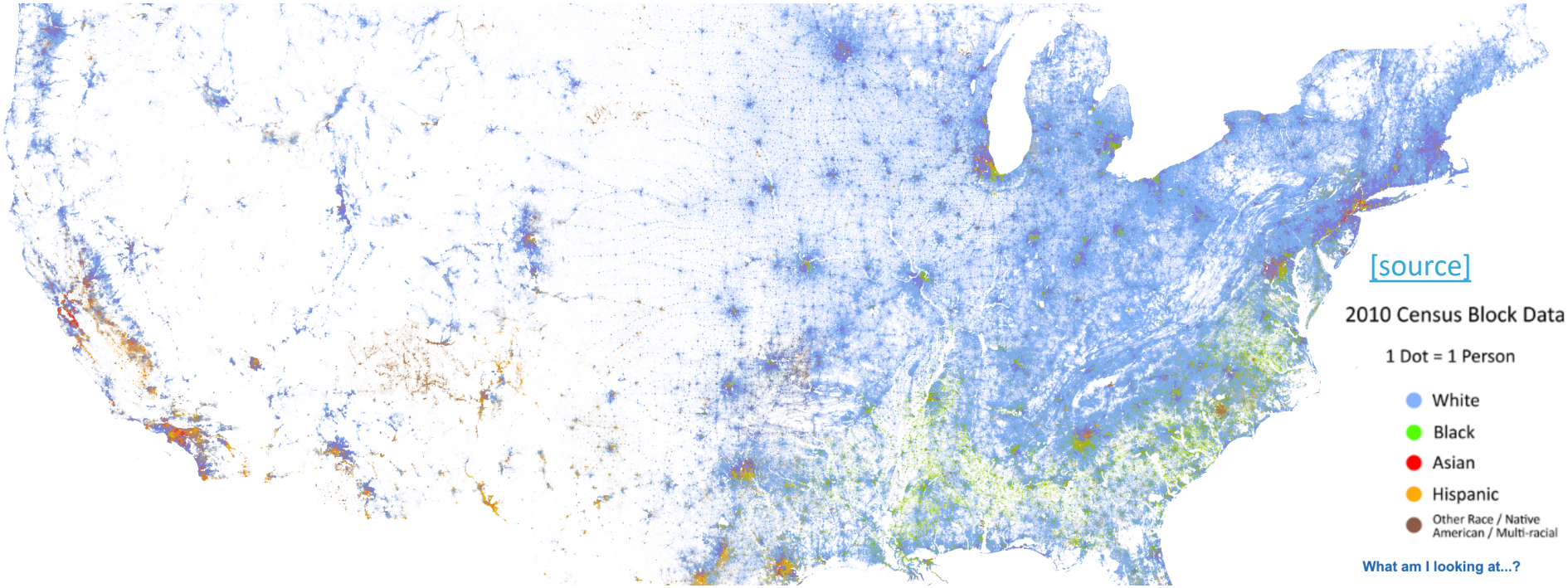
Introduction to *Urban* Data Science

Networks & Space

(EPA1316)

Lecture 7

Trivik Verma



Groups (28/09-16/10)

Form groups of ***4 students*** each. Each student needs to be part of a group. If you are going to drop the course, don't sign up for the final project as that may delay the progress of other students.

Last Time

- Geo-Visualisation
- Dangers of Geo-Vis
- Mapping Data
  - MAUP
  - Choropleths

Today

- Introduction to Networks
- The need to represent space formally
- Spatial weights matrices
  - What
  - Why
  - Types
- The spatial lag

Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

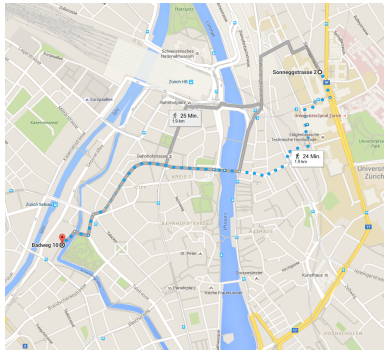
- Which facility is closest?

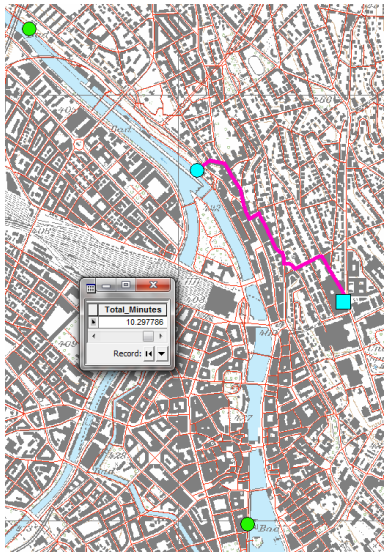
Vehicle-routing-problems

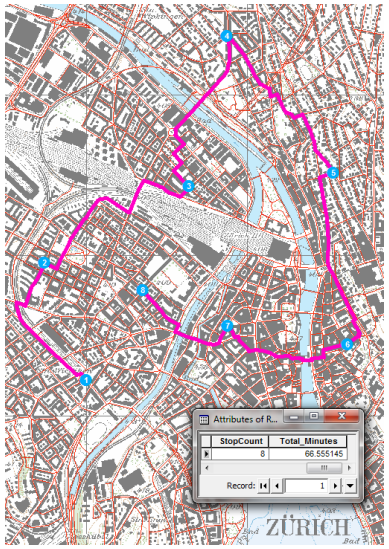
- Which sequence of stops is most effective?

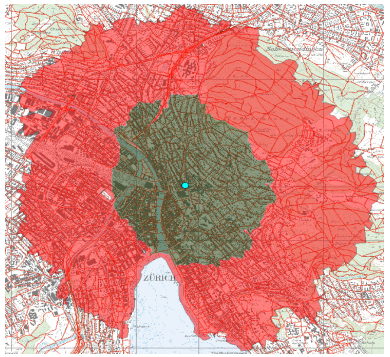
Accessibility of Zones

- What is the accessible area in a given time?









Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

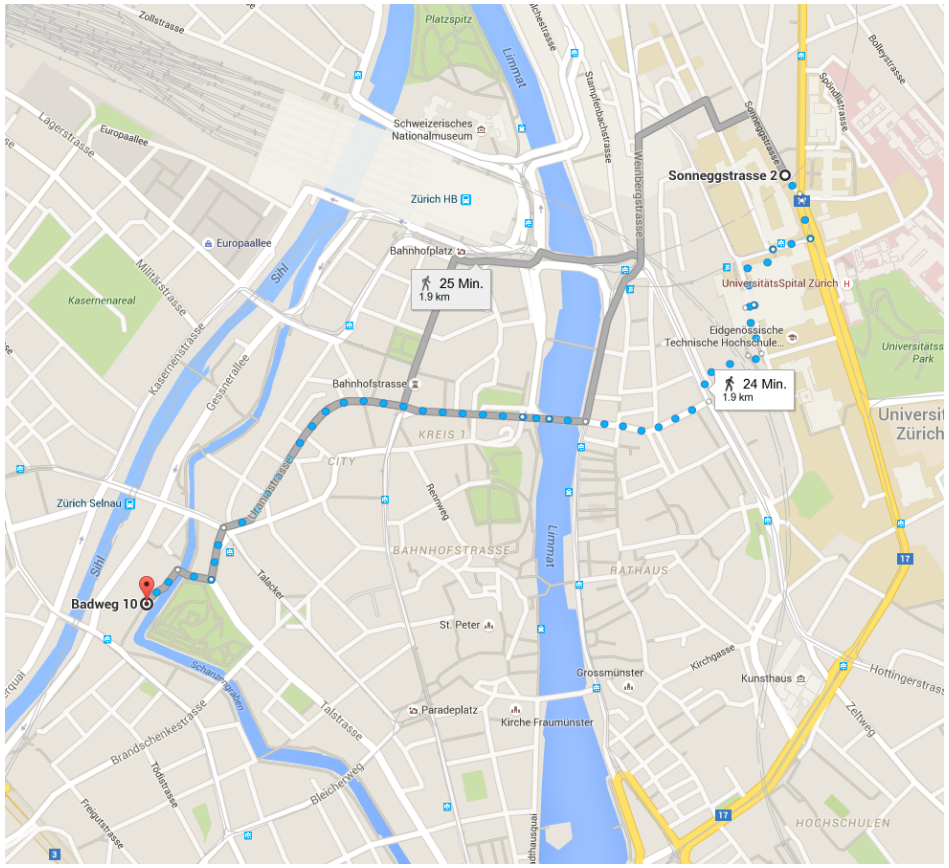
- Which facility is closest?

Vehicle-routing-problems

- Which sequence of stops is most effective?

Accessibility of Zones

- What is the accessible area in a given time?



Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

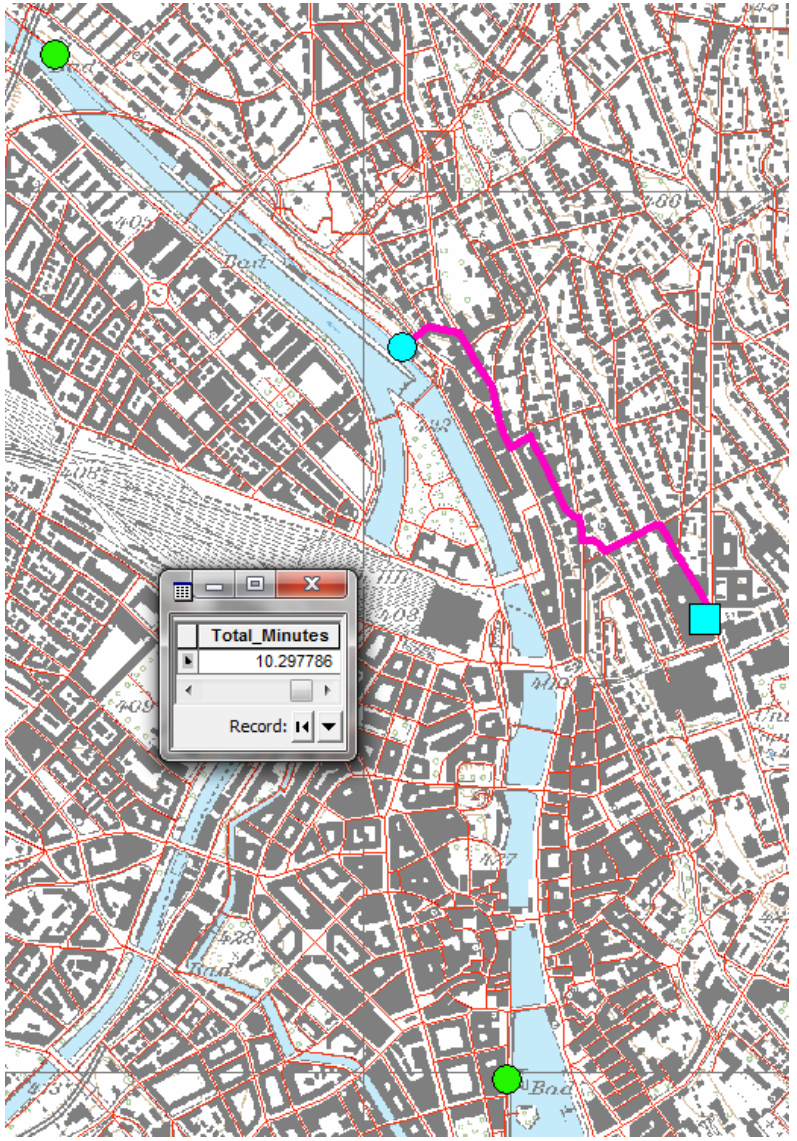
- Which facility is closest?

