Geoinformation

Prof. Dr.-Ing. Liqiu Meng

Chair of Cartography and Visual Analytics

WS 24/25

Contents of Lectures

L1: Introduction

L2: Spatiotemporal representations and databases

L3-L4: Spatial data analysis

L5: Cartographic techniques

Spatial data analysis

I. Data quality

II. Pitfalls of spatial data analysis

III. Types of spatial data analysis

III. Types of spatial data analysis

1. The analysis of point patterns
2. Vertical analysis
3. Network analysis

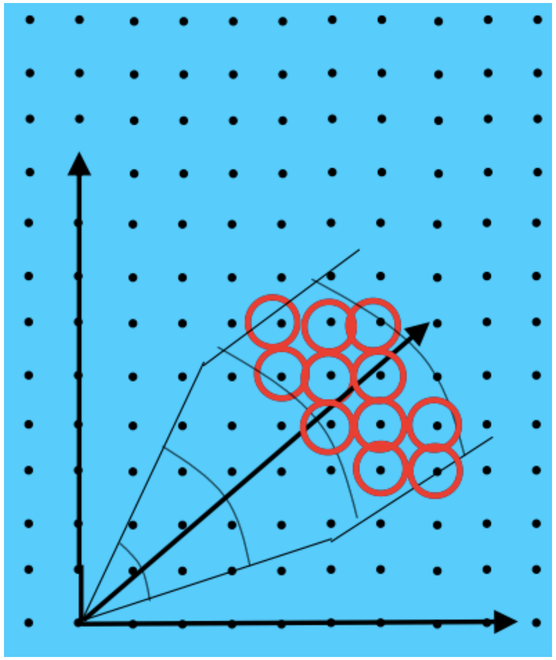
III.1 The analysis of point patterns

Analytical measures of autocorrelation

Variograms

They are used for exploring autocorrelation between all pairs of points within a specified distance.

The autocorrelation can be isotropic or direction dependent.

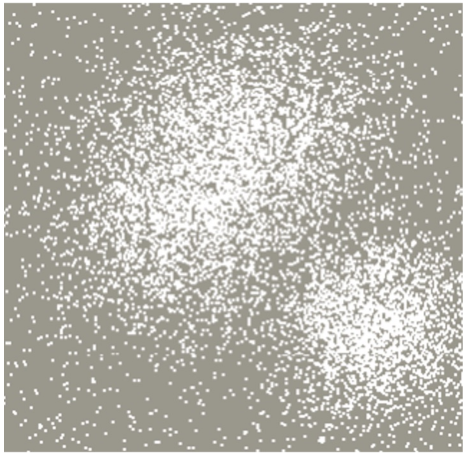


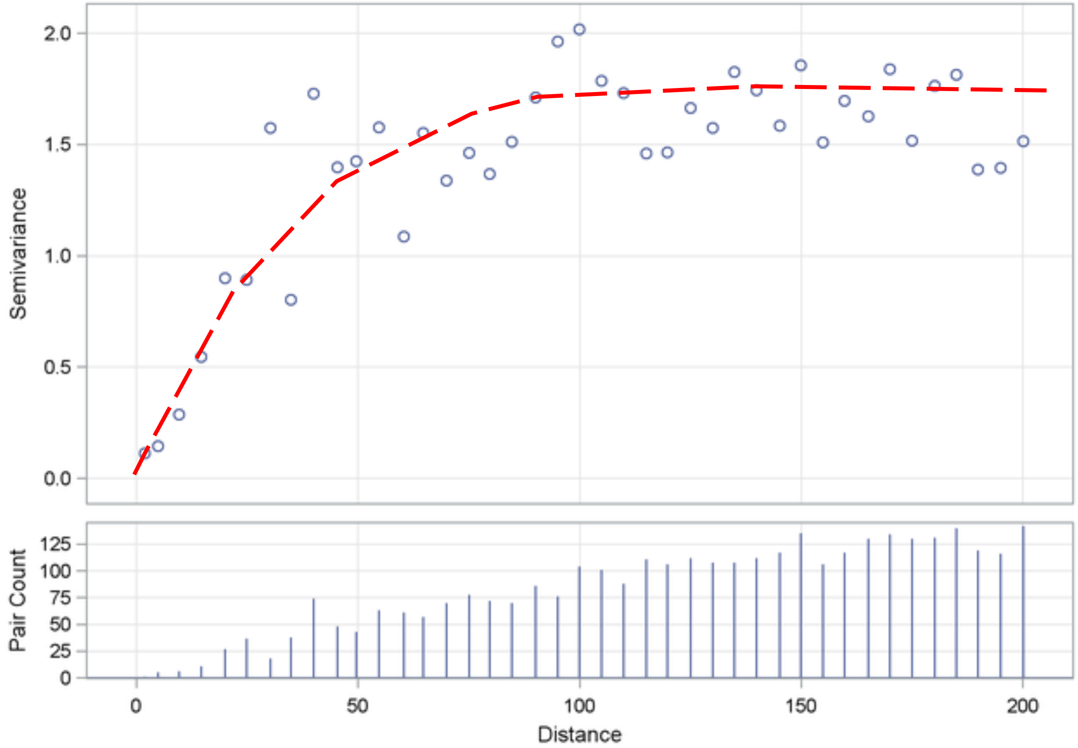
Semivariogram - a representative variogram

A point pair at the distance *h* can be regarded as a vector. The head and tail can have the same or different values. The semivariance measure is:

$$\gamma(h) = \frac{1}{2N_h} \sum_{d_{ij} \approx h}^{N_h} \left(v_i - v_j\right)^2$$

where is $\boldsymbol{N_h}$ the number of pairs within the tolerance ***h***





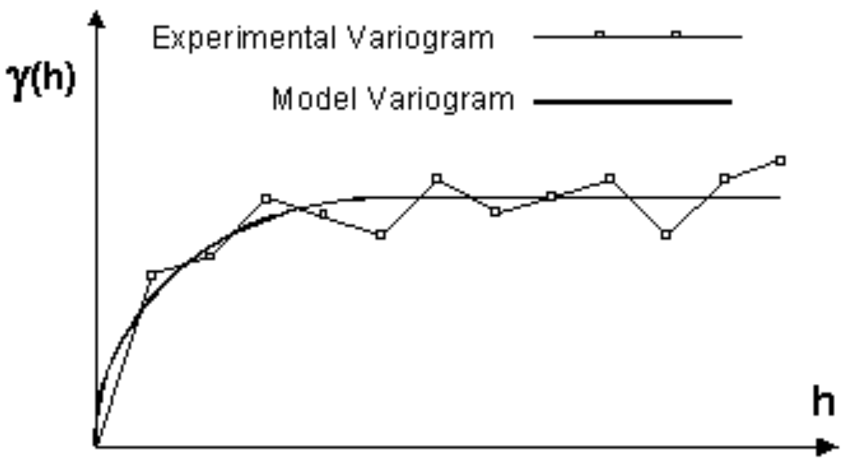
The meaning of a variogram

– If the variogram does not diminish as distance h → 0, then no spatial interpolation is meaningful.

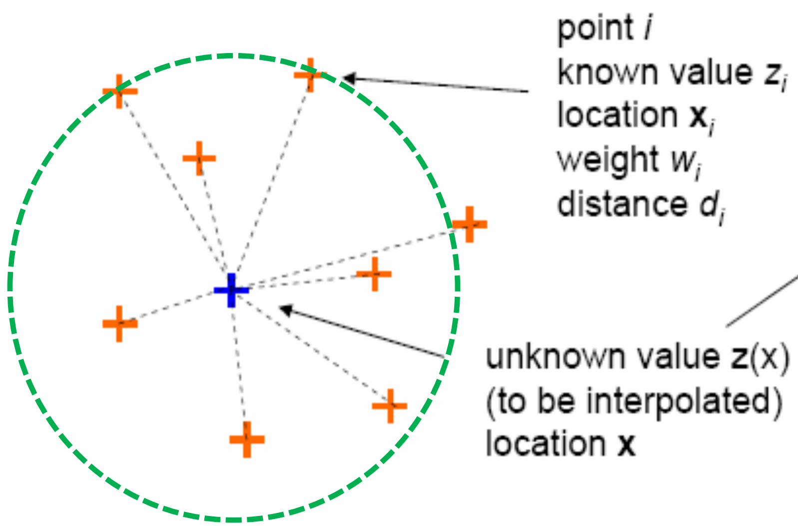
– A noisy variogram suggests that too few points have been used. At least 50-100 data points are necessary.

– The range provides information about the search window. If the range is large, then the long-range variation dominates.

– When the curve reaches the sill, there is no correlation to the points beyond.

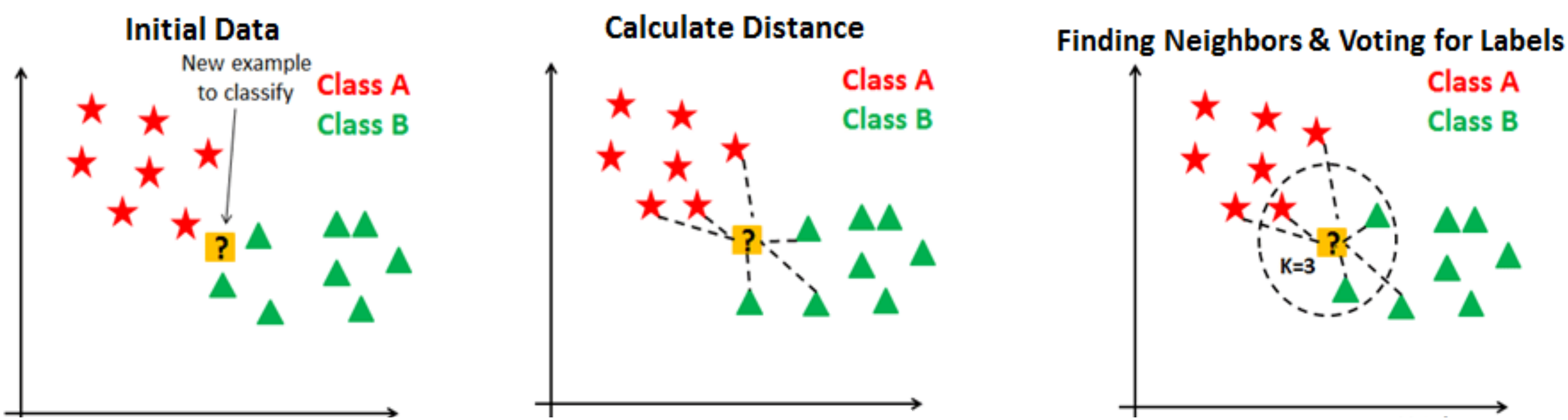


Interpolation guided by variogram



$$z(x) = \sum_i w_i z_i / \sum w_i$$

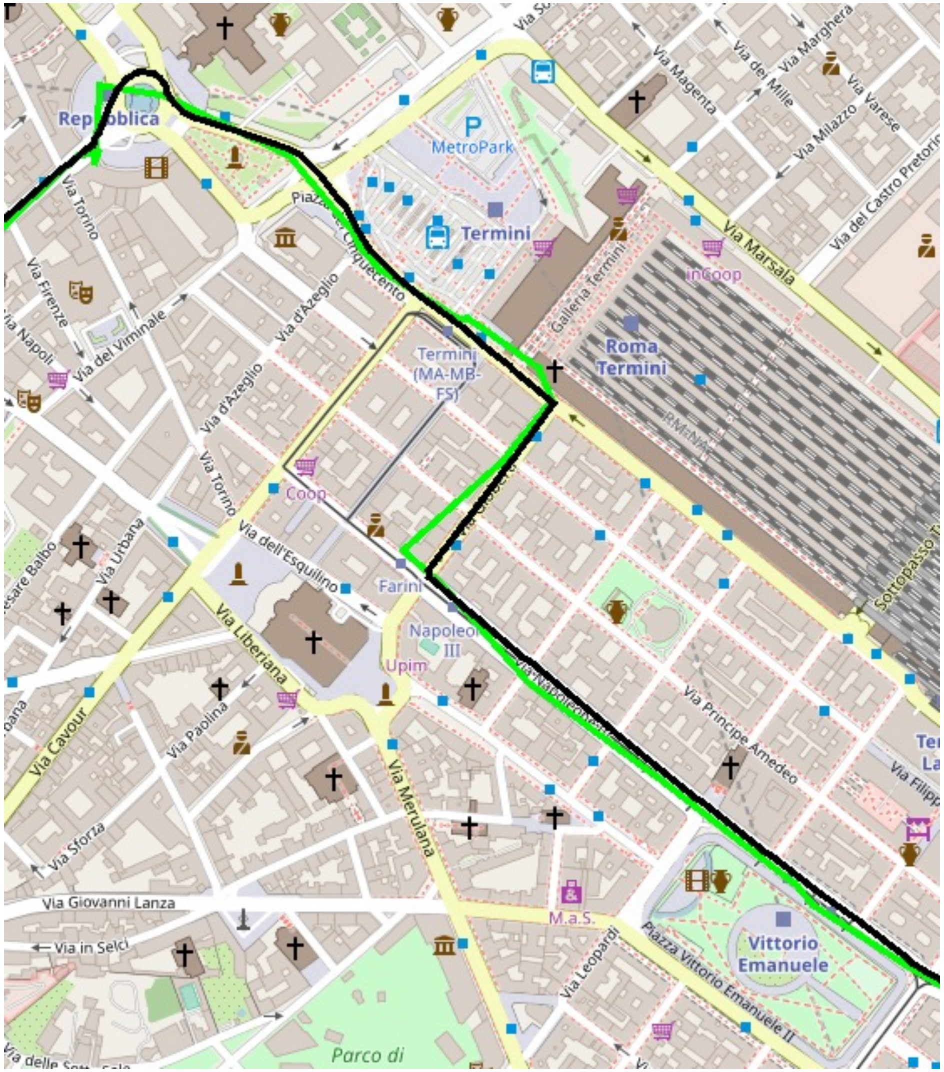
$$w_i = 1/d_i$$

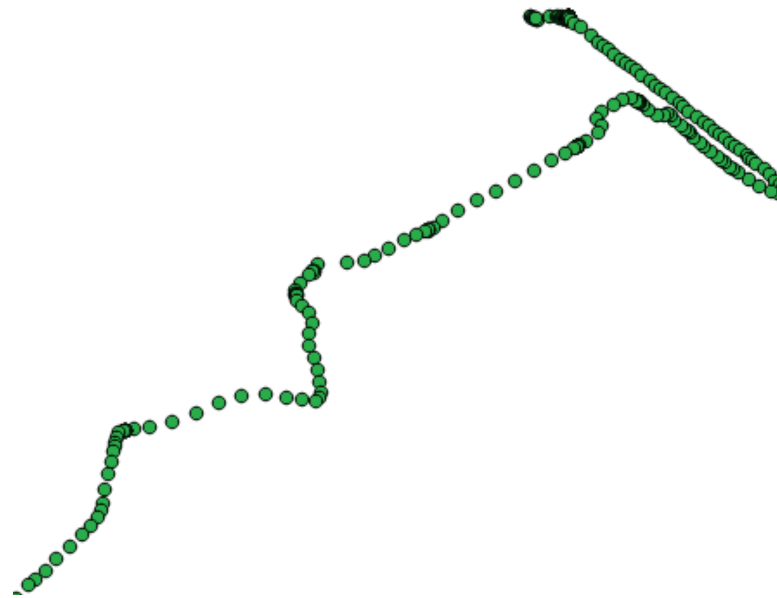


Analysis of mobility points (GNSS points)

