Spatial Big Data

Dr.Waleed M.Ead

Lec2

Data science

- Data science, also known as data-driven science, is an interdisciplinary field of scientific methods, processes, algorithms and systems to extract knowledge or insights from data in various forms,

either structured or unstructured, similar to data mining.

Spatial Data science is multi-disciplinary

- Database
- Big data system (Hadoop)
- Statistics
- Machine learning
- Data mining
- Data visualization

Spatial data science

- An interdisciplinary field about process and systems to extract knowledge or insights as well as causal relationship of spatial phenomena from spatial and other data in various forms

Spatial Data science is multi-disciplinary

- Data science
- GIS
- Spatial DBMS
- Spatial Statistics

- spatial Data mining
- Spatial Data visualization

Why do the major IT companies have maps?

- Excellent contents
- Spatialization of information
  - Useful in sentiment analysis
- Location based advertisement
- Interface between human and robot
  - Nokia sells their maps unit to BMW and Mercedes for $3 billion
- Platform of everything (spatial big data)

what is Spatial Big Data?

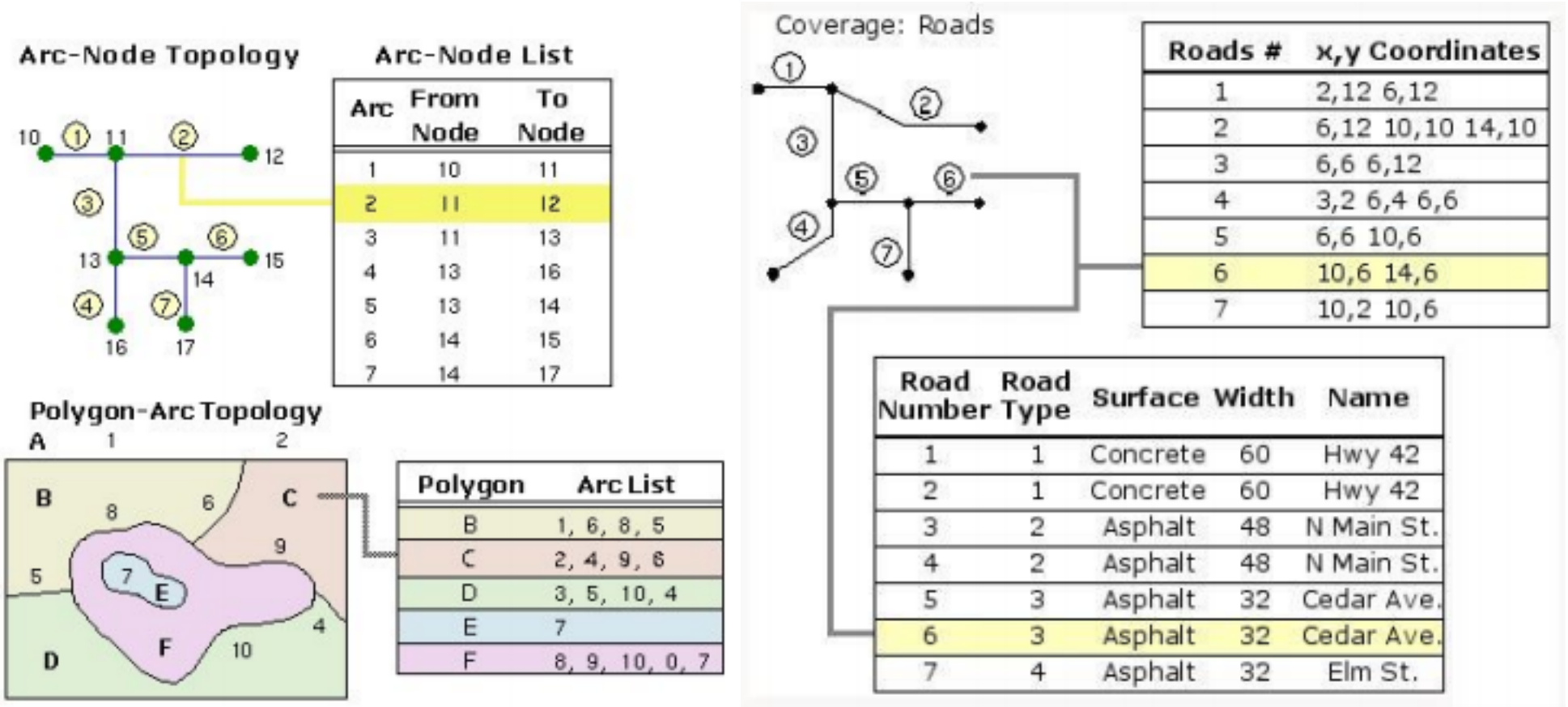
- All types of data objects or elements that have geographical information present
- Enables the global finding and locating of individuals or devices
- Also known as geospatial data, spatial information, geographic information

what is Spatial Big Data?