Vehicle-routing-problems

- Which sequence of stops is most effective?

Accessibility of Zones

- What is the accessible area in a given time?



Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

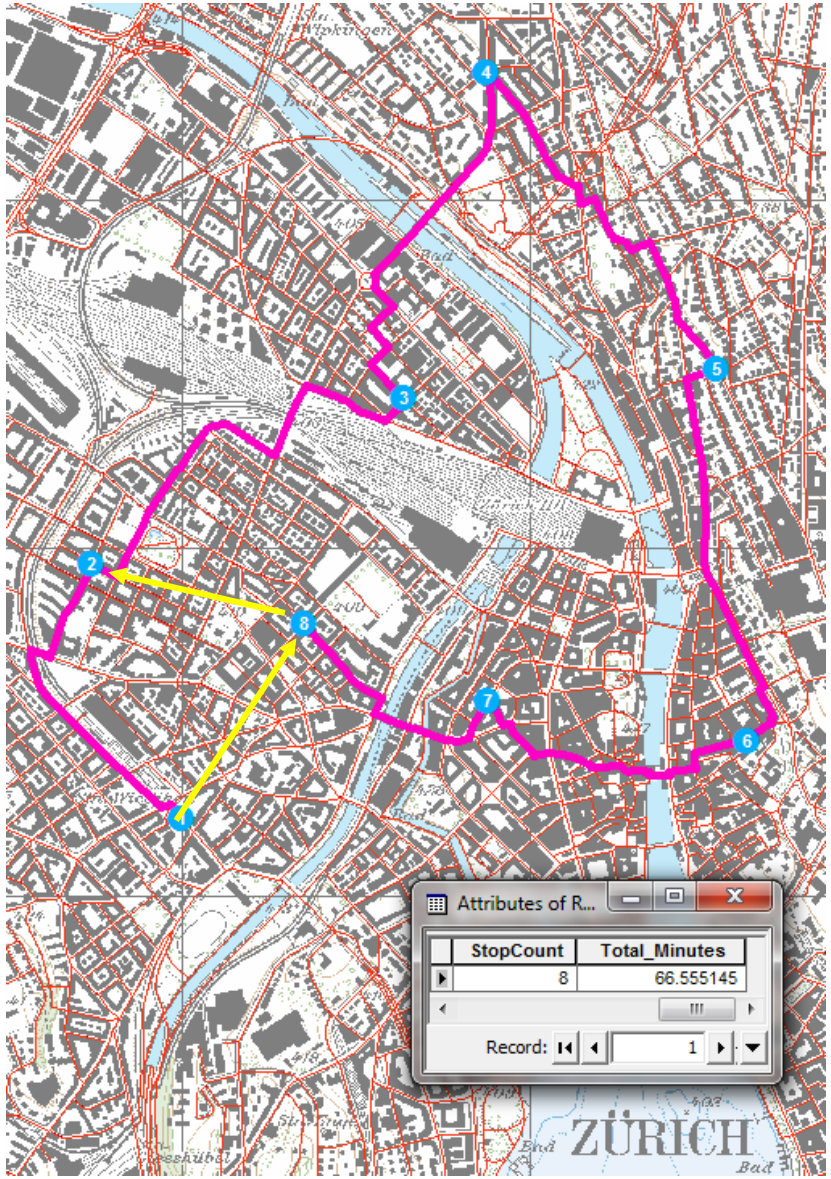
- Which facility is closest?

Vehicle-routing-problems

- Which sequence of stops is most effective?

Accessibility of Zones

- What is the accessible area in a given time?



Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

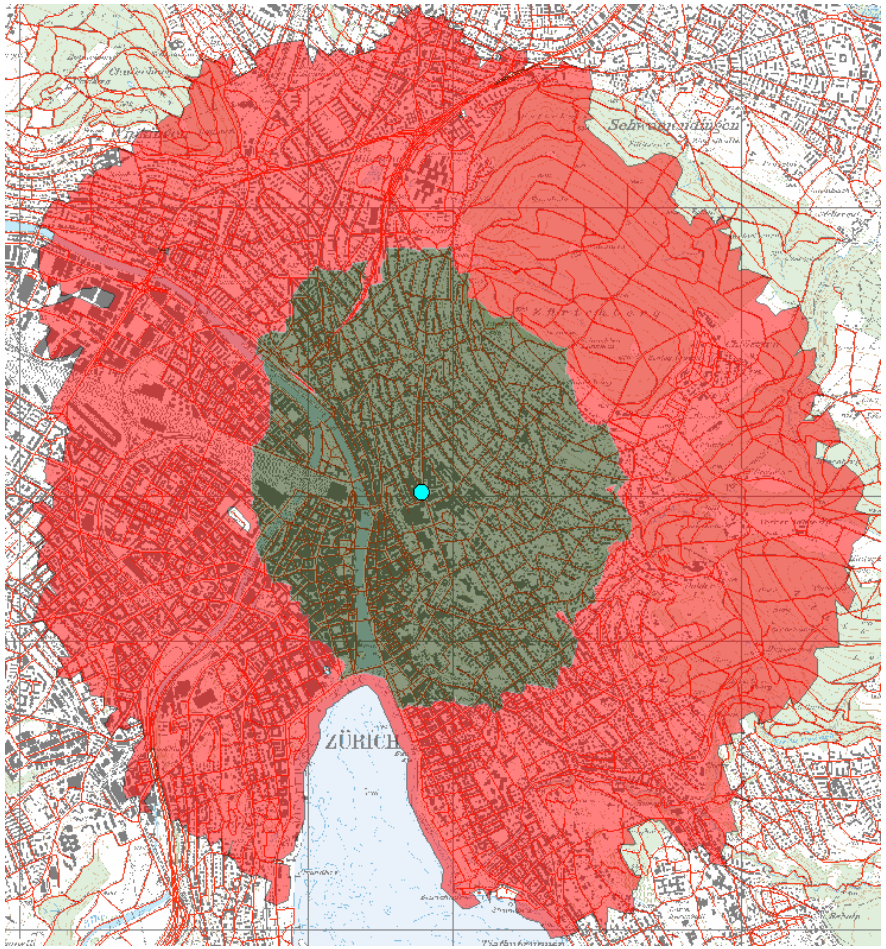
- Which facility is closest?

Vehicle-routing-problems

- Which sequence of stops is most effective?

Accessibility of Zones

- What is the accessible area in a given time?



Introduction to Networks

Application Examples

Routes and driving time

- Which is the fastest way to …?

Accessibility of objects

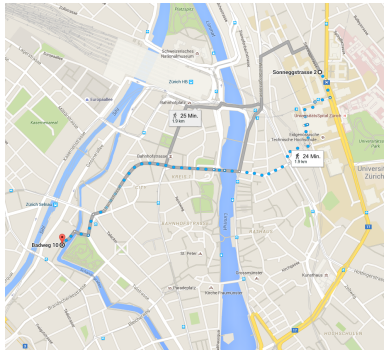
- Which facility is closest?

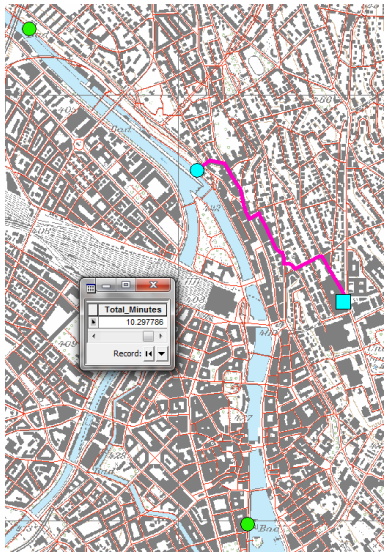
Vehicle-routing-problems

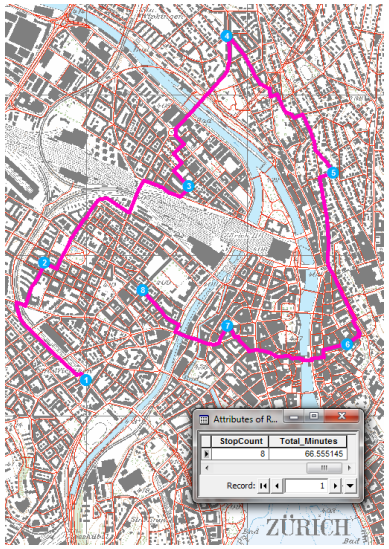
- Which sequence of stops is most effective?

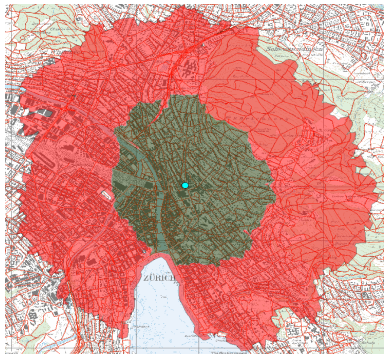
Accessibility of Zones

- What is the accessible area in a given time?