Basic attributes:
Point id: unique key of a point
Latitude: latitude of the point
Longitude: longitude of the point
Time: datetime when the point was registered
Track id: unique key of the trajectory the point belongs to
Sampling rate: point density on the trajectory …

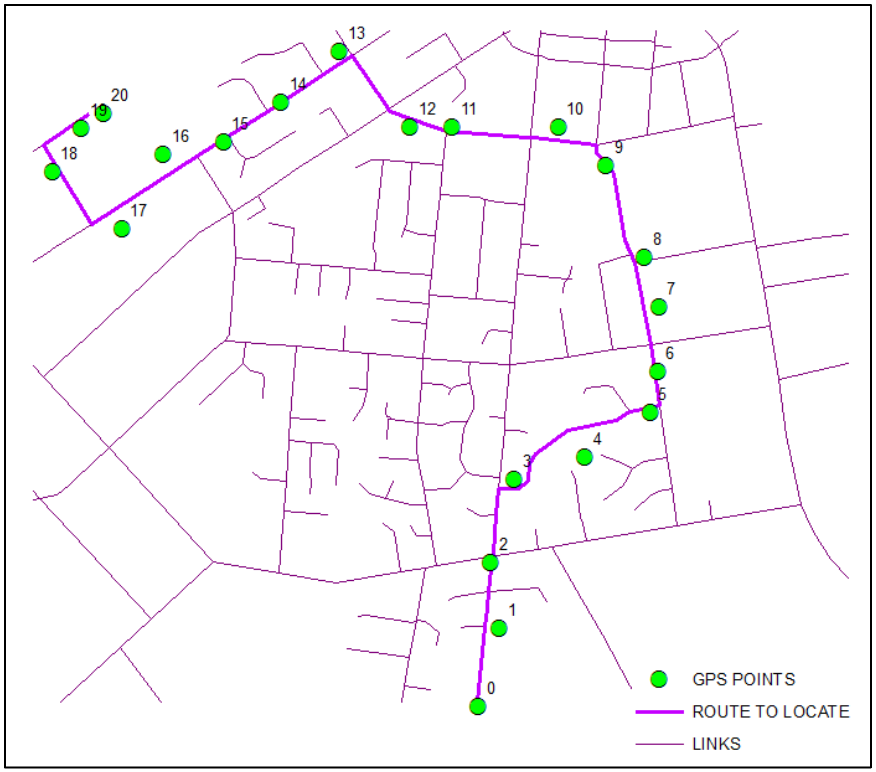
Derived attributes from a moving object:
Heading: derived from the last two point positions
Speed: the actual or average speed along a track
Distance: the total length of a track …