- Raster data
  - Geoimages (obtained by satellites for example)
  - 3D objects

- Vector data
  - Points, Lines, Polygons
- Graph data
  - Road networks (an edge = a road segment and a node = intersection)
  - Topological coverage

Topological Coverage



Contains both the location and attribute data

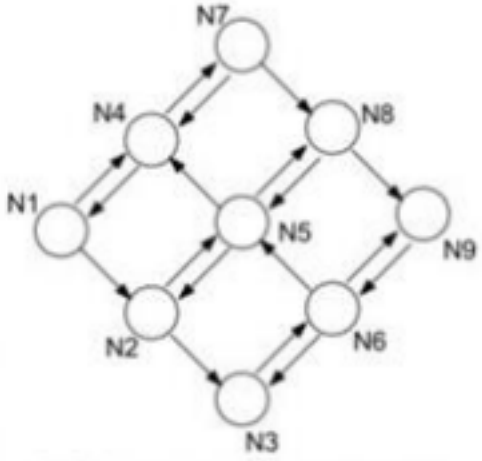
Spatial big data

- Spatial Big Data exceeds the capacity of commonly used spatial computing systems
  - due to volume, variety and velocity
- Spatial Big Data comes from many different sources • satellites, drones, vehicles, geosocial networking services, mobile devices, cameras
- A significant portion of big data is in fact spatial big data

Types of Spatial Big Data

- Speed every minute for every road segment
- GPS trace data from cell-phones

(a) A Road Map [22]



(b) Graph Representation

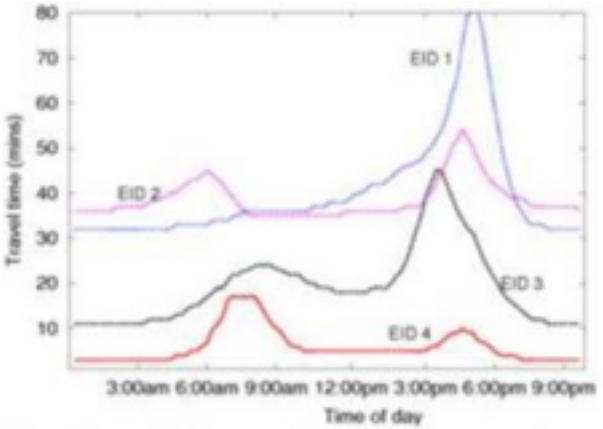
- Engine measurements of fuel consumption (can be estimated from fuel levels, distance travelled and engine idling from engine RPM)
- Greenhouse gas emissions

Use cases for spatial Big Data

- Eco routing
- Tracking Endangered Species
- Better crop production, reducing costs
- Detecting extreme events