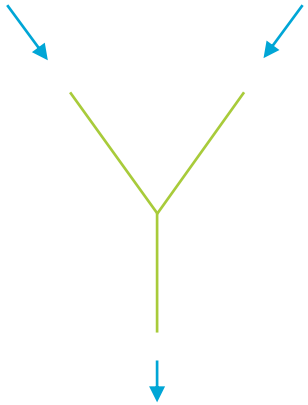




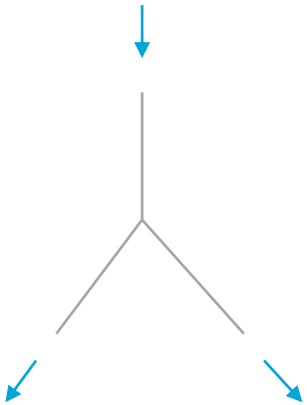


Applications

Connection



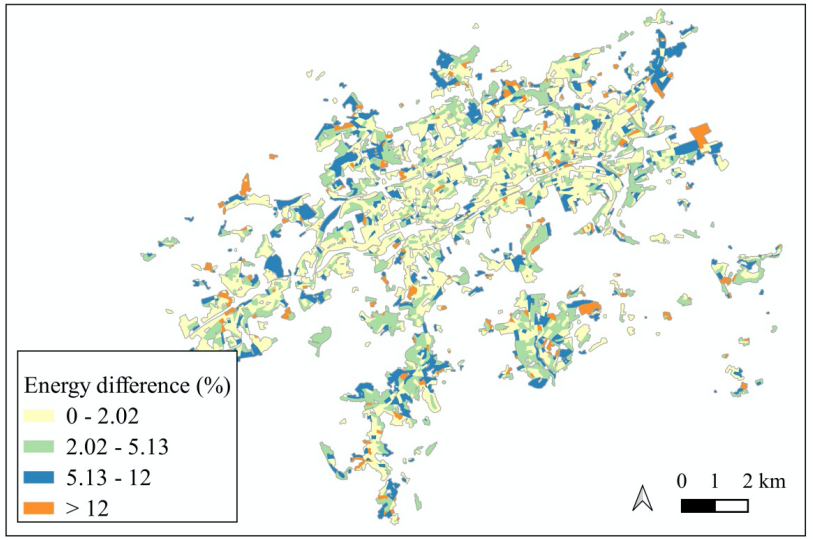
Confluence

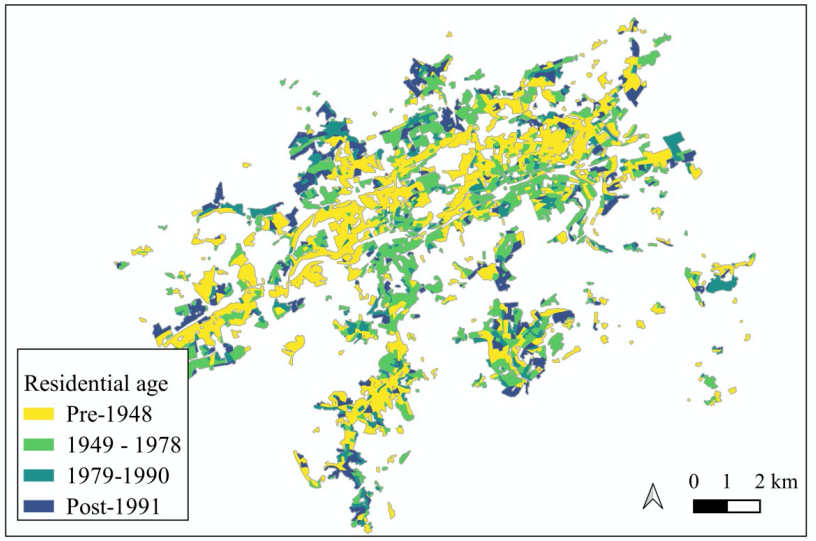


Bifurcation

River Networks
- by Bram van Meurs

Applications

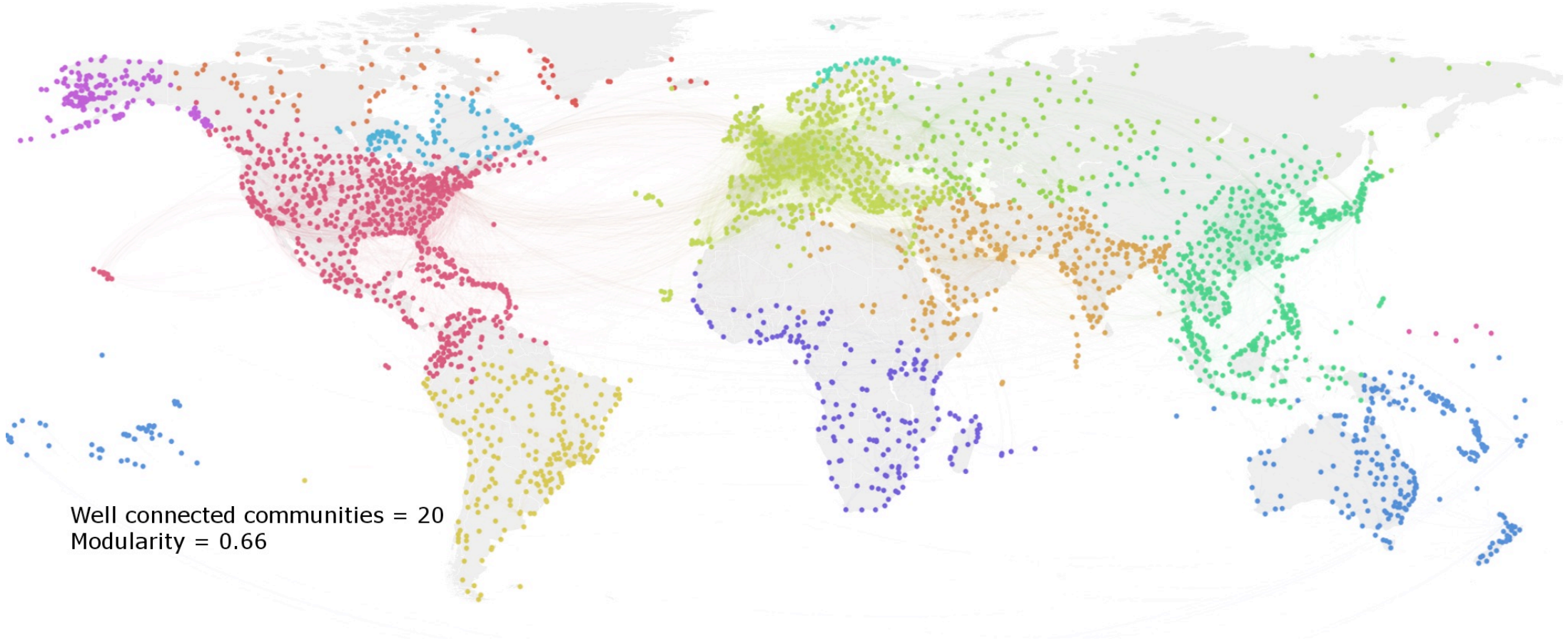




Street Networks
- by Oana Garbasevschi

Applications

Airline Networks – by me (Trivik)



Overview of Networks

What is a network?

- Definitions
- Elements, representation
- Types of graphs
- Structural properties

Definitions

- **Network**:

"*A geometric-topological arrangement of **nodes** and **edges**, e.g. in the form of a **graph** […]*”

Definitions

- **Network**:

"*A geometric-topological arrangement of **nodes** and **edges**, e.g. in the form of a **graph** […]*"

- **Network analysis**:

"*A basic group of analysis functions […] based on **line-like phenomena** to calculate and determine relations […].*

Definitions

- **Network**:

"*A geometric-topological arrangement of **nodes** and **edges**, e.g. in the form of a **graph** […]*"

- **Network analysis**:

"*A basic group of analysis functions […] based on **line-like phenomena** to calculate and determine relations […].*

This includes

  - *shortest path analysis,*
  - *searching for the nearest neighbour or the best location,*
  - *calculating a minimum spanning tree or*
  - *the solution of the travelling salesman problem.*

Definitions

- **Network**:

"*A geometric-topological arrangement of **nodes** and **edges**, e.g. in the form of a **graph** […]*"

- **Network analysis**:

"*A basic group of analysis functions […] based on **line-like phenomena** to calculate and determine relations […].*

This includes

- *shortest path analysis,*
- *searching for the nearest neighbour or the best location,*
- *calculating a minimum spanning tree or*
- *the solution of the travelling salesman problem.*

*Important for this group of analysis function is the correct representation of topological relations because mathematical methods of **topology** and **graph theory** are used.*"

Definitions

"***Topological characteristics*** *…*

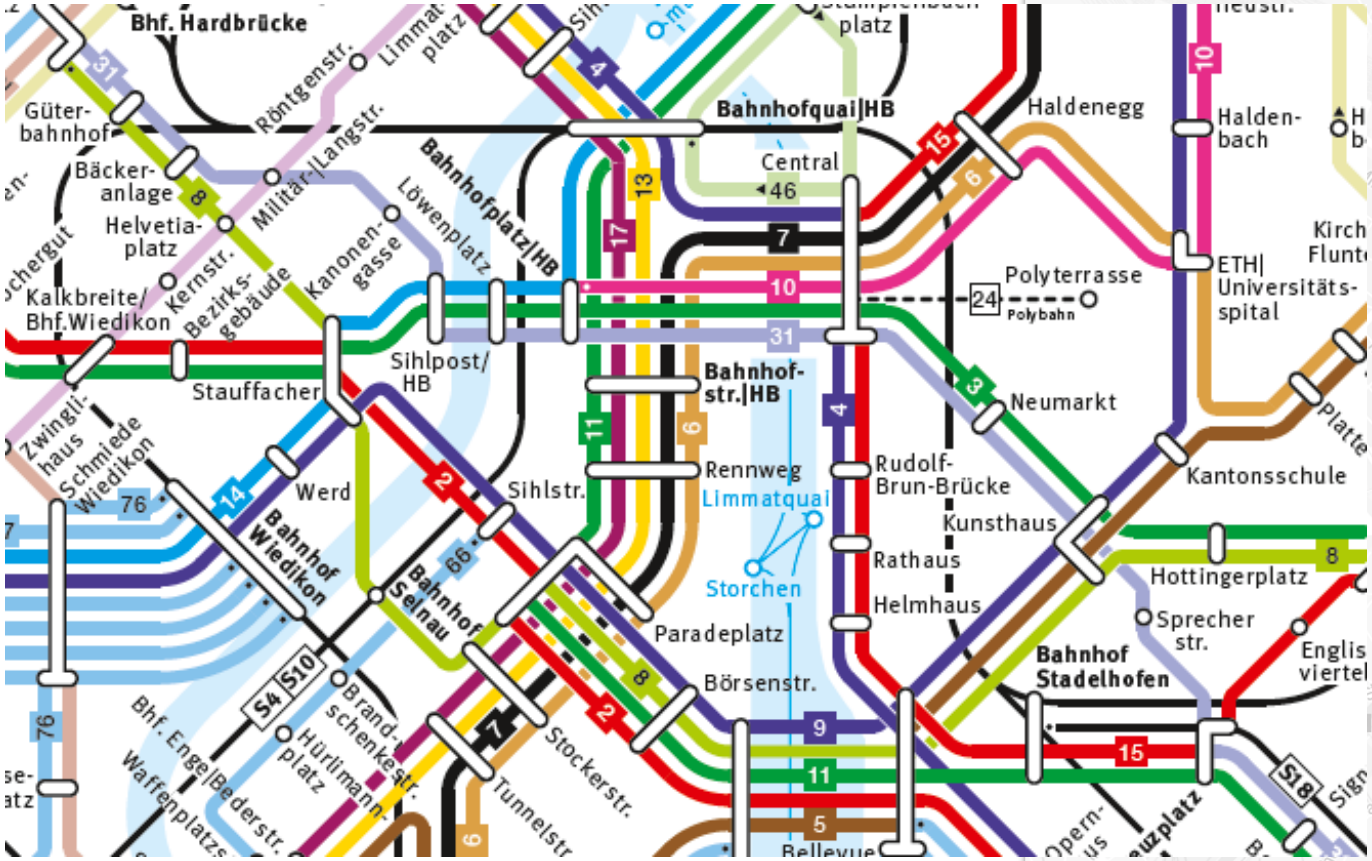
… describe the ***relative spatial relations between objects*** *[…]. Typical topological relations are related to* ***adjacencies****, (e.g. if two areas are adjacent),* ***containedness*** *(e.g. if a house is located on a certain parcel) or the* ***intersection*** *(e.g. if two roads cross).*"