The moving speed varies

A trajectory is typically aligned with an existing road



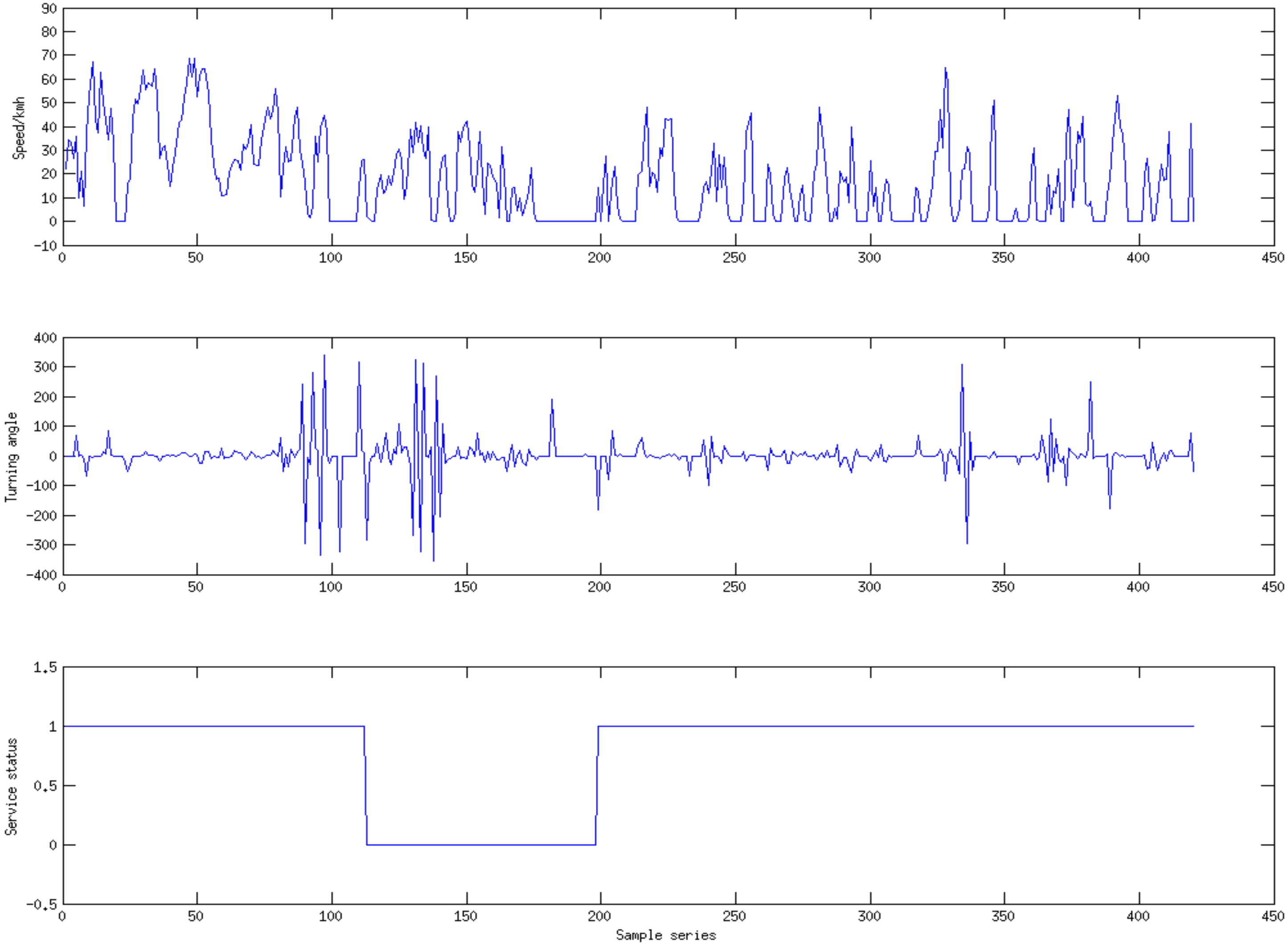
A trajectory with basic attributes

A trajectory with extended attributes

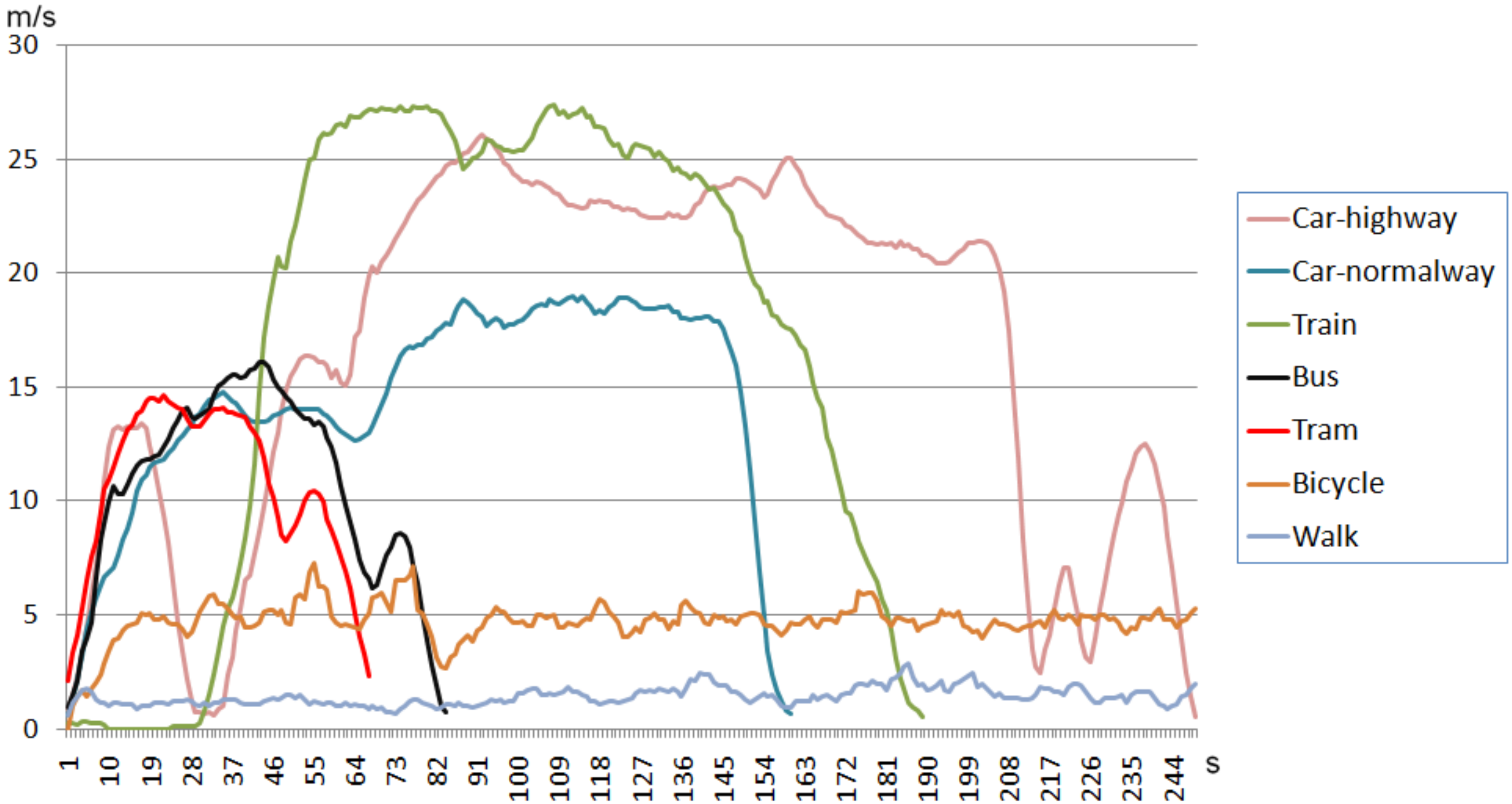
Speed

Turning angle

Service status

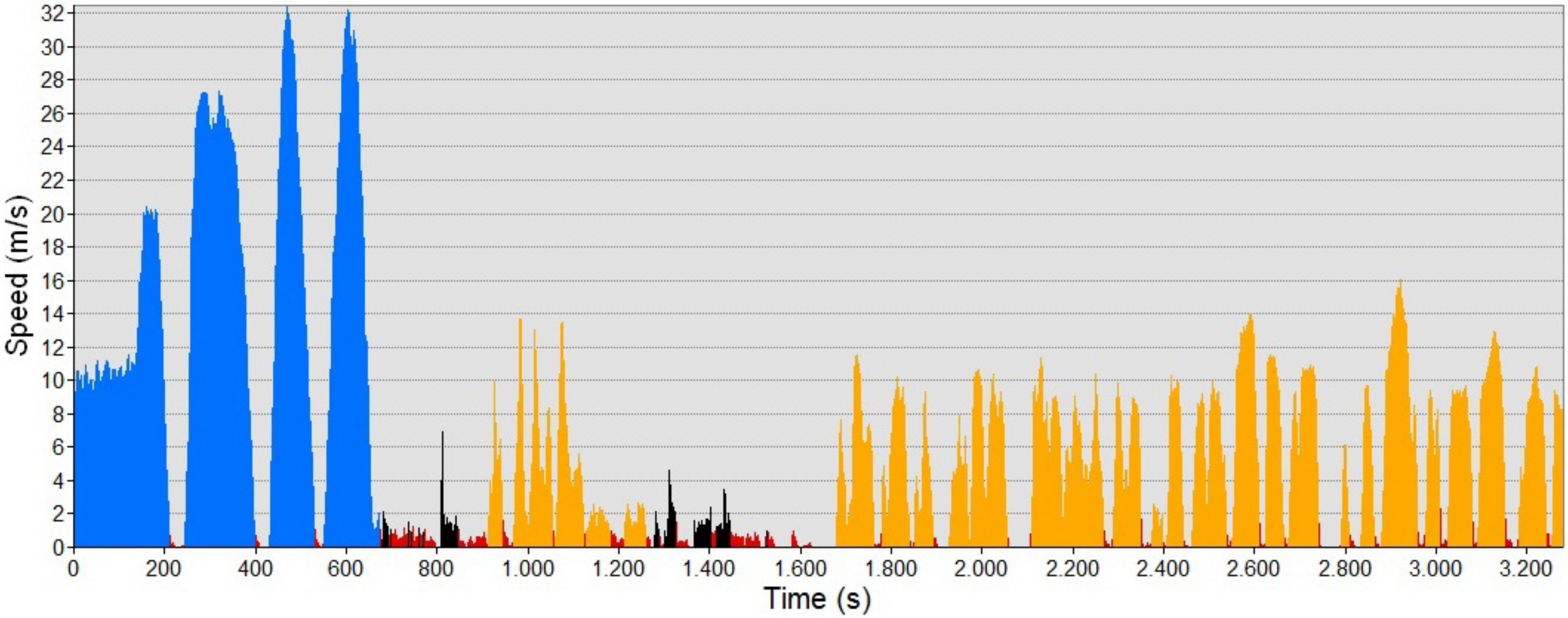


Trajectory dynamics of a taxi trajectory from 12:25 to 15:19 on 2010.4.1

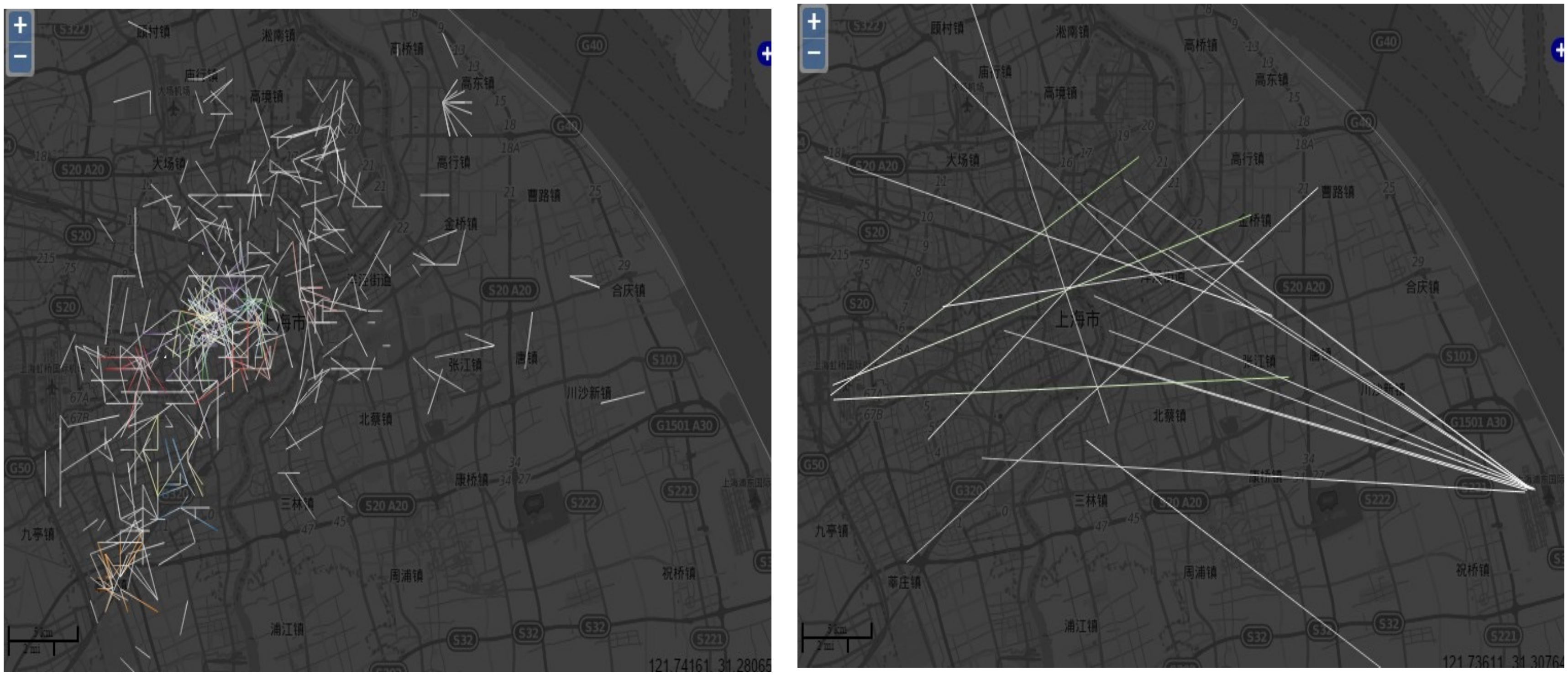


Speed patterns of different travel modalities

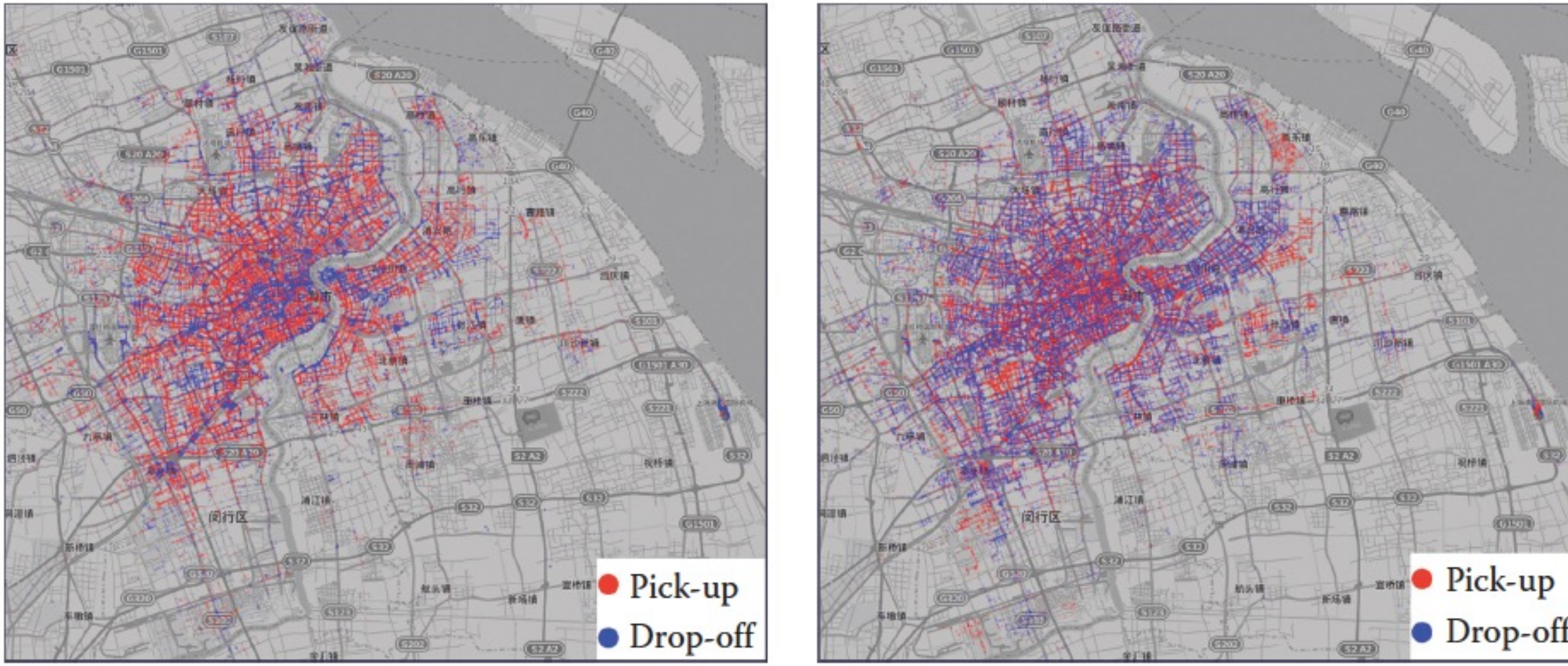
Recognition of road types from trajectories of multiple travel modalities



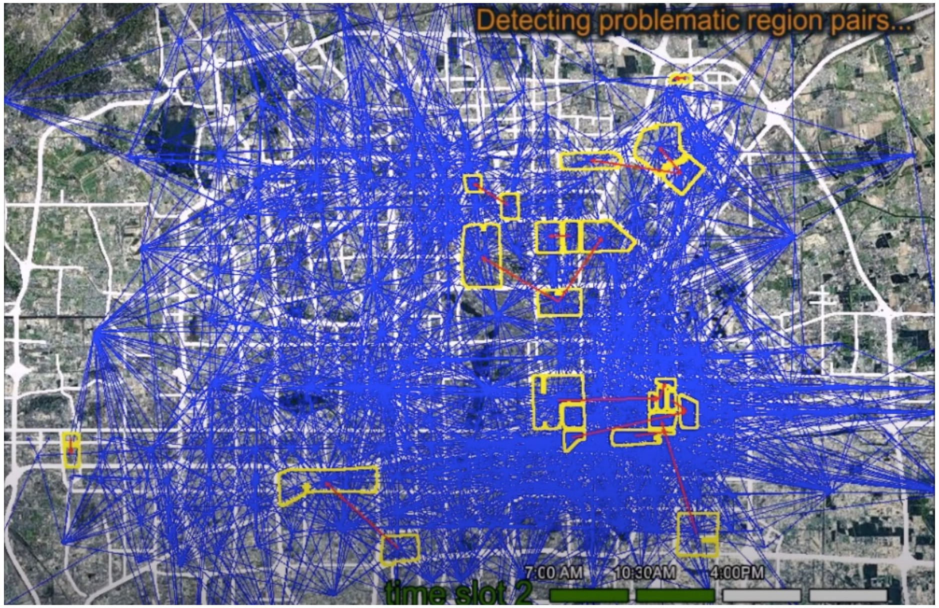
Recognition of travel modalities of a trajectory: train (blue), bus (yellow), walking (black), stops (red)

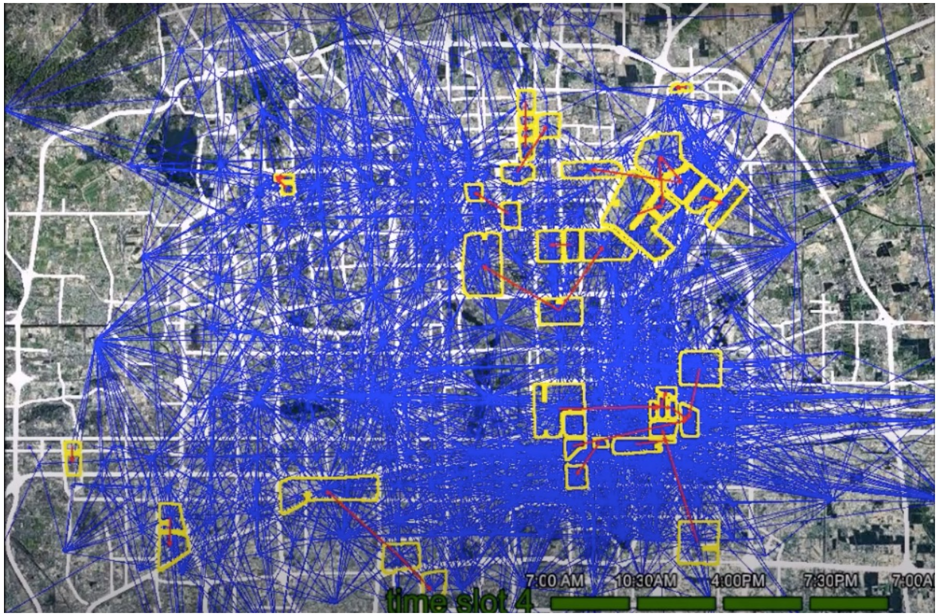


The origin-destination distribution of taxi travels of < 3km (left) and >30 km (right)

