocks	0.2km
5	Turn right at Klaratunneln after 1 blocks	122km
6	Turn left at 223 after 146 blocks	11.1km
7	Turn left at 910 after 9 blocks	34.2km
8	Turn right at Unnamed Road after 5 blocks	24.2km

The video player interface includes the YouTube logo, a search bar, and navigation links like "Home", "Videos", "Channels", "Subscriptions", "History", and "Upload". The video title is "FromToMap demo" and it has 2 views. The uploader's name is "tauja" and there is a "Subscribe" button. The video progress bar shows 1:54 / 2:46.

---

## A few reflections

- Geospatial research should go beyond the data scale of kilobytes and megabytes.
  - Do not assume space is normally distributed (homogeneity), and space with a massive large data scale is likely to bear enormous heterogeneity (Goodchild 2009, Anselin 2006) .
  - OSM can be a benchmark data for geospatial research.
  - Following up the third point, geospatial community should setup a data repository to archive data, algorithms and source codes which can be shared among researchers.
-

- 
- Thank you for your attention