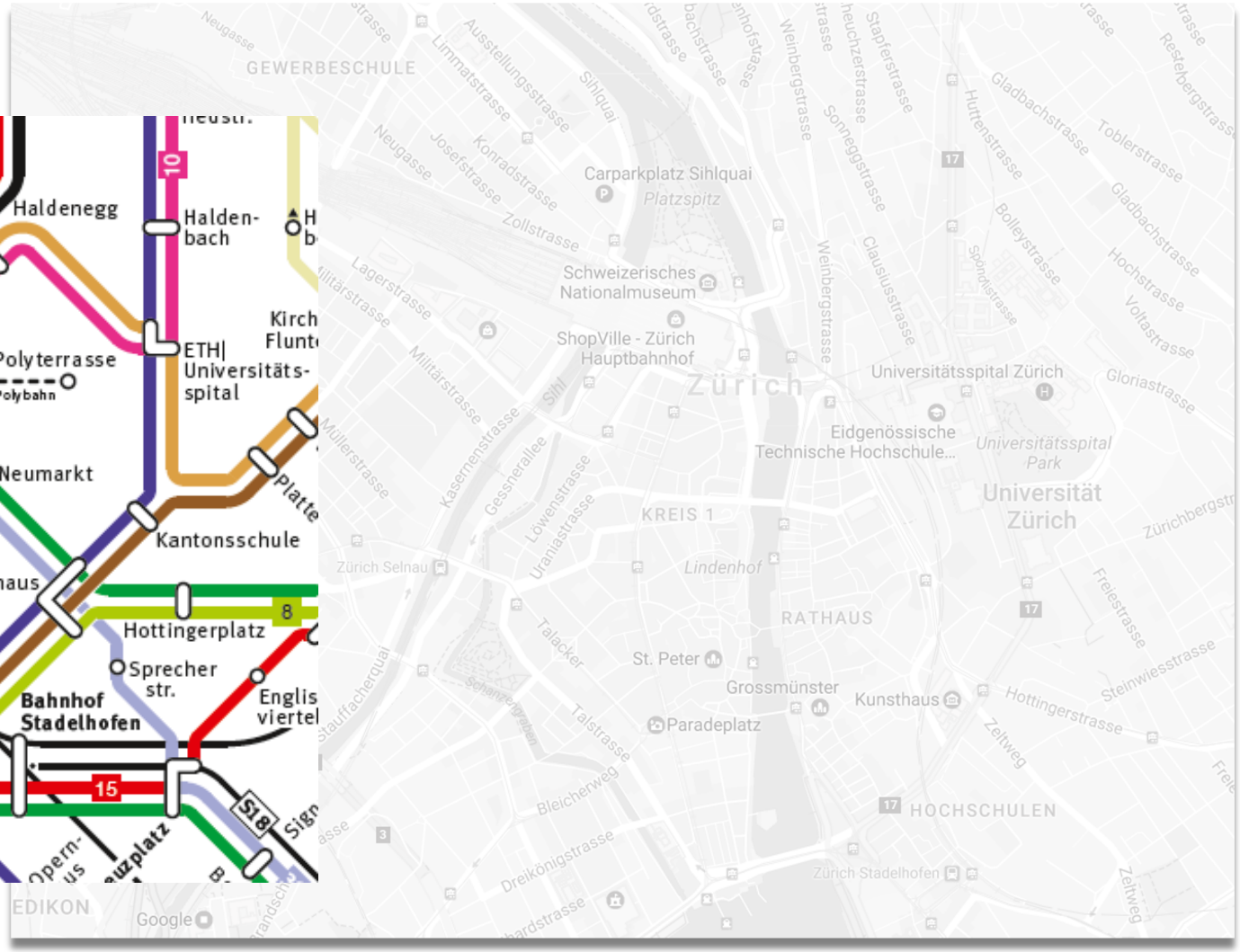
It's about **mutual positions** and **arrangement** of geometrical objects in space …

not about metrical relations.

Definitions

"*Topological characteristics …*





What is a network?

Elements/representation

- Vertex, edge
- Adjacency matrix

Types

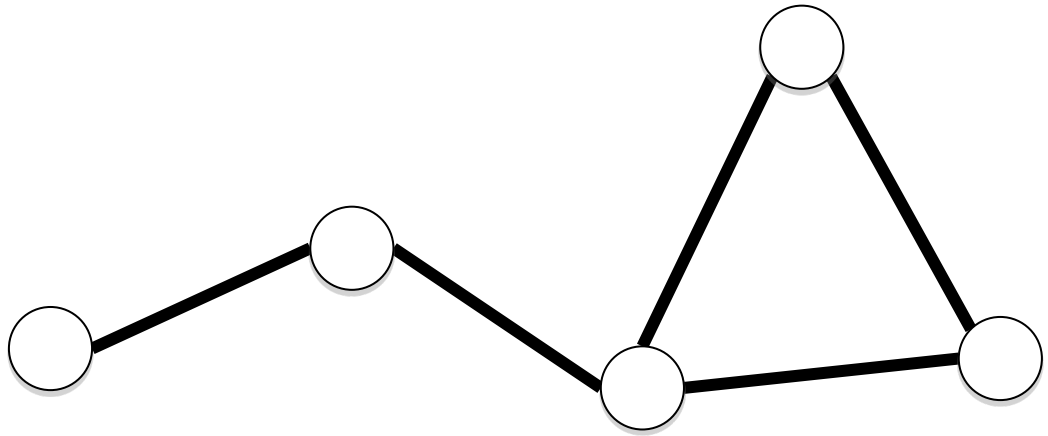
- Weighted vs. unweighted
- Directed vs. undirected
- Connected vs. disconnected
- Cyclic vs. acyclic
- Complete, tree, cubic, star

Elements/representation

Graph G = (V, E)

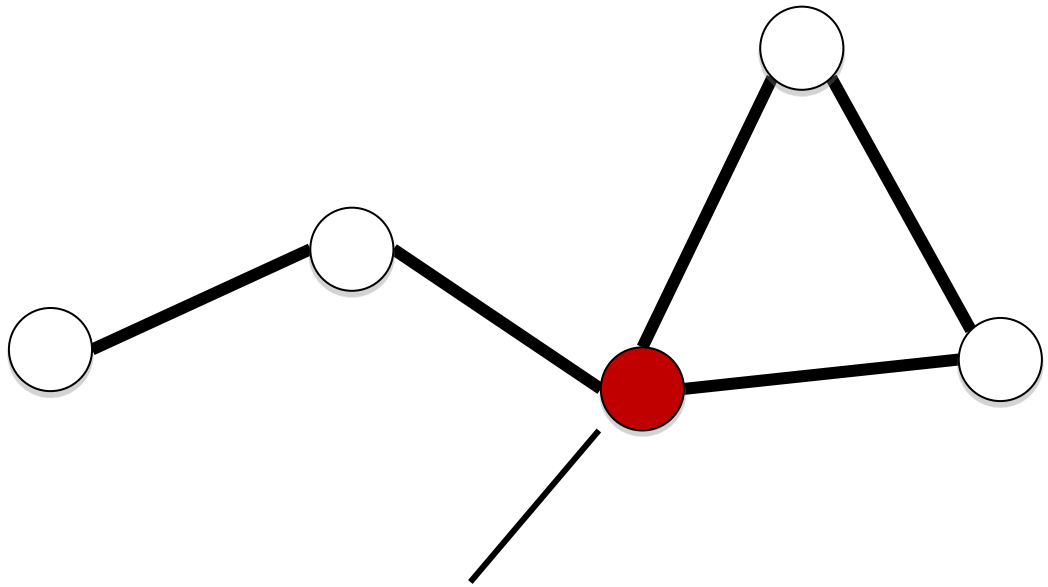
V = vertices (singular vertex)