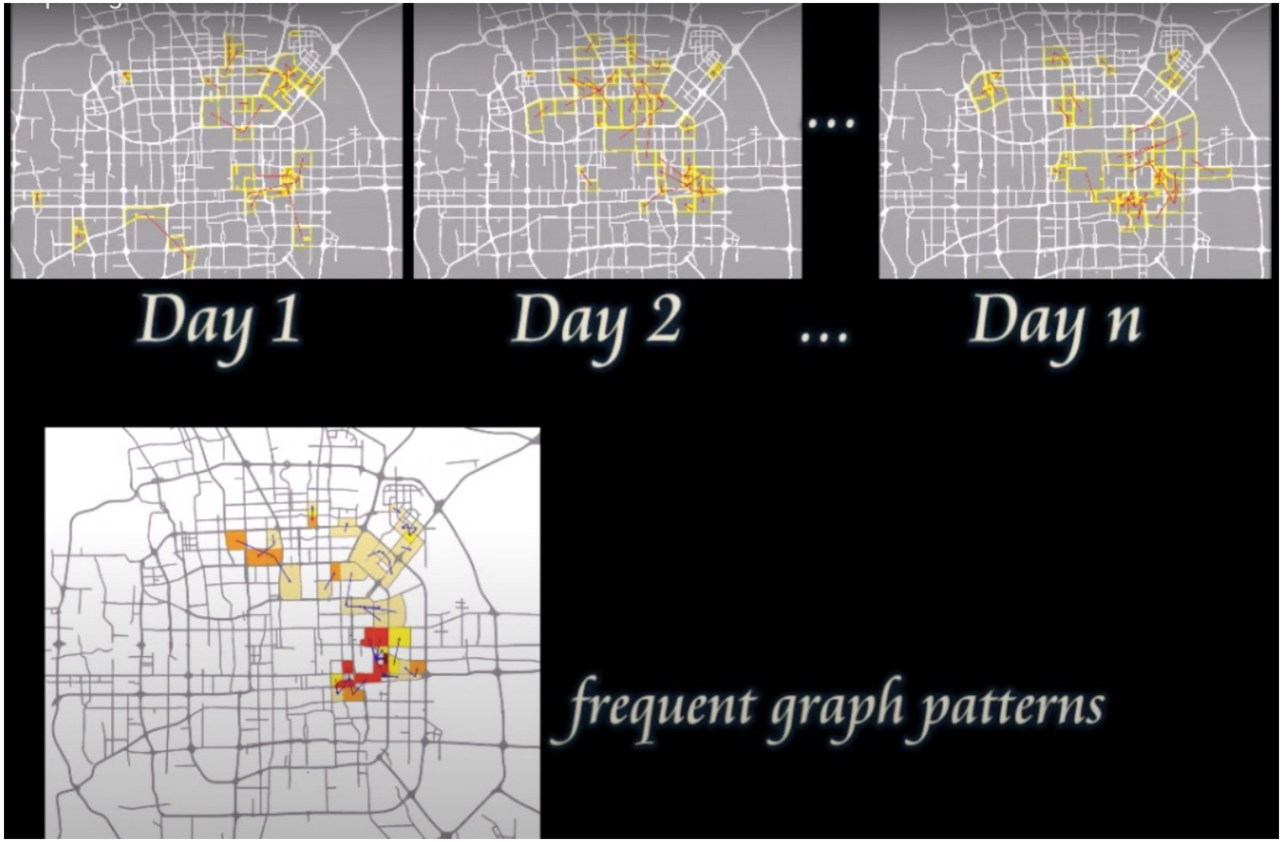


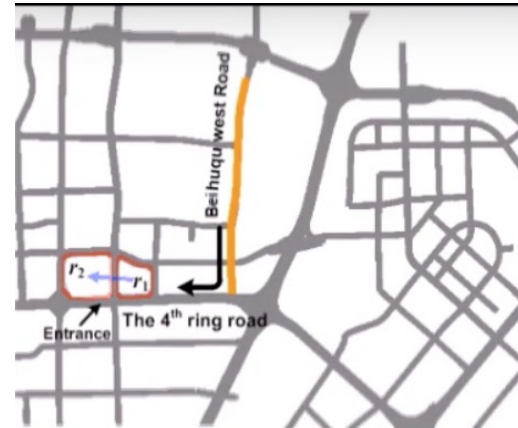
Spatial distribution of pick-up and drop-off points at 7-8 am (left) and at 6-7 pm (right)





Detection of problem pairs (e.g. entering jam) at different times during working days





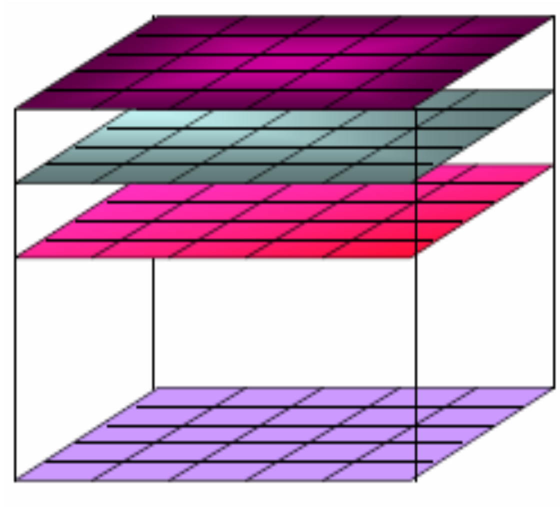
Find problem to trigger actions

Find evidence to justify actions

III.2 Vertical analysis

1. Map algebra

Analytical method that combines rasterized layers cell by cell using



Boolean operators

Where is both A and B

Where is A or B

Where is B but not A

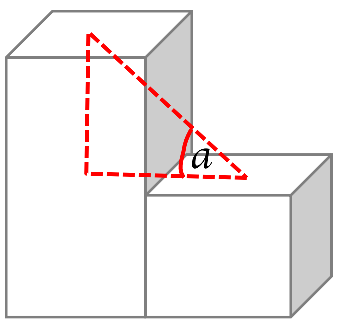
Where is neither A nor B

Algebraic operations (+,-, *, /, log, etc)

2. Slope and aspect

They are calculated by comparing the cell's elevation to that of its neighbors. The slope is the steepest descent or ascent at that point, its value is between 0° and 90°.

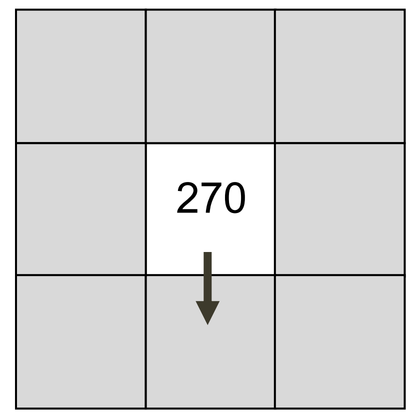
The aspect refers to the direction to which a slope faces. It takes a value between 0° and 360°



For each neighboring cell:
arctan (dz/dx) = arctan (dz/dy)

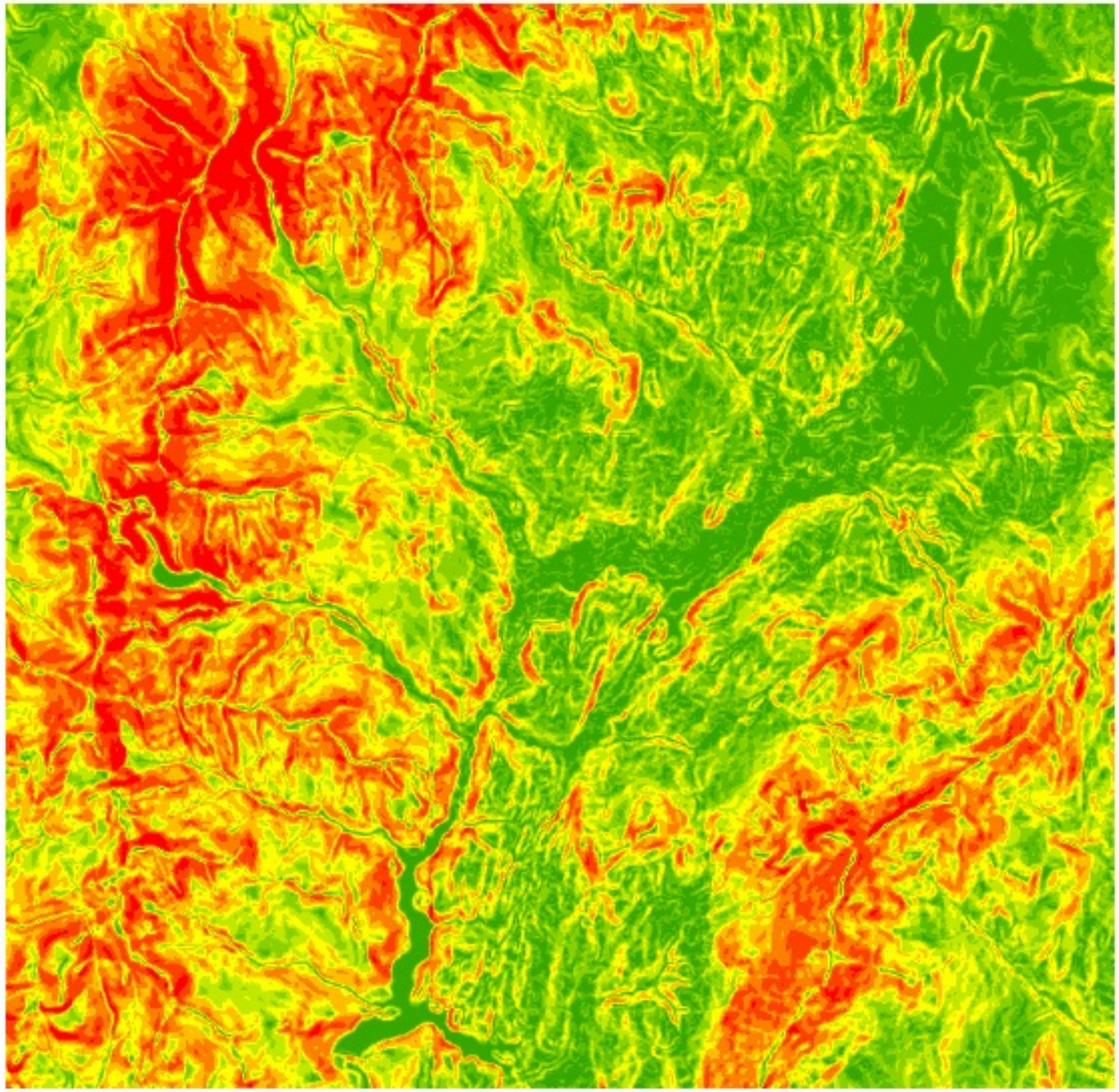
23°	23°	21°
25°	31	20°
20°	31°	12°

slope



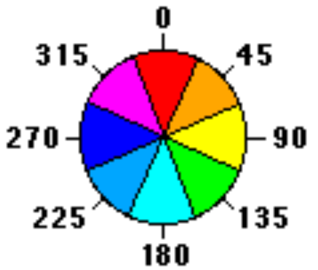
aspect

Slope
(degrees)

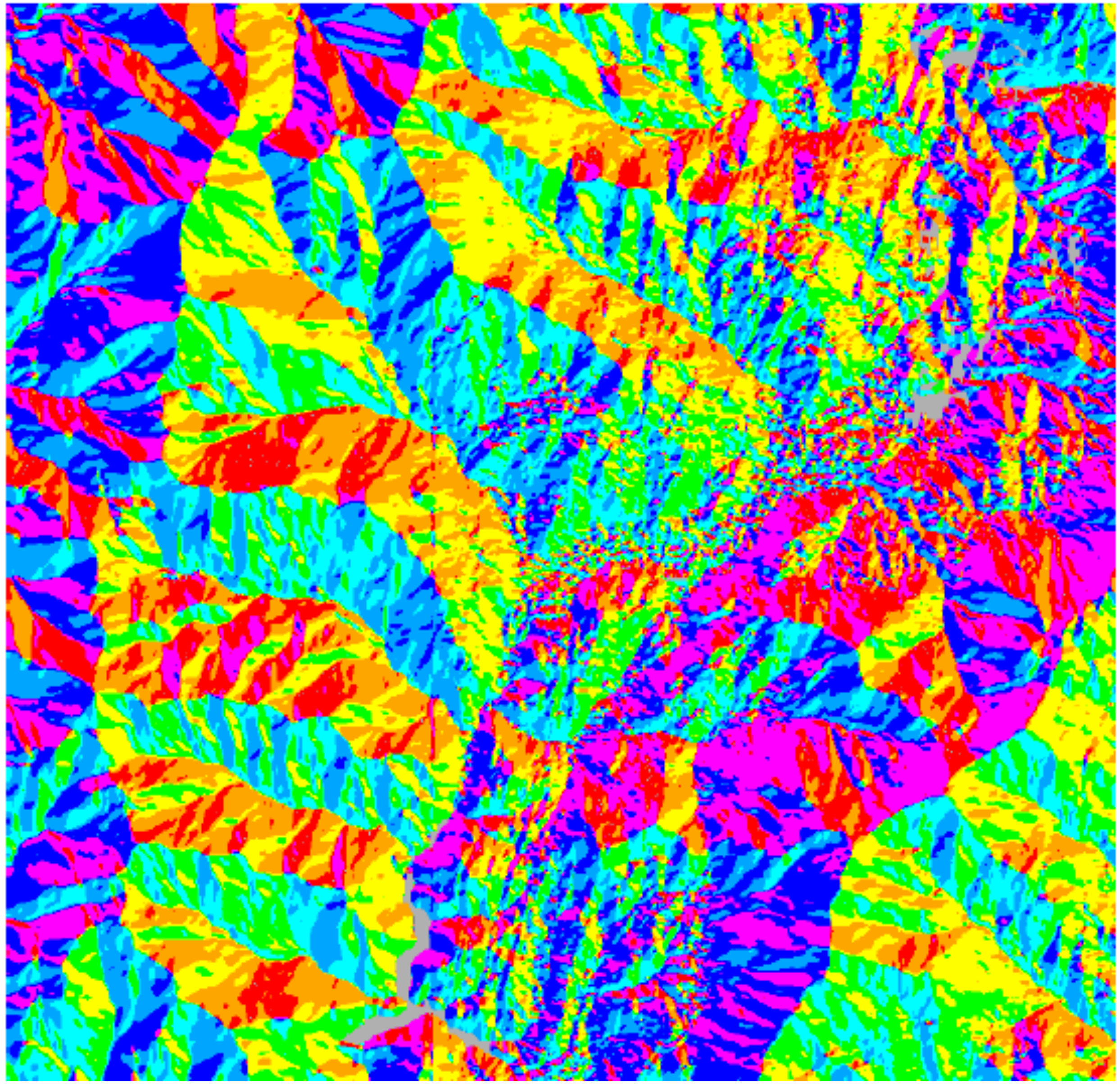


Slope of elevation
- 0 - 9.071943395
- 9.071943396 - 16.93429434
- 16.93429435 - 24.19184905
- 24.19184906 - 31.14700566
- 31.14700567 - 38.10216226
- 38.10216227 - 44.75492075
- 44.75492076 - 51.40767924
- 51.40767925 - 58.96763207
- 58.96763208 - 77.11151886