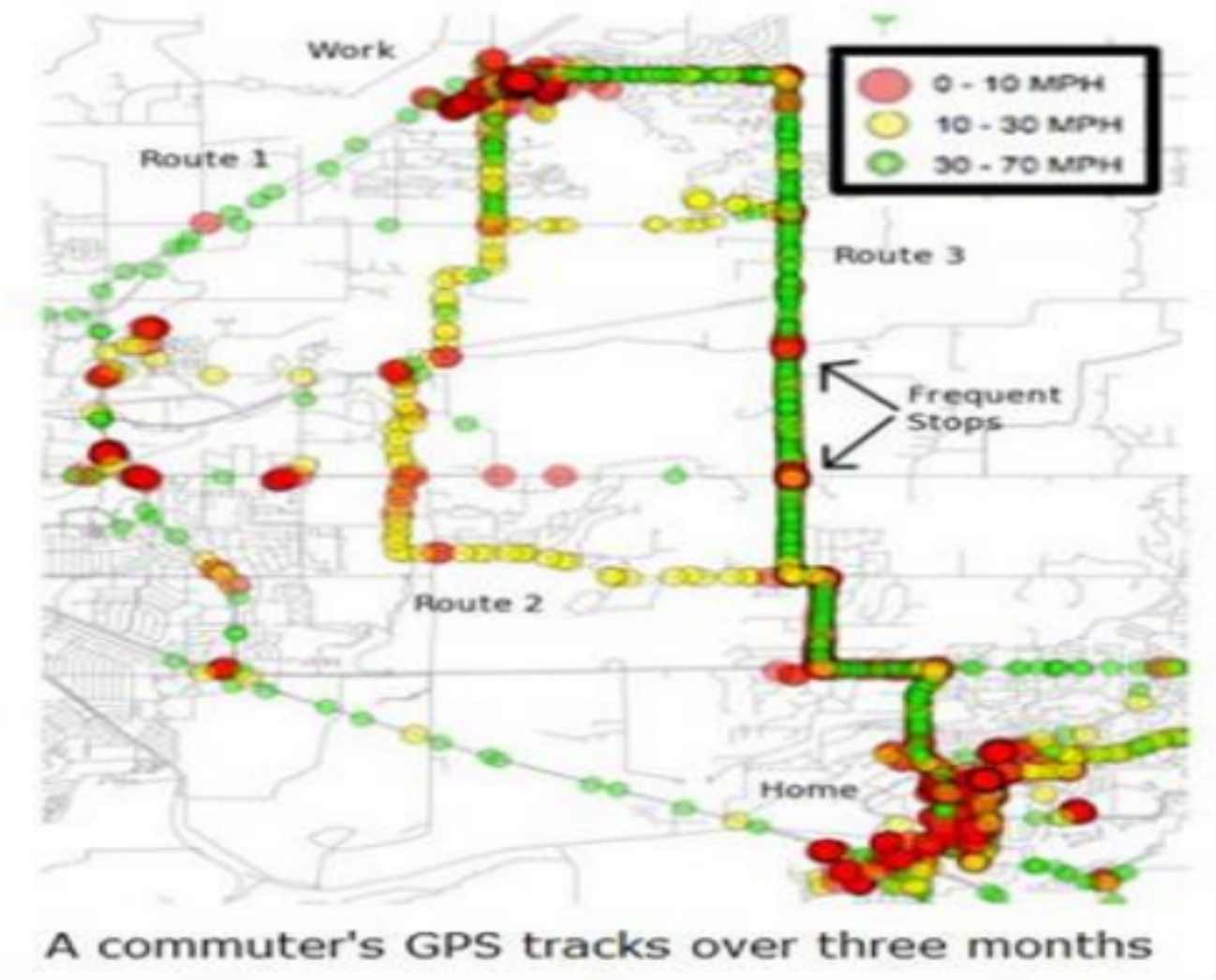
Eco routing

- Next generation routing service
  - avoids congestion
  - reduces idling at red lights
  - avoids left turns
- Estimation: in 2020 about $600 billion is saved a
- Takes into account various datasets
  - real-time and historic traffic data of engine measurements
  - speed-limits
  - road types
  - “rush hour vs non-rush hour”



(a) Travel time along four road segments over a day.

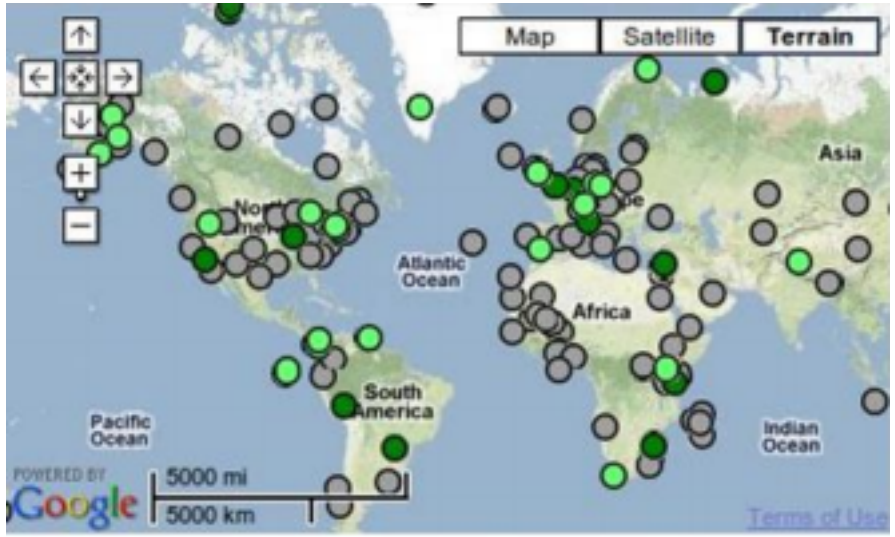
Eco routing



A commuter's GPS tracks over three months

Tracking endangered species

- 2013: 970 studies over 250 contributors, 41,170 tracks and 61 million locations
- Movebank: a free online database of animal tracking data



Better crop production

- “If you can grow crop fast in these circumstances, query for similiar places”

Detecting extreme events

- Earthquakes
- Wildfires
- Flooding

- Other calamities

How to detect

- Built-in motion detectors in mobile phones • Using

unstructured data sets can be used such as tweets

Future

- New Datasets -> need to rapily integrate new datasets and algorithms
- Computational cost increases as the diversity of Spatial Big Data grows
- Easy to collect, sensors (or sensor networks) are becoming more and

more common (Internet of things)

Features of spatial Big data

- Access of data depends on the daytime of where it is used
- Changes dynamically
- Recent Spatial Big Data is usually being generated at a very high speed

Challenges of spatial Big data

- Retaining computational efficiency
- Storing Spatial Big Data into the cloud
- Applying new data when Spatial Big Data or change old data => repartitioning is needed