E = edges



Elements/representation

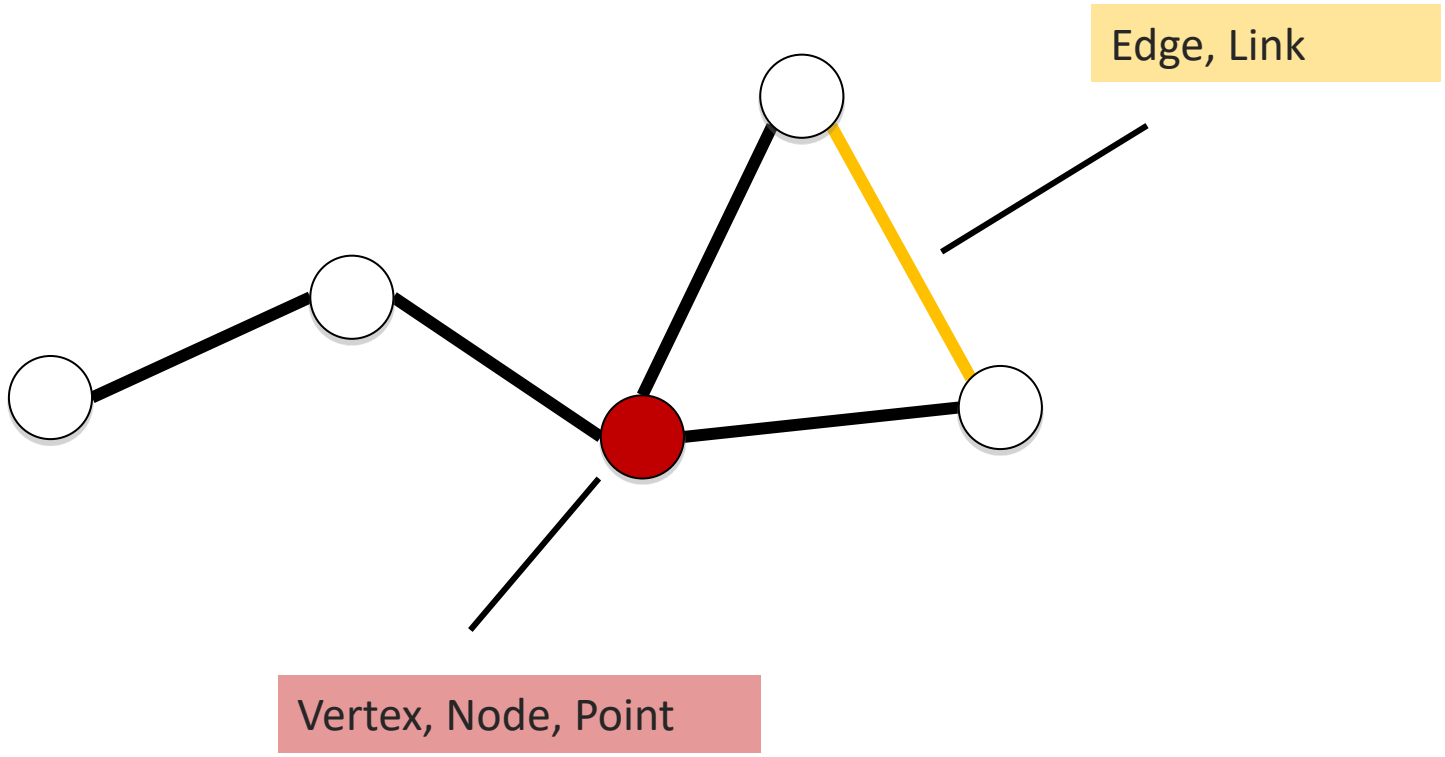
Vertex V, edge E



Vertex, Node, Point

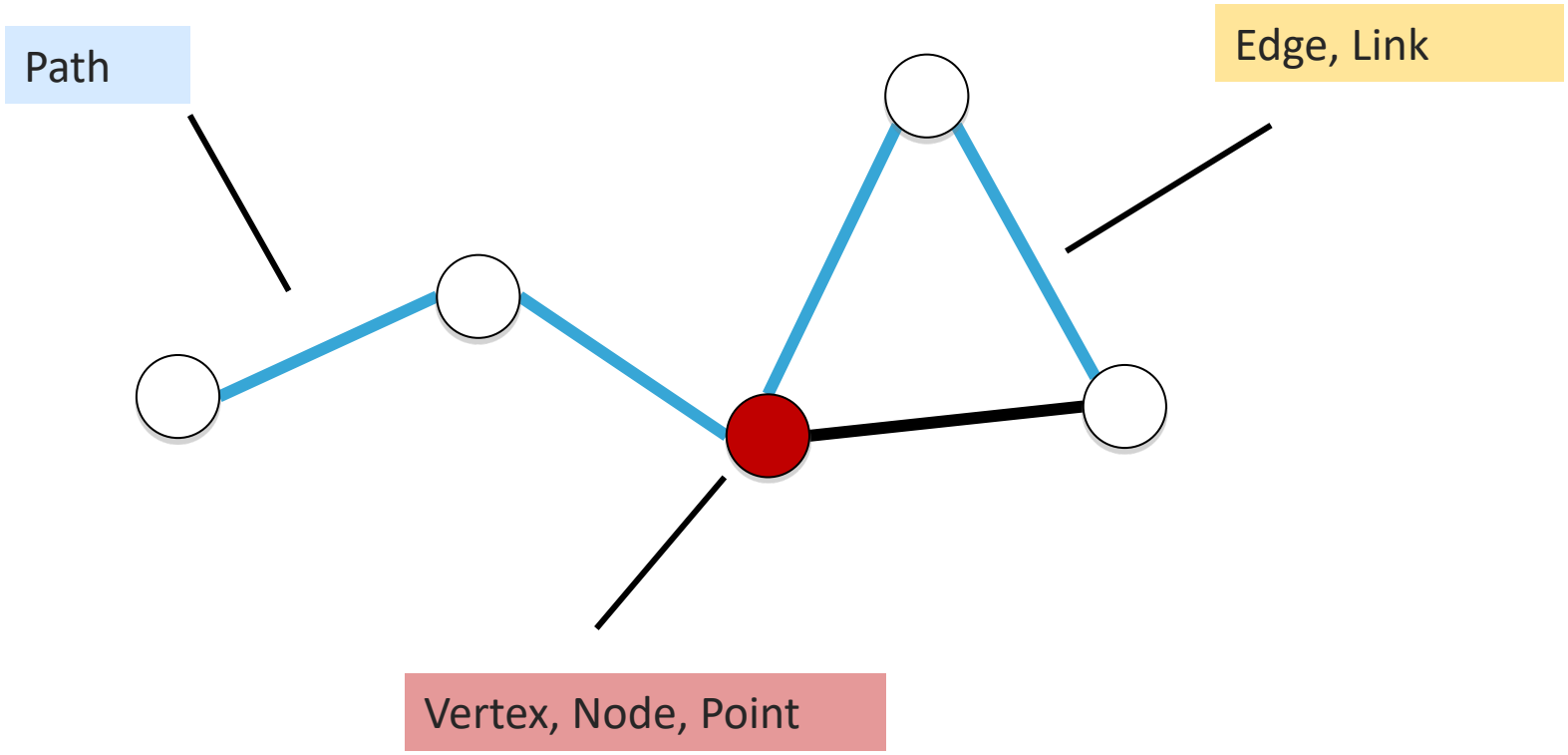
Elements/representation

Vertex V, edge E



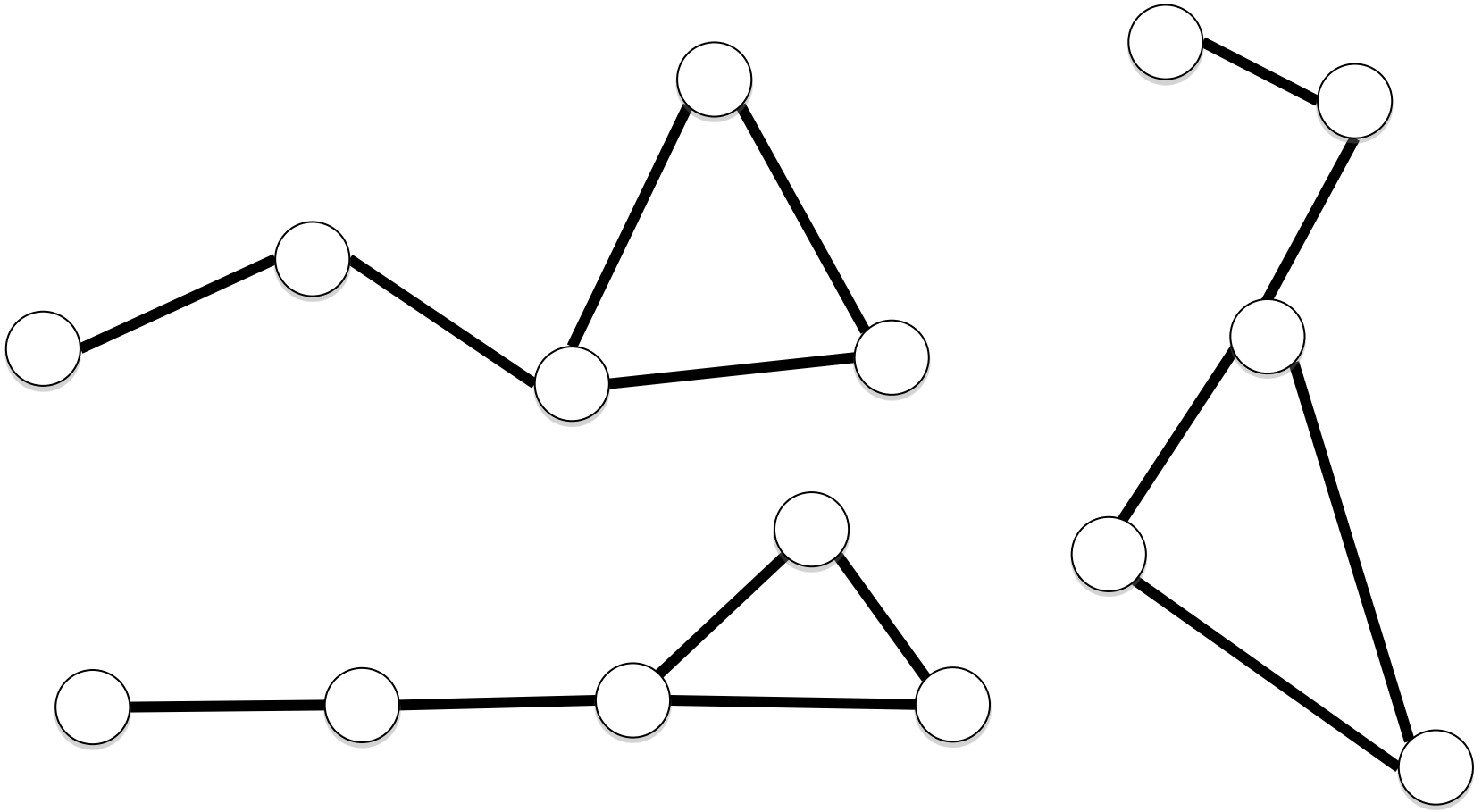
Elements/representation

Vertex V, edge E

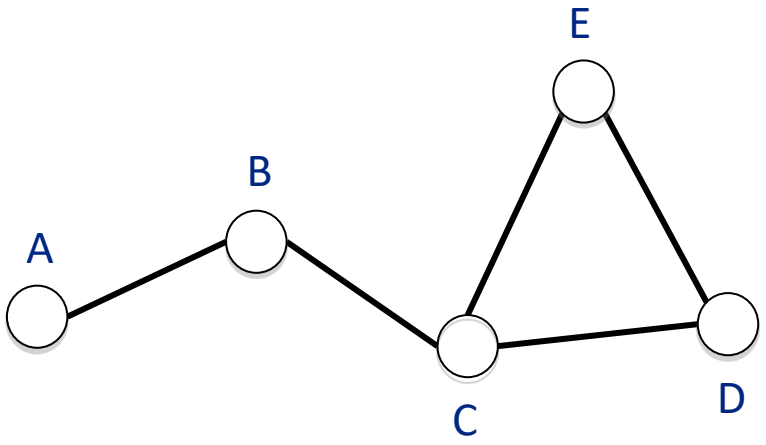


Elements/representation

Graph = independent from its visualisation



Elements/representation

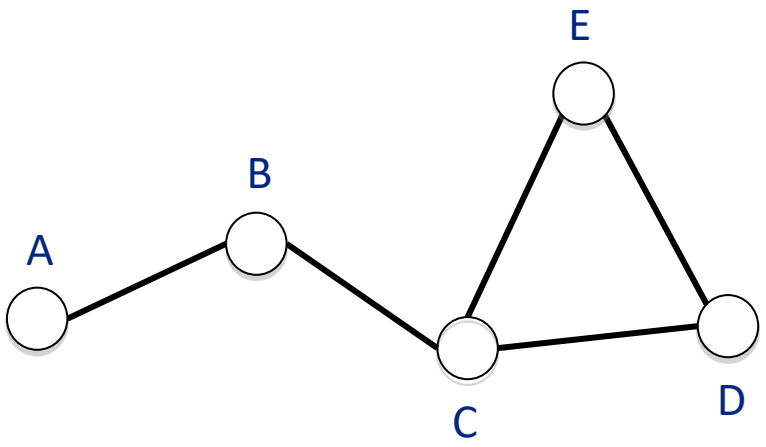


Adjacency matrix

	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	0	0
C	0	1	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

Elements/representation

Degree of a vertex:
= the number of edges connecting to a vertex



Adjacency matrix

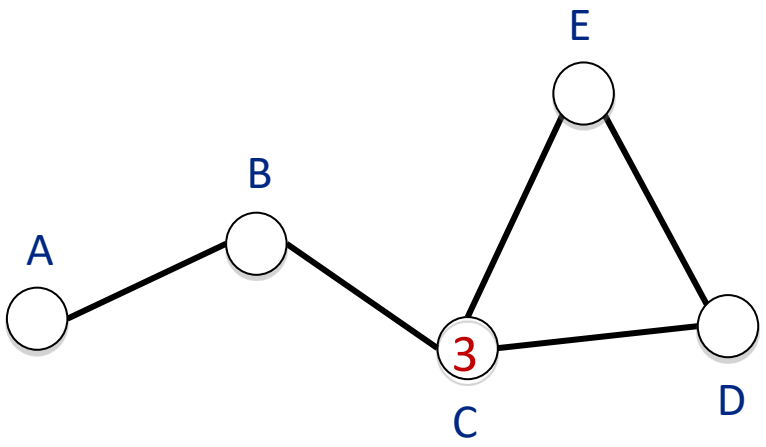
	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	0	0
C	0	1	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

3

3

Elements/representation

Degree of a vertex:
= the number of edges connecting to a vertex



Adjacency matrix

	A	B	C	D	E
A	0	1	0	0	0
B	1	0	1	0	0
C	0	1	0	1	1
D	0	0	1	0	1
E	0	0	1	1	0

3

3

What is a network?