Aspect
(direction of slope)



Aspect of elevation
- Flat (-1)
- North (0-22.5)
- Northeast (22.5-67.5)
- East (67.5-112.5)
- Southeast (112.5-157.5)
- South (157.5-202.5)
- Southwest (202.5-247.5)
- West(247.5-292.5)
- Northwest (292.5-337.5)
- North (337.5-360)

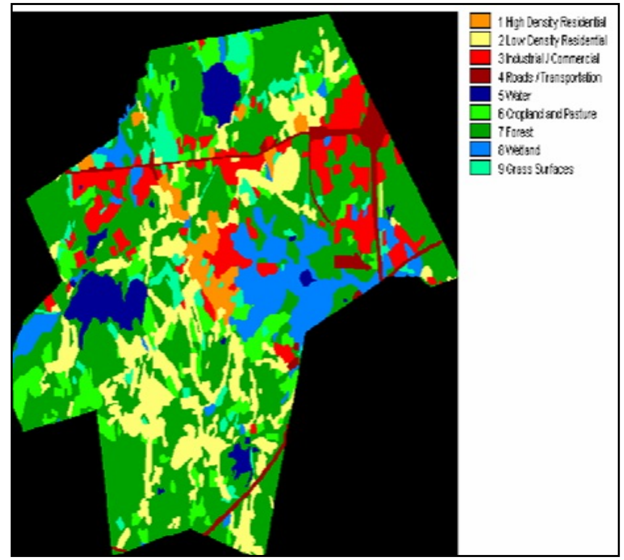


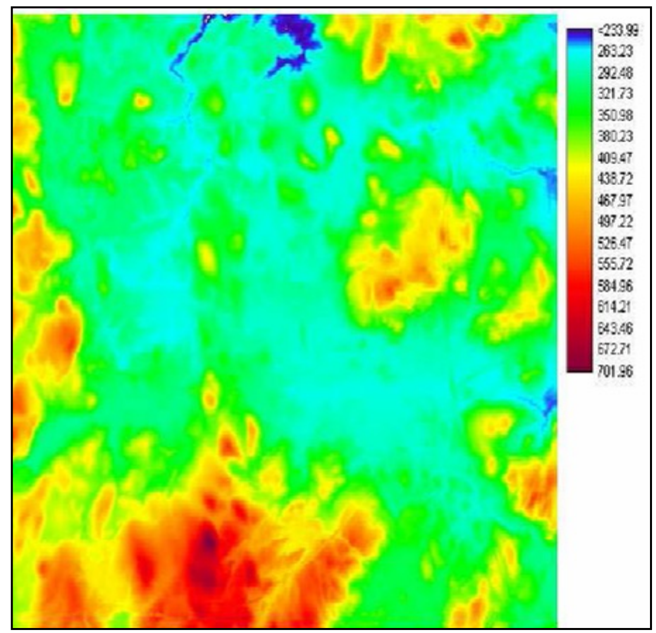
Example:

Find areas suitable for a new residential area considering the following conditions:

- the area should be located higher than 200m above sea level
- the slope should be less than 5°
- land use type should be forest

Data sources:
- Land use map
- Digital Elevation Model, DEM



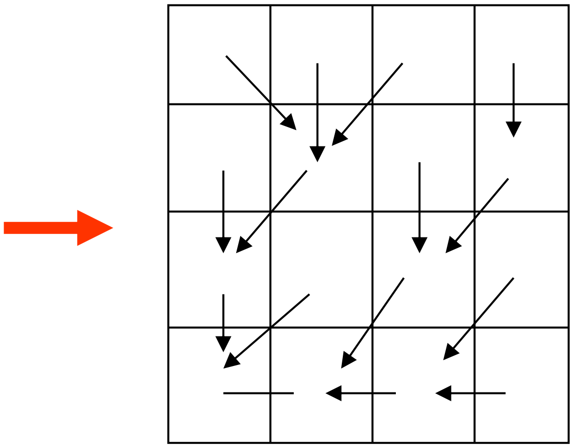


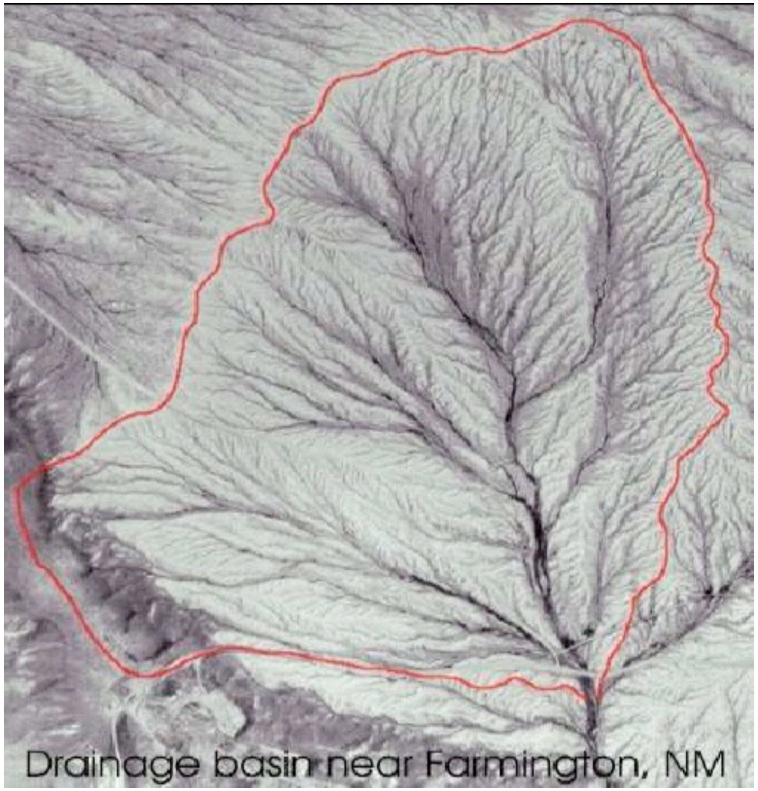
3. Flow analysis in hydrology

By estimating the flow distribution from all cells in a terrain model to its eight neighbors, the flow pattern over a drainage basin can be modeled. Traditionally, the water flows to one of eight neighbors. The local drain direction map can be used to calculate:

– wetness in the cell
– sediment transport through the cell
– stream channels and catchment areas

9	8	9	10
9	7	8	8
5	6	6	7
2	4	5	6





Examples of catchment areas / drainage basins

Flow analysis is based on simplified assumptions:

- From any point in the terrain, the water flows according to the *topographic form* of the cell and its eight neighbor cells;
- Water is evenly distributed over the grid cells;
- The infiltration capacity over the surface is set to zero;
- The surface is bare;
- The evapotranspiration is set to zero.

Estimation of flow accumulation (Pilesjö method**)**

1) Start by identifying cells with the highest height value

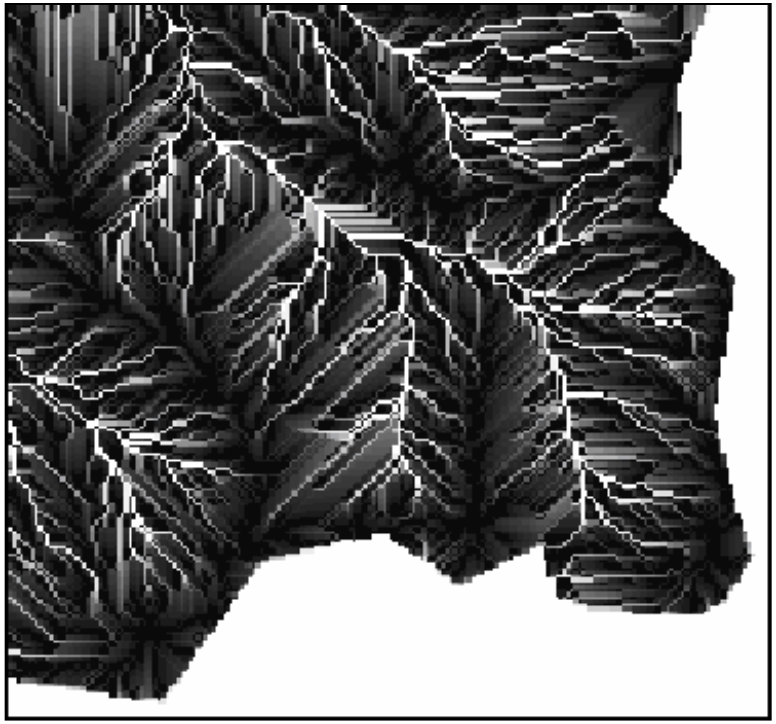
2) Compute the output flow from the cell

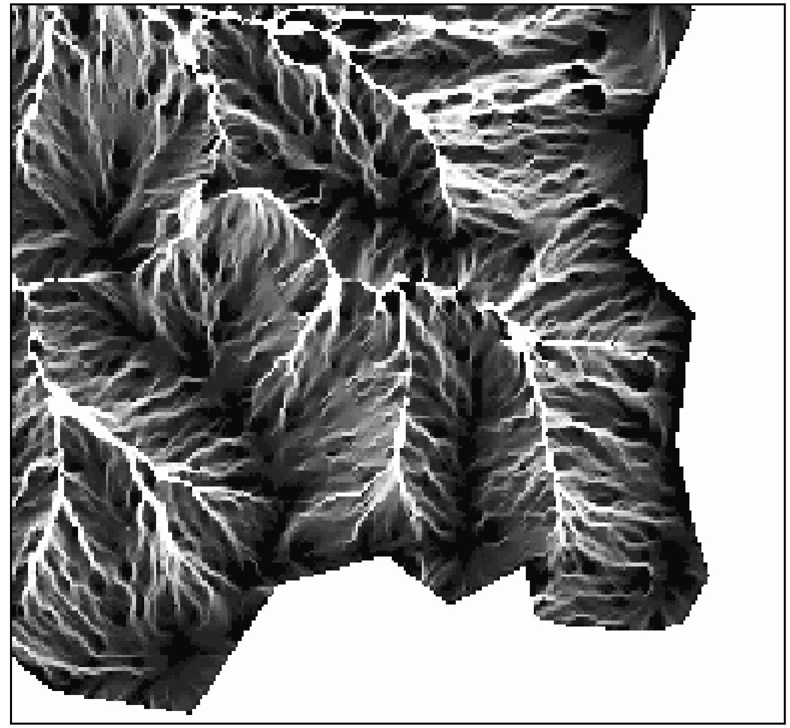
$$F_{out} = F_{in} + F_{local}$$

where F_{in} is the flow received by the cell, and F_{local} is the flow stored in the cell (e.g. precipitation).

3) Compute to which neighboring cells the output flow will go

4) Repeat 1-3 until all cells have been visited. Special treatment is needed for flat areas and sinks in terrain model.





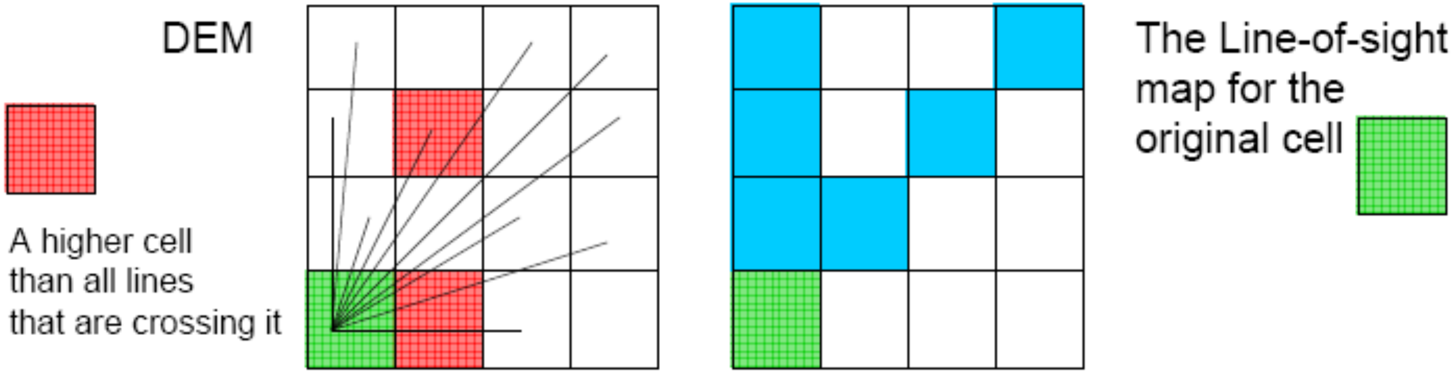
Left: flow accumulation estimated by the assumption that all flow to the lowest neighboring cell.
Right: flow accumulation estimated by Pilesjö method allows a more general water flow.