Elements/Representation

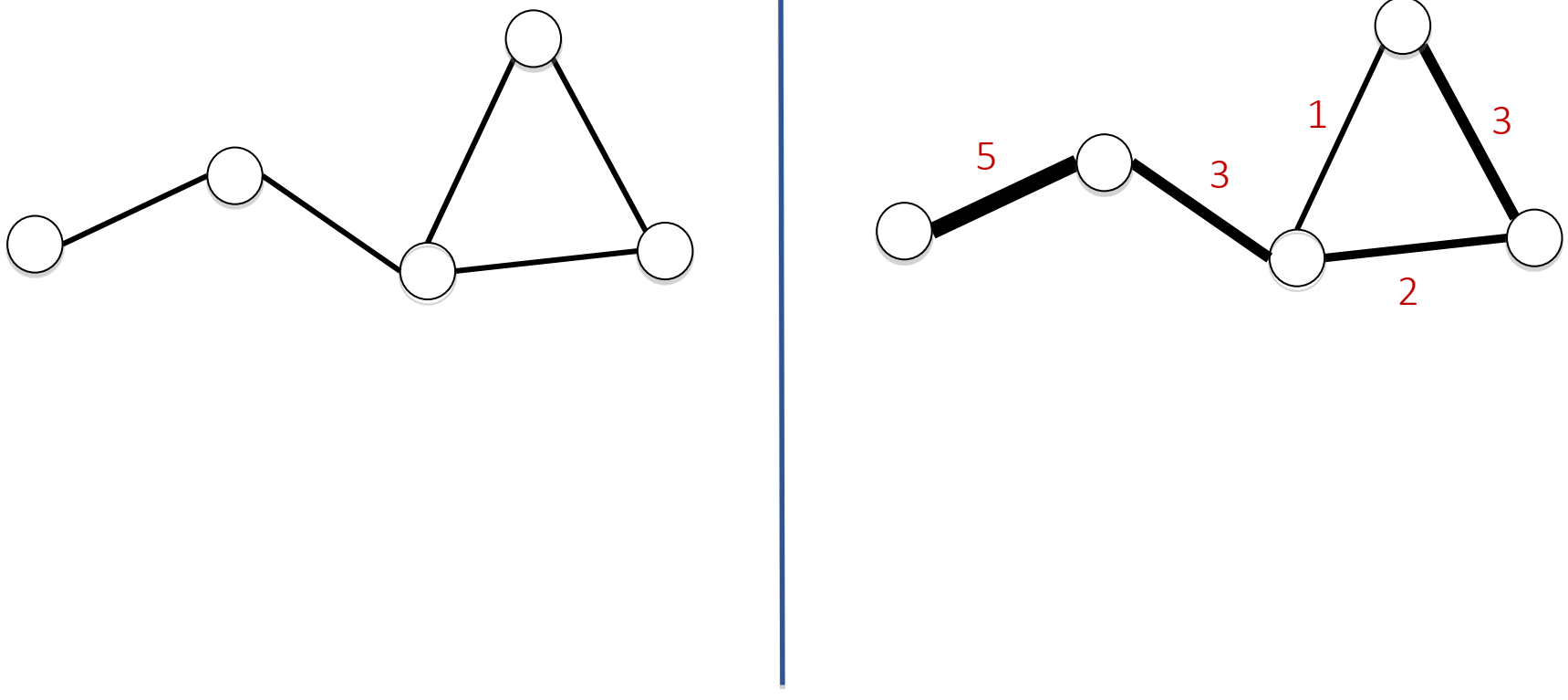
- Vertex, Edge
- Adjacency matrix

Types

- Weighted vs. unweighted
- Directed vs. undirected
- Connected vs. disconnected
- Cyclic vs. acyclic
- Complete, tree, cubic, star

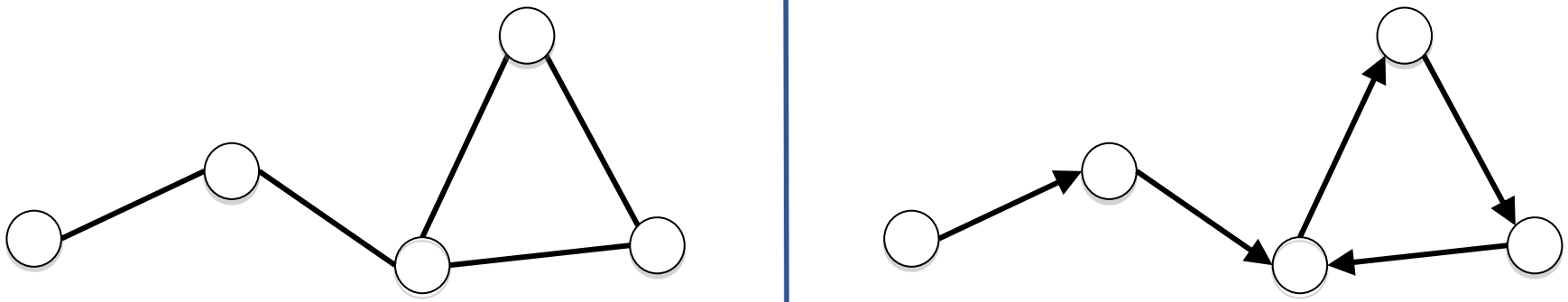
Types of graphs

unweighted graph vs. weighted graph



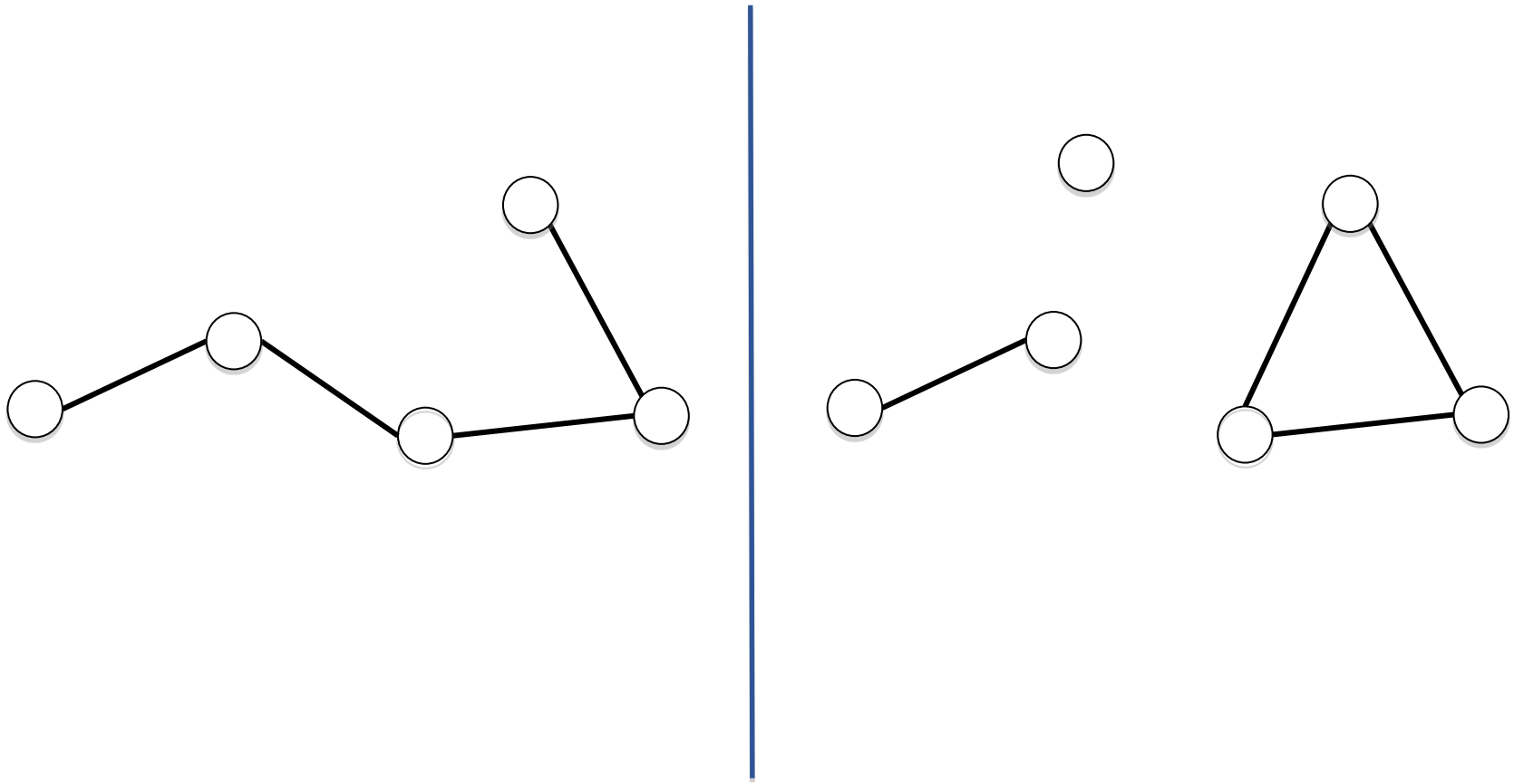
Types of graphs

undirected graph vs. directed graph (digraph)