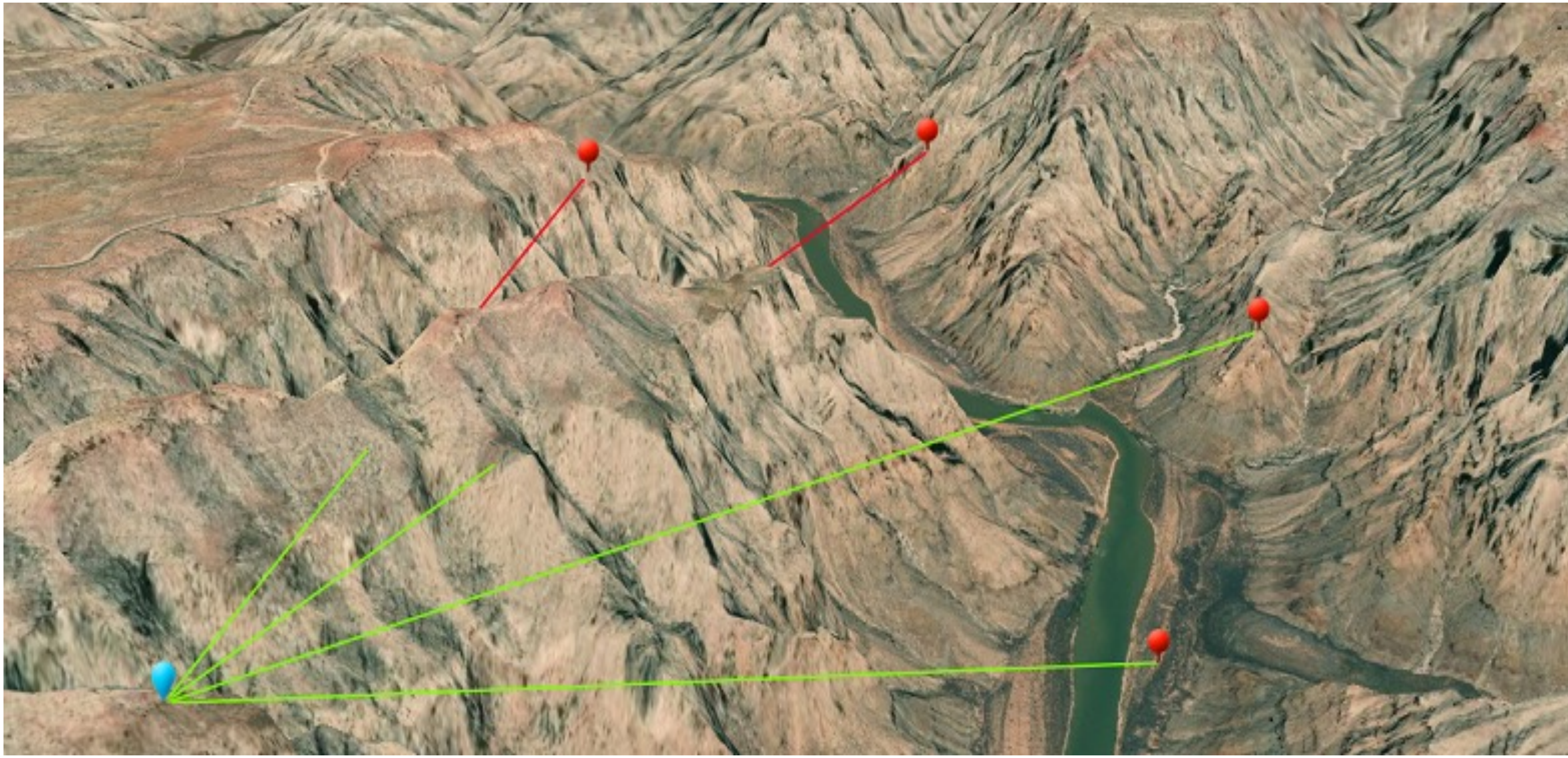
4. Line-of-sight

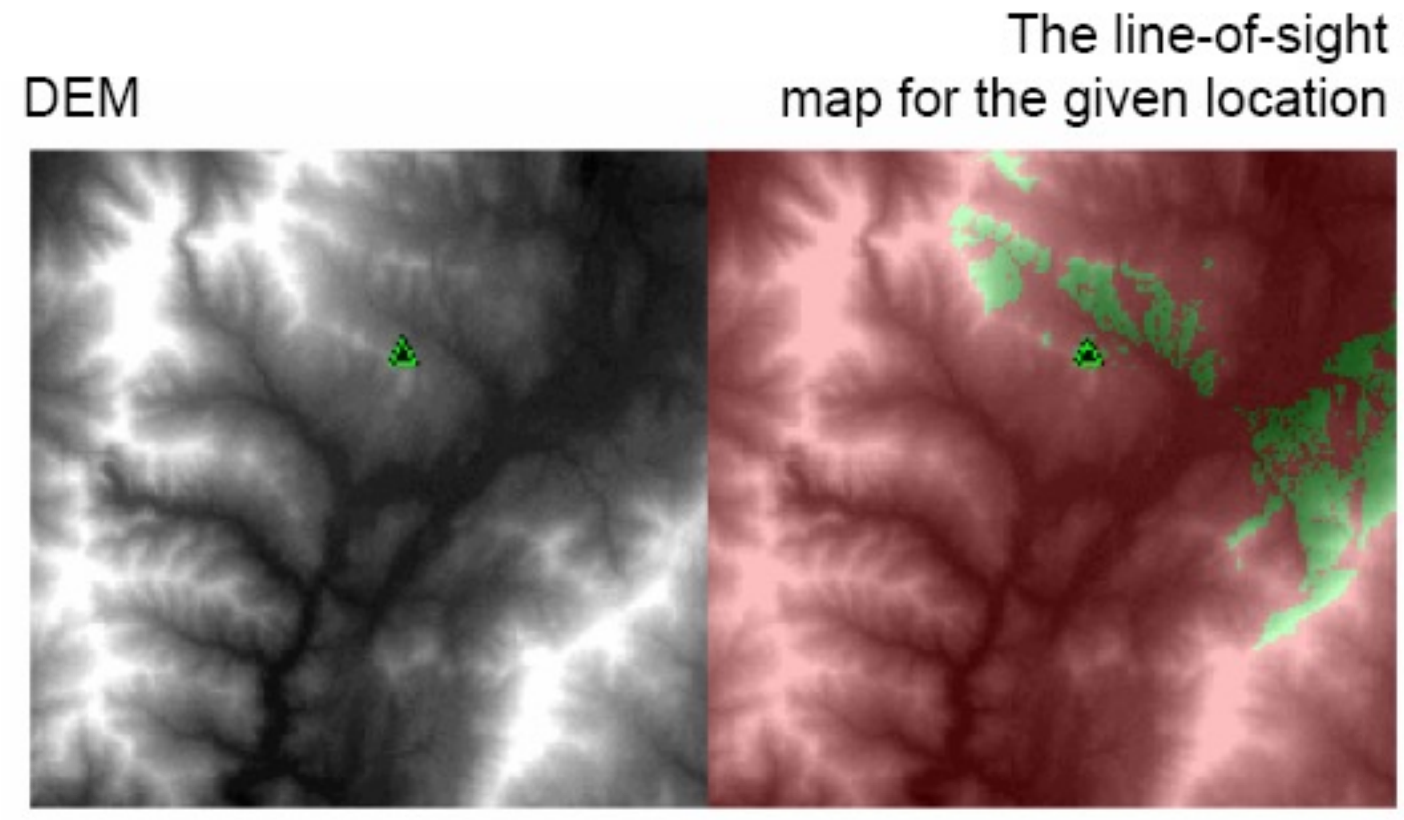
It answers the question of which parts of the landscape can be seen from a given original point.

A simple algorithm:

- draw a straight line between the given point (original cell) and each other cell in the DEM
- determine if the line crosses any cell that is higher than the line
- if yes, then you can not see the current cell from the original cell – set the current cell 0, otherwise set it 1.





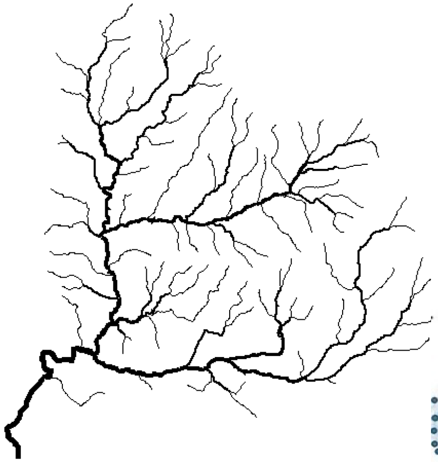


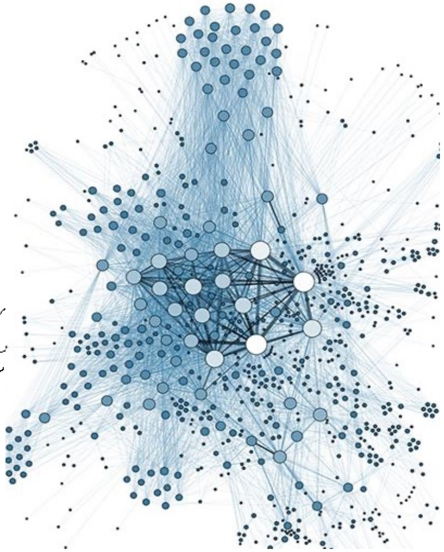
Applications:

- spreading of the traffic noise from the roads
- locating positions for the masts / antennas for the mobile telephony in a mountainous terrain

III.3 Network analysis

A network, or graph, has two main constitutes: the *nodes* and the *edges*. The network can represent for example a street network, river networks, or social network.

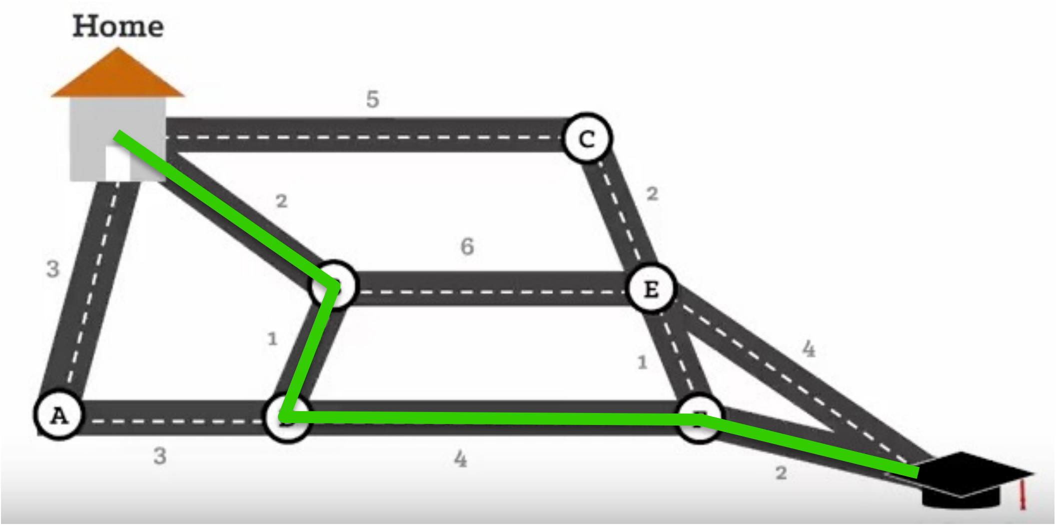




Network analysis considers primarily the connections and distances. The geometric locations of the nodes are less interesting.

1. Dijkstra's algorithm

It is used to find the shortest path using *adjacency matrix* and *state matrix*. The *adjacency matrix* stores the direct distances between the vertexes. The *state matrix* changes during the algorithm execution.



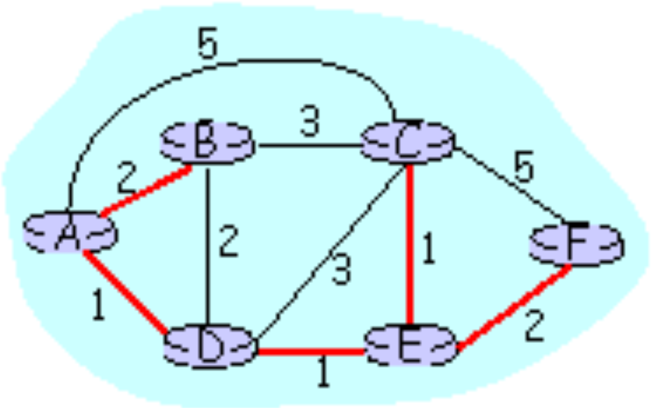
https://brilliant.org/wiki/dijkstras-short-path-finder/

The *state matrix* keeps for each vertex *V* the cost *d*[*V*] of the shortest path found so far from the starting node.

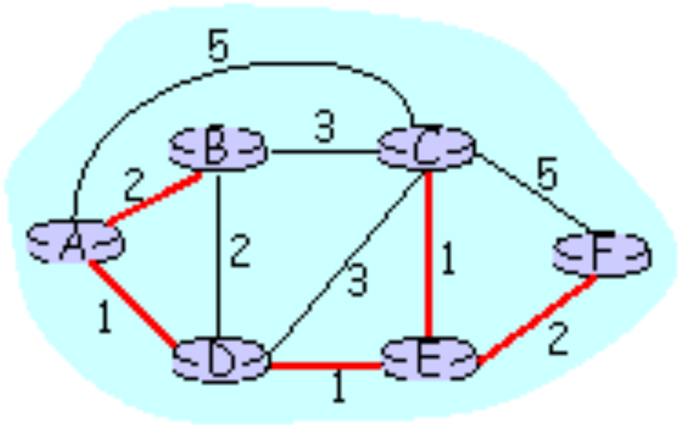
Initially, this value is 0 for the starting vertex *A* (*d*[*A*]=0), the value is ∞ between *A* and vertex *V* if there is no direct path (*d*[*V*]=∞).

The algorithm searches stepwise among the connected vertexes till the destination vertex is reached. When the algorithm finishes, *d*[*V*] will be the cost of the shortest path from *A* to *V*.