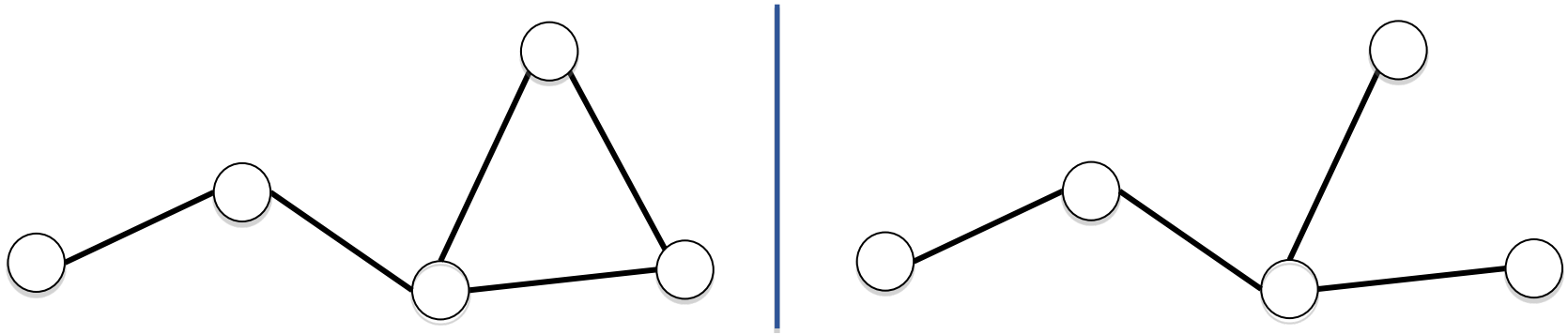
Types of graphs

connected graph vs. disconnected graph



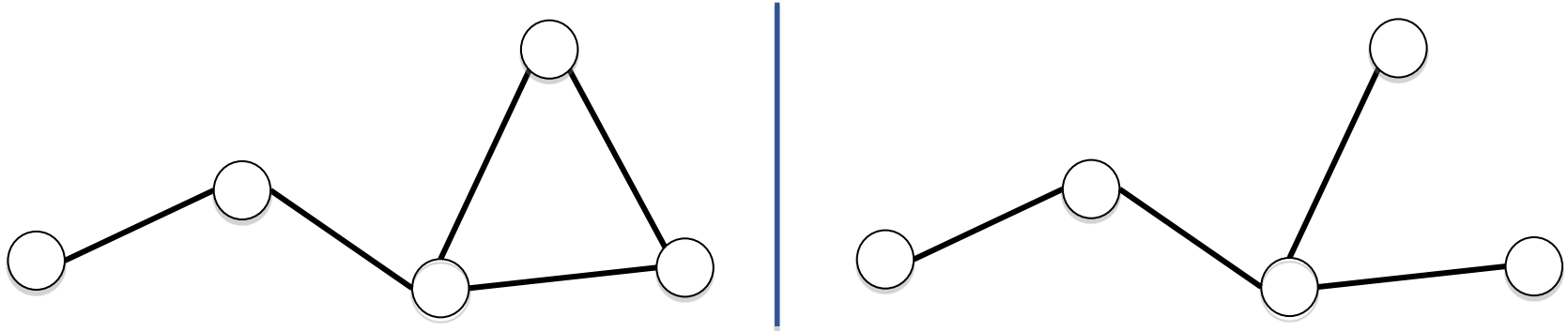
Types of graphs

cyclic graph vs. acyclic graph

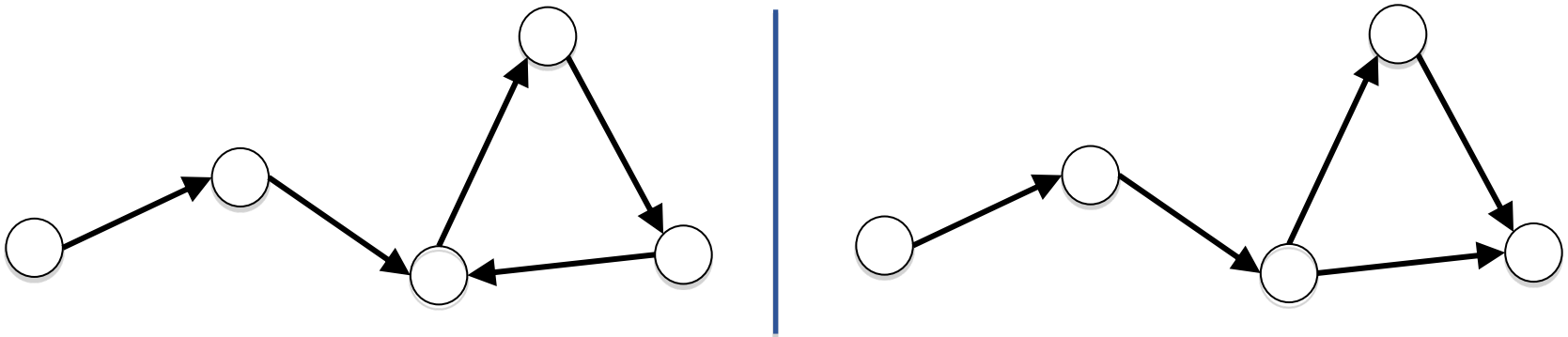


Types of graphs

cyclic graph vs. acyclic graph

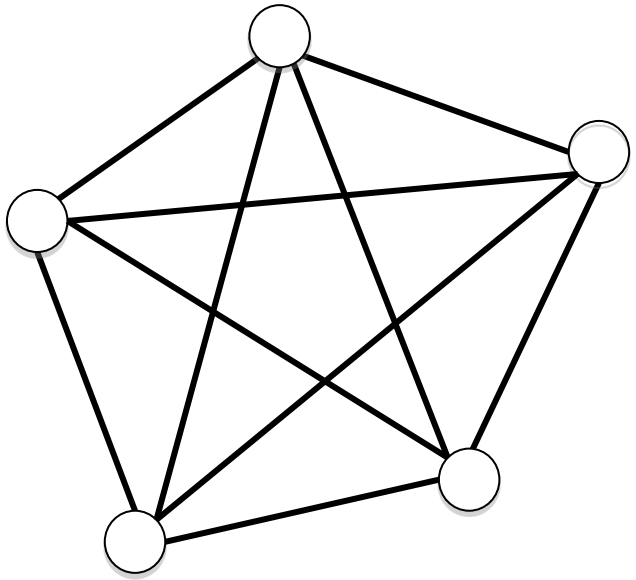


directed cyclic graph vs. directed acyclic graph

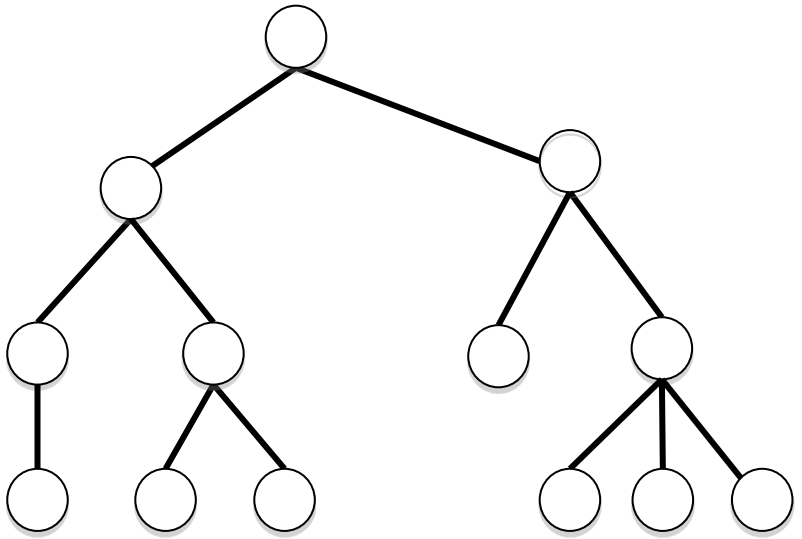


Types of graphs

complete graph



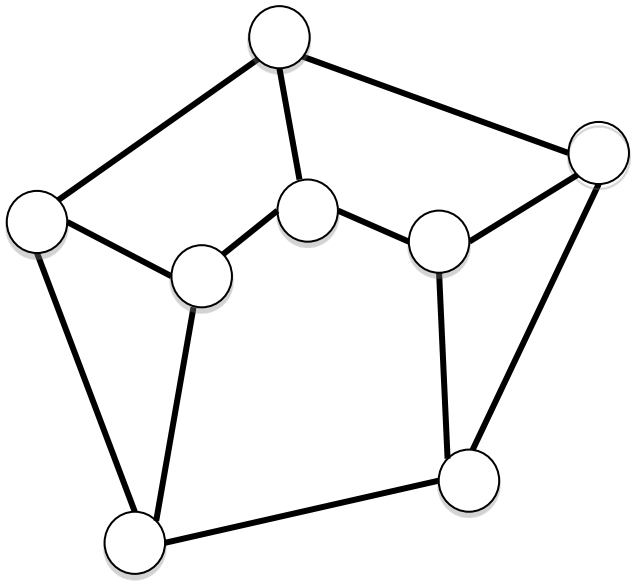
tree (connected acylic graph)



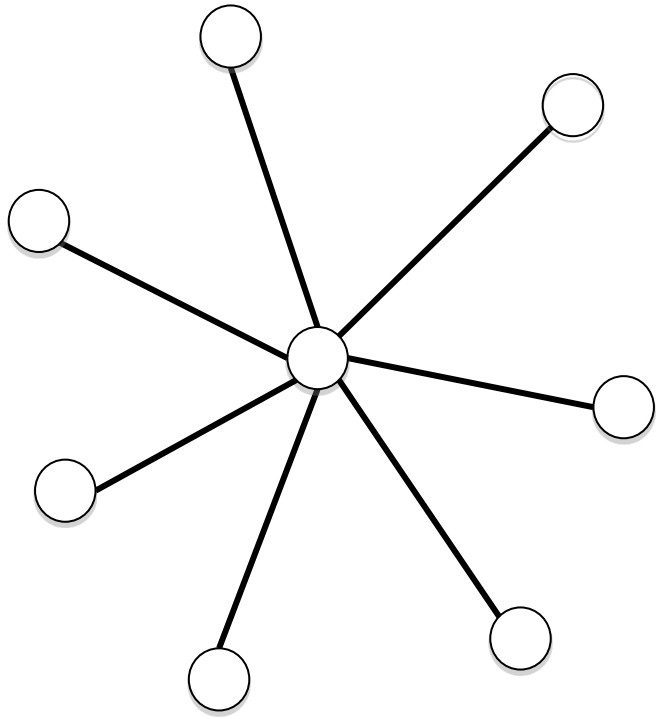
Types of graphs

cubic graph

degree = 3



star graph



Break

WATER

WALK

COFFEE OR TEA

MAKE FRIENDS

Space, formally

For a statistical method to be **explicitly spatial**, it needs to contain some representation of the geography, or **spatial context**

One of the most common ways is through **Spatial Weights Matrices**

- **(Geo)Visualization**: translating numbers into a (visual) language that the human brain "*speaks better*"
- **Spatial Weights Matrices**: translating geography into a (numerical) language that a computer "*speaks better*".

Core element in several spatial analysis techniques:

- Spatial autocorrelation
- Spatial clustering / geodemographics
- Spatial regression

W as a formal representation of Space

W

N x N positive matrix that contains spatial relations between all the observations in the sample

$$w_{ij} = \begin{cases} x > 0, & if\ i\ and\ j\ are\ neighbours \\ 0, & otherwise \end{cases}$$

$w_{ij} = 0$ *by convention*

…What is a neighbour???

* kindly do not lie with maps

Types of W

A neighbour is “somebody” who is

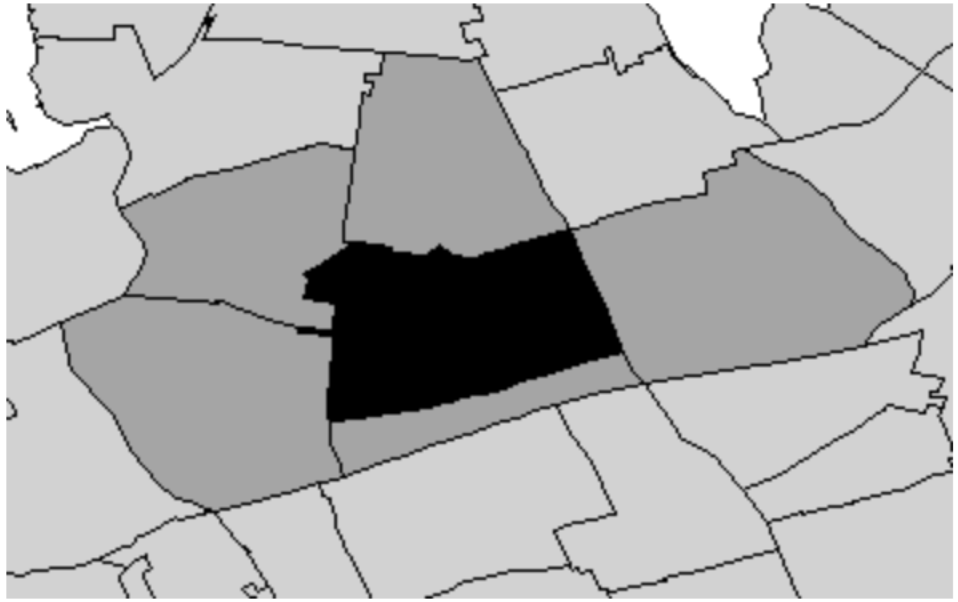
- Next door → **Contiguity**-based Ws
- Close → **Distance**-based Ws
- In the same “place ” as us → **Block** weights

Contiguity-based weights

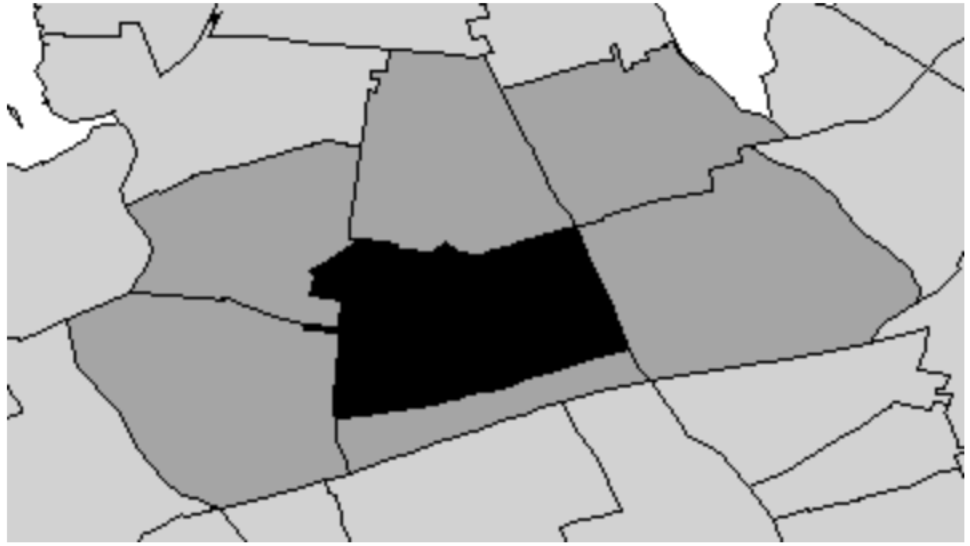
Sharing **boundaries** to any extent

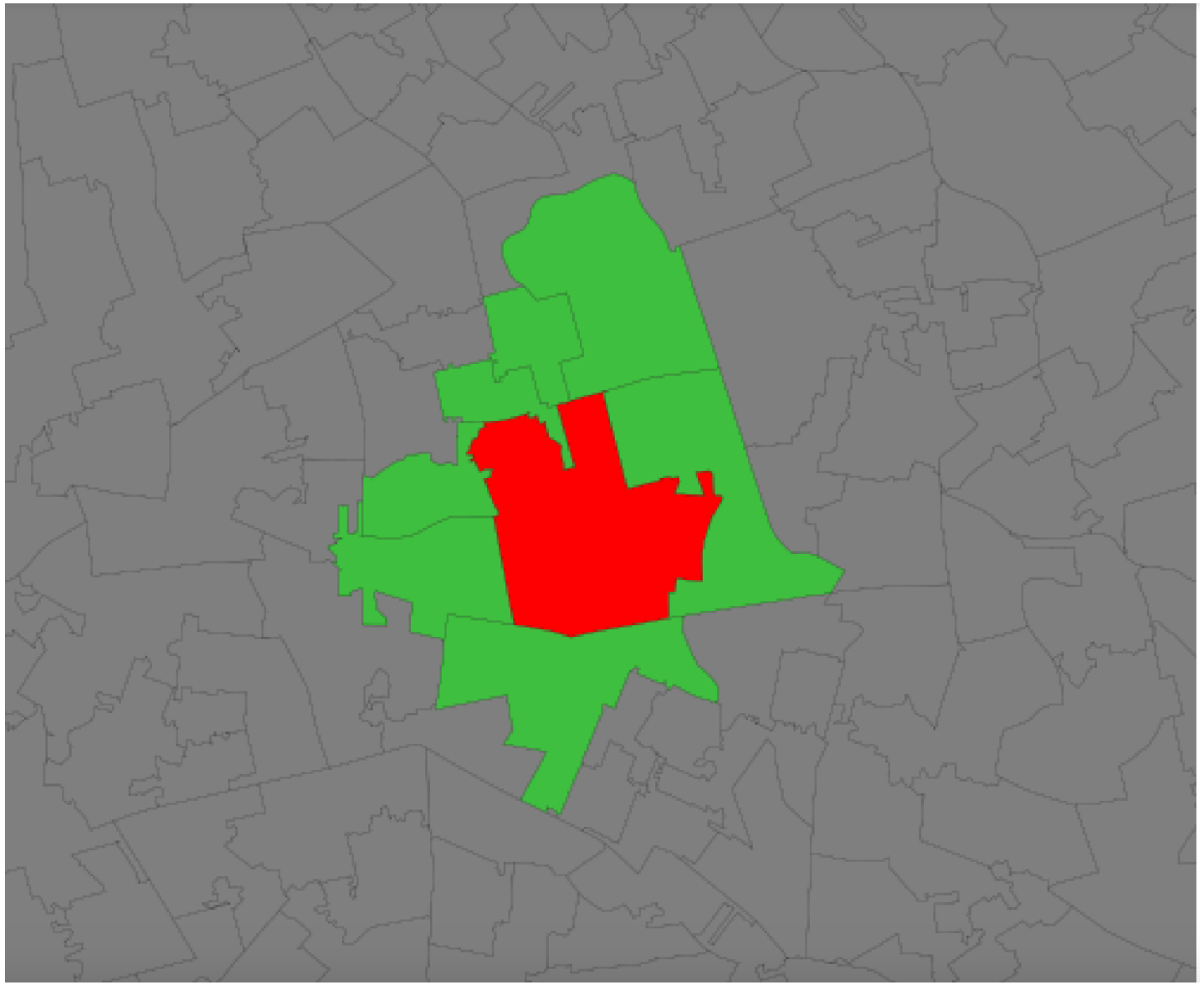
- Rook
- Queen
- …

Rook



Queen

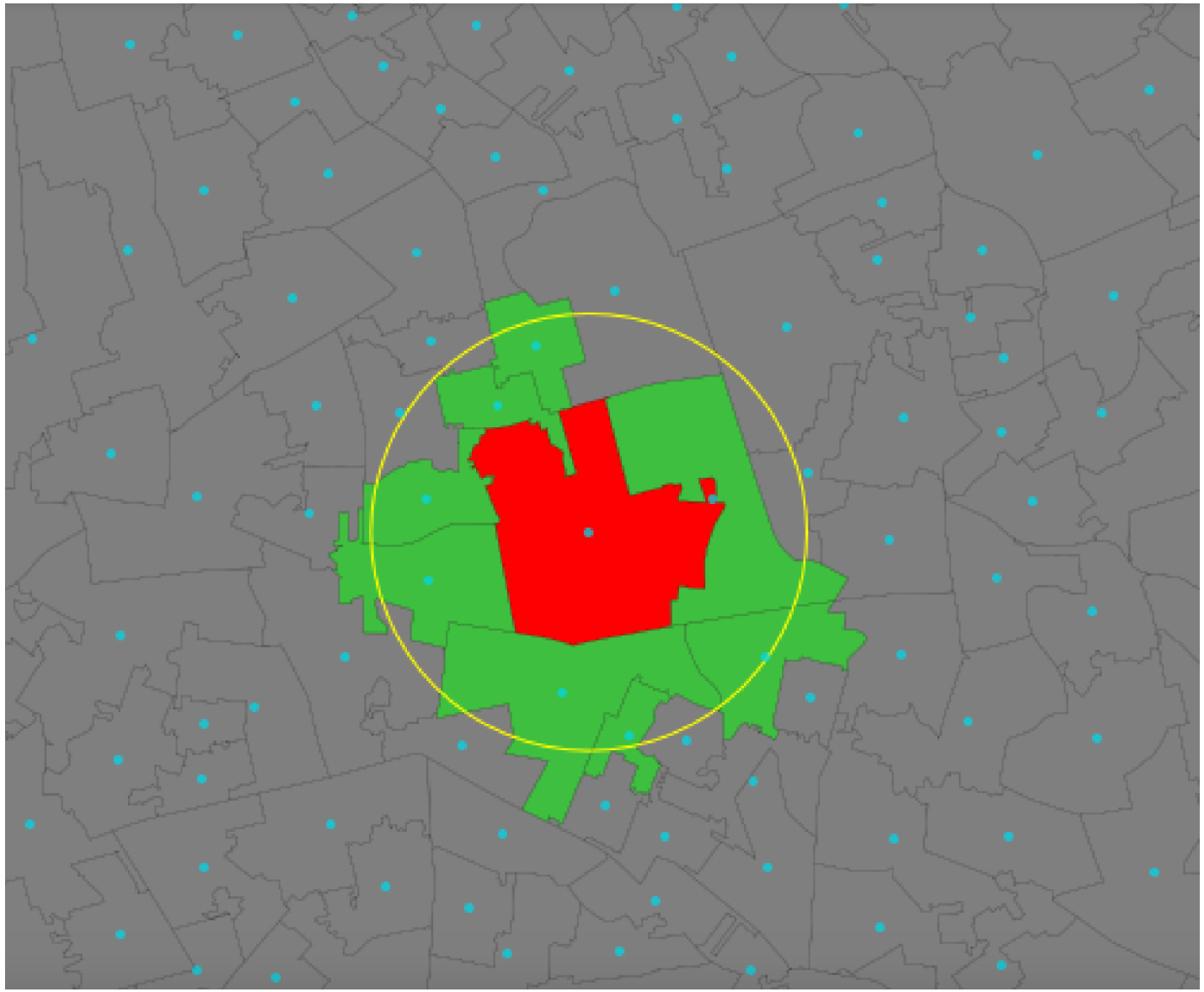