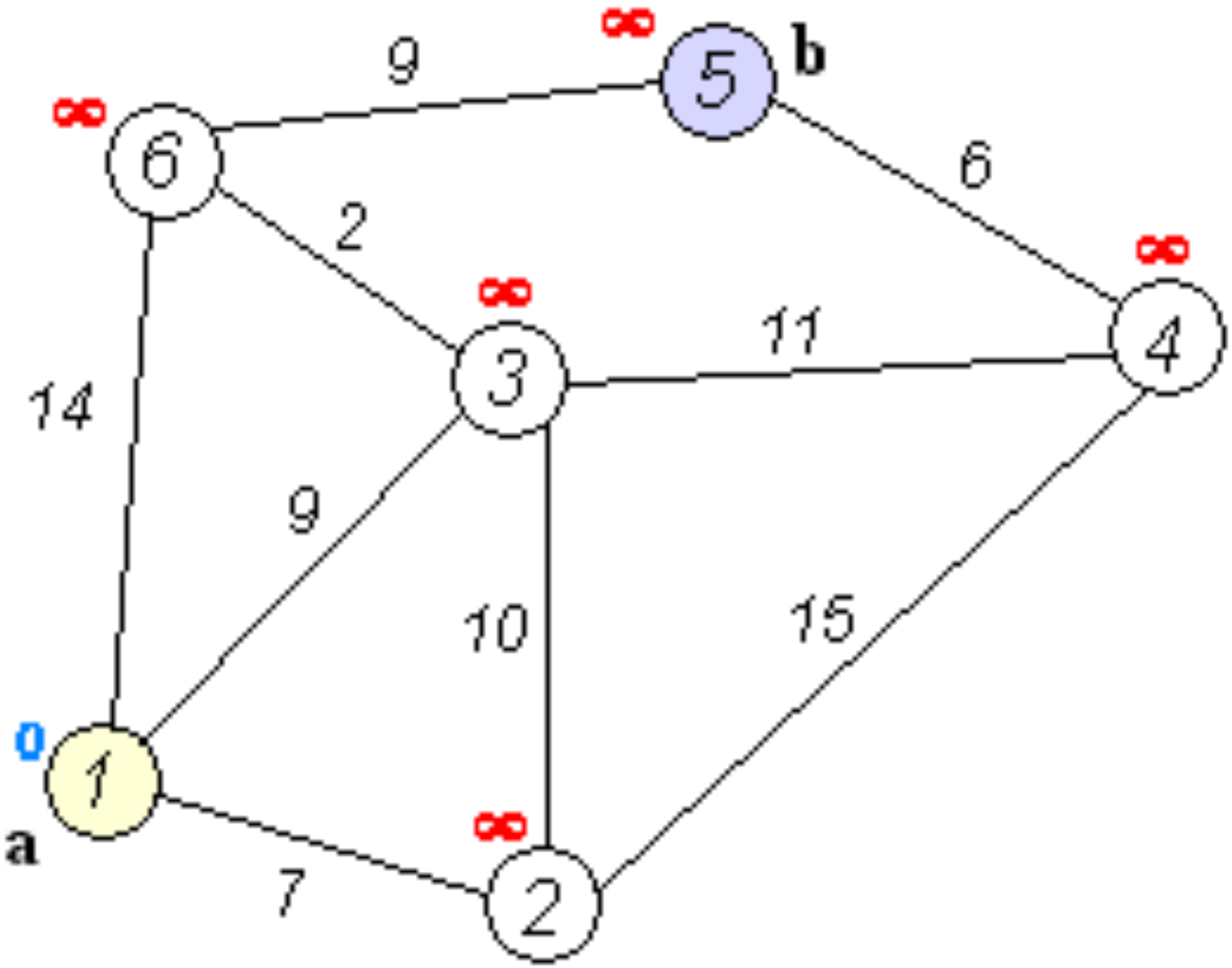
Step	start N	d(B)	d(C)	d(D)	d(E)	d(F)
0	A	2,A	5,A	1,A	infinity	infinity

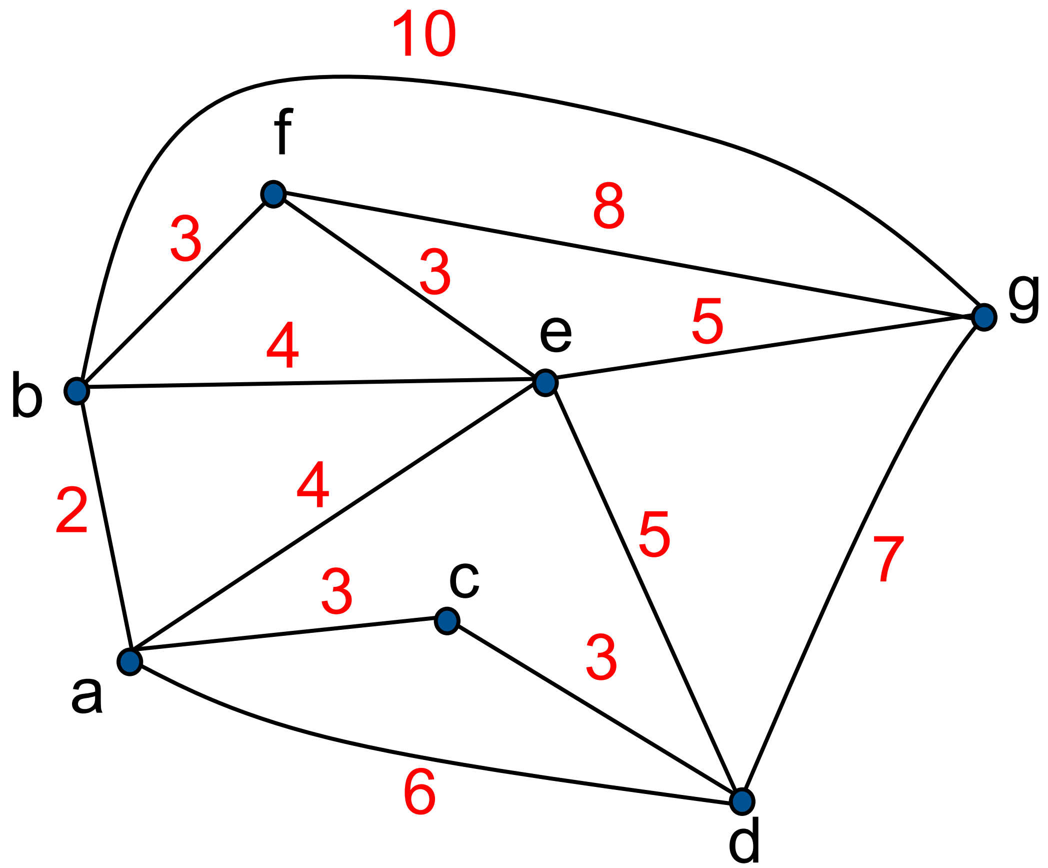


Step	start N	d(B)	d(C)	d(D)	d(E)	d(F)
0	A	2,A	5,A	1,A	infinity	infinity
1	AD	2,A	4,D		2,D	infinity
2	ADE	2,A	3,E			4,E
3	ADEB		3,E			4,E
4	ADEBC					4,E
5	ADEBCF					



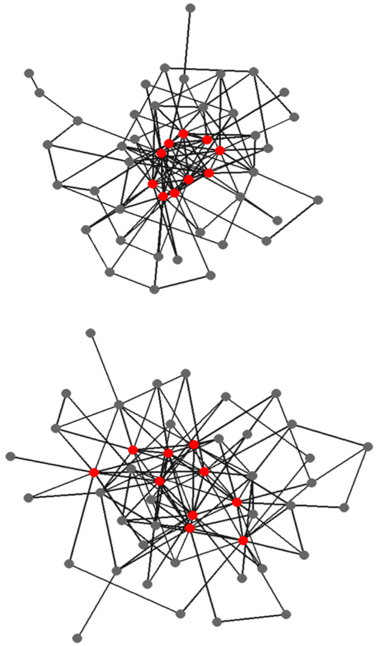
From each step, the shortest distance will be taken and the corresponding node will be marked as visited.

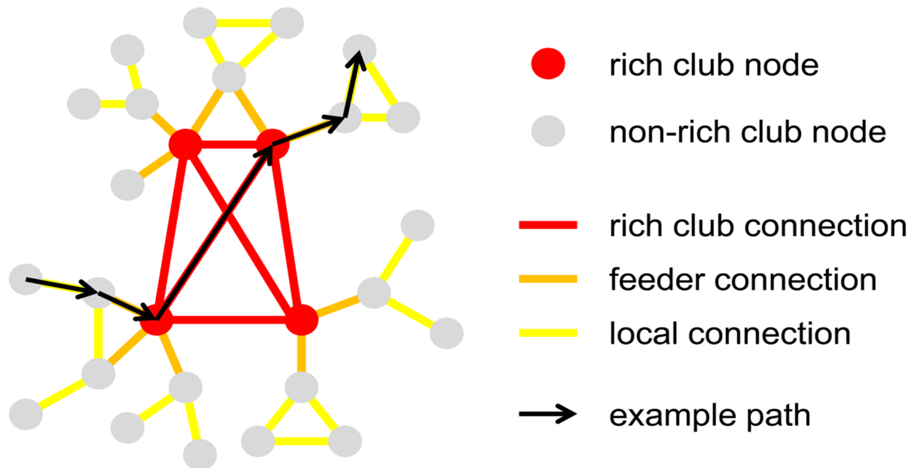

∞
∞
∞
∞
∞
0
9
5
b
6
2
6
4
11
3
14
9
10
15
1
a
7
2



Please find the shortest road from *a* to *g*

2. Rich-club phenomenon





- Nodes are ranked in terms of node degree
- Edges are ranked in terms of traffic volume, distance etc.
- A rich club exists in a network, if there is a set of nodes that are rich in connections and are more densely interconnected among themselves than lower-degree nodes in the network.

Indicator of richness

Rich-club coefficient

the number of edges between the nodes which have a degree greater than or equal to *k*

$$\phi(k) = \frac{2E_{>k}}{N_{>k}(N_{>k}-1)}$$

the number of nodes with degree greater than or equal to *k*

$$\rho_{rand}(k) = \frac{\phi(k)}{\phi_{rand}(k)}$$

Rich-club coefficient of a randomized network, e.g. with the same node degree distribution

Rich-club exists only when $\rho_{rand}(k) > 1$

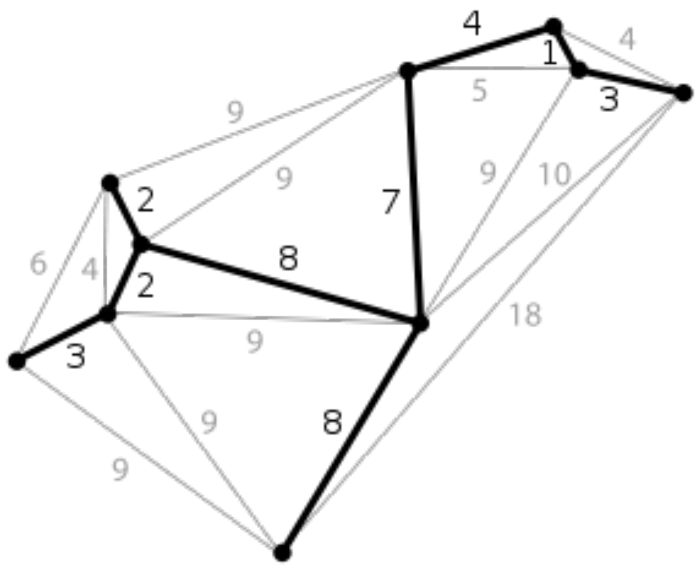
3. Minimum Spanning Tree (MST)

A connected network with minimized total sum of edges

1. Find the shortest edge in the network. If there are more than one, pick any one randomly. Highlight this edge and the nodes connected.

2. Pick the next shortest edge, unless it forms a cycle with the edge already treated. Highlight this new edge and its nodes.

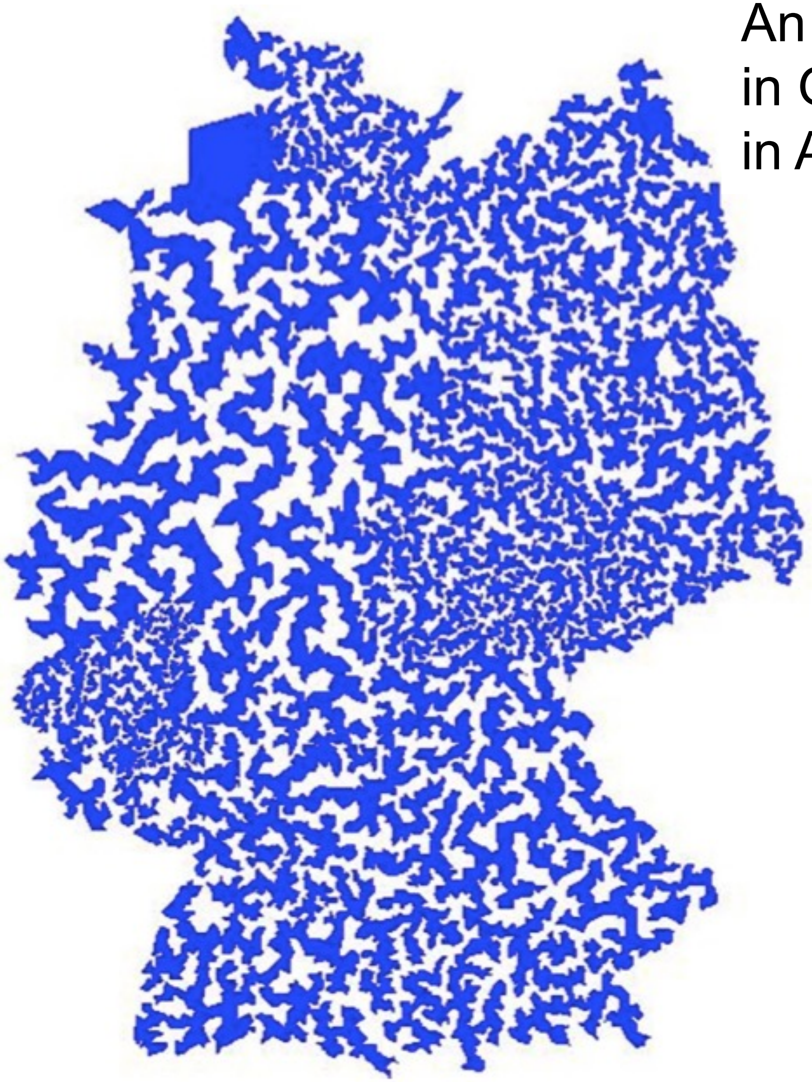
3. If all edges are connected, they are are done. Otherwise, repeat Step 2.



Applications of MST:

Solve Traveling Salesman Problem by finding the shortest route through all the nodes of the network so that each node is visited exactly once.

The maximum length of the traveling salesman problem is at most 2 times the total length of the MST.



An optimal tour through 15,112 cities in Germany was proved possible in April 2001.

4. Cost surface & least cost path

How much does it cost to cross each raster cell?
What is the minimum cost to reach each raster cell from the starting point?

3	5	6	6
6	6	2	2
4	4	4	2
2	6	3	4

Cost surface

0	0	0	0
0	0	0	0
0	●	0	0
0	0	0	0

Starting point

+

3	5	6	6
6	6	2	2
4	●	4	2
2	6	3	4

Cost surface

→

13	11	12	14
10	6	6	8
4	●	4	6
6	6	7	10

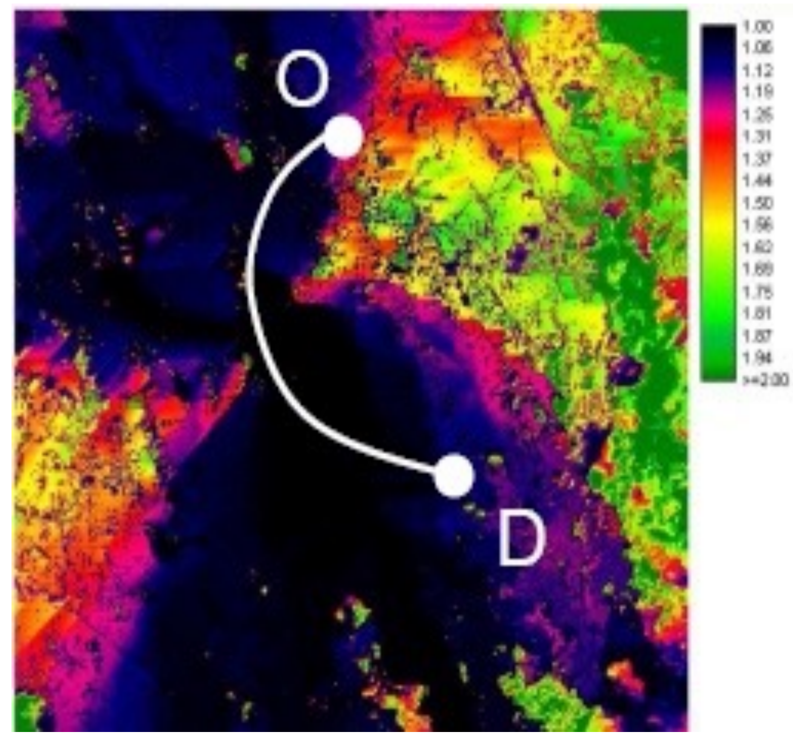
Accumulated cost surface

4 + 6 + 3 = 13
6 + 6 + 3 = 15
6 + 5 + 3 = 14

minimum = 13

What is the minimum cost to reach each raster cell from the starting point?

Example of least cost path: the cheapest path across a continuous cost surface between two points O and D (new road, power line, pipeline, etc.)



Cost can be:

- construction cost,
- land acquisition,
- environment impact,
- operating cost.

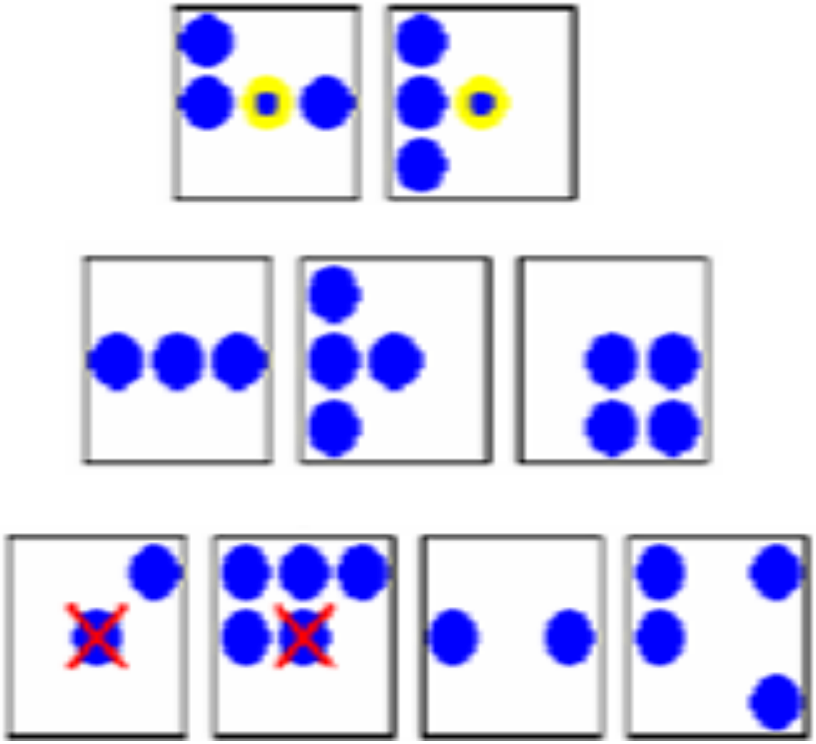
5. Cellular automata

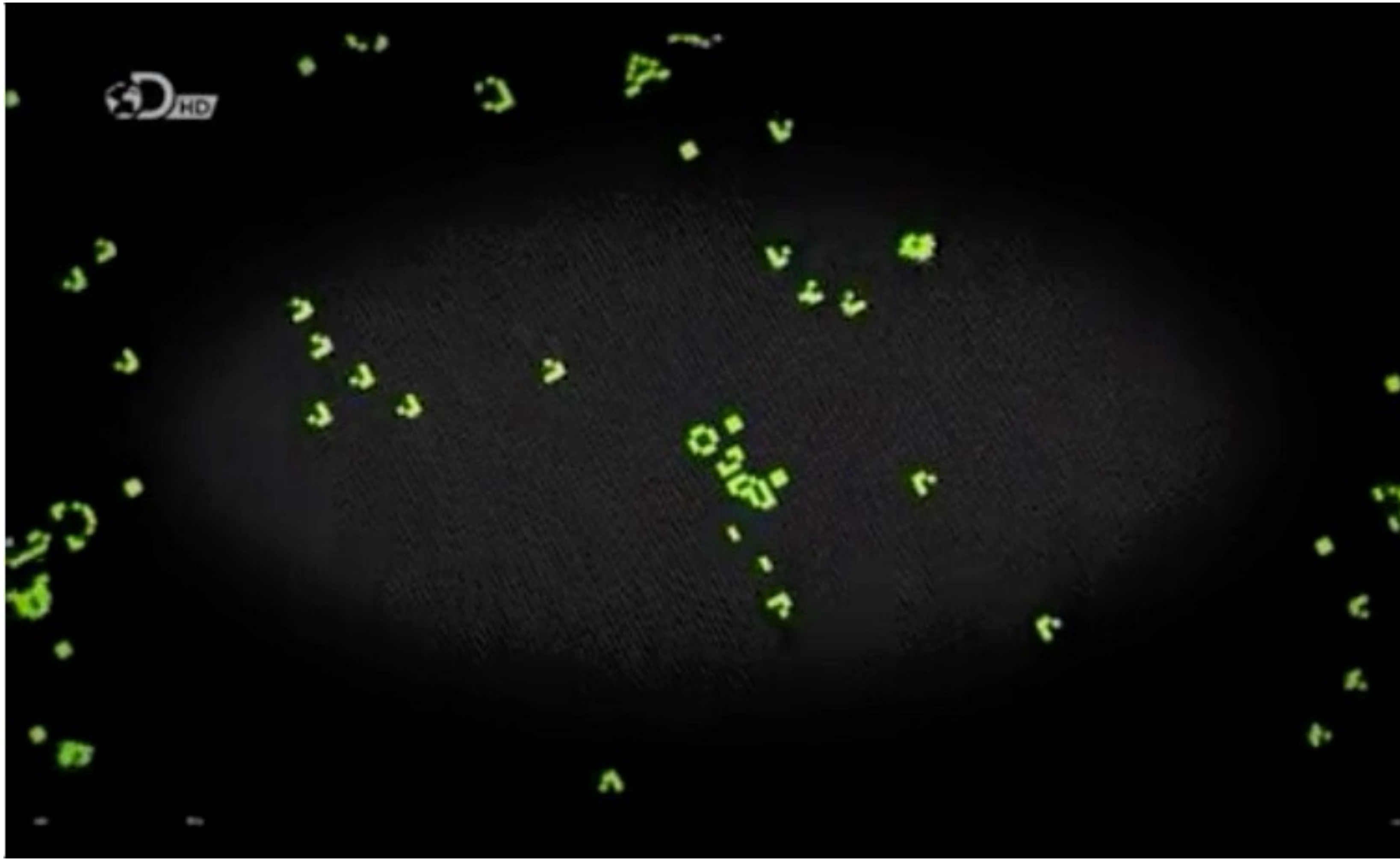
Active simulations of the behavior of a network system: model a system using a raster

- each cell can be in one of a number of states
- the state of each cell changes with time
- change through time is represented by change of cell state
- change is defined by a series of rules, depending on the state of the cell and its neighbors

An example: the life is played on a matrix. A cell can be live or dead. A live cell is shown by putting a marker on its square. A dead cell is shown by leaving the square empty. Each cell has 8 neighbors.

- A dead cell with exactly three live neighbors becomes a live cell.
- A live cell with two or three live neighbors stays alive
- In all other cases, a cell dies or remains dead (overcrowding or loneliness)

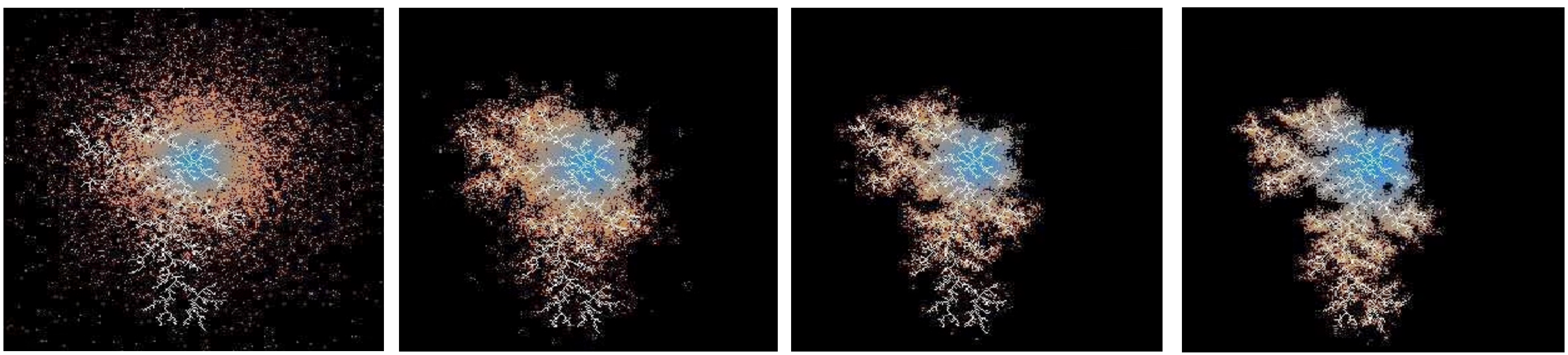




www.youtube.com/watch?v=CgOcEZinQ2I

Applications of cellular automata:

- simulation of the movement of pedestrians / vehicles on the streets
- modeling the urban-growth of a city around a street network



1. How do you understand the variogram?
2. What are basic attributes and derived attributes of GNSS points?
3. What can you discover from trajectories?
4. What can you derive from origin-destination pairs of GNSS points?
5. What is map algebra? Which operations do you know?
6. What are slope and aspect?
7. What is flow analysis in hydrology? Explain the working principle of a typical algorithm for flow analysis.
8. What is line-of-sight? And where can it be applied?
9. How does the Dijkstra’s algorithm work?
10. How do you understand the rich-club phenomenon?
11. What is a minimum spanning tree?
12. What is a cost surface?
13. What is a cellular automata and where can it be applied?

Additional references

- Cao Y, J Yuan, S Xiao & Q Xie, 2019: TPM: A GPS-based Trajectory Pattern Mining System. arxiv.org/pdf/1907.02678.pdf
- Ding L 2016: Visual Analysis of Large Floating Car Data - A Bridge-Maker between Thematic Mapping and Scientific Visualization. PhD thesis, TU Munich
- Zhang J, F Chen, Z Wang et al, 2017: Spatiotemporal Patterns of Carbon Emissions and Taxi Travel Using GPS Data in Beijing, Energies 2018, 11, 500; doi:10.3390/en11030500
- Zhang L, 2014: Mining GPS-Trajectory Data for Map Refinement and Behavior Detection. PhD thesis, University of Hannover
- Zhang W, X Wang and Z Huang 2019: A System of Mining Semantic Trajectory Patterns from GPS Data of Real Users. Symmetry 2019, 11, 889; doi:10.3390/sym11070889
- Zheng Y, 2018: Urban Computing (Information Systems). The MIT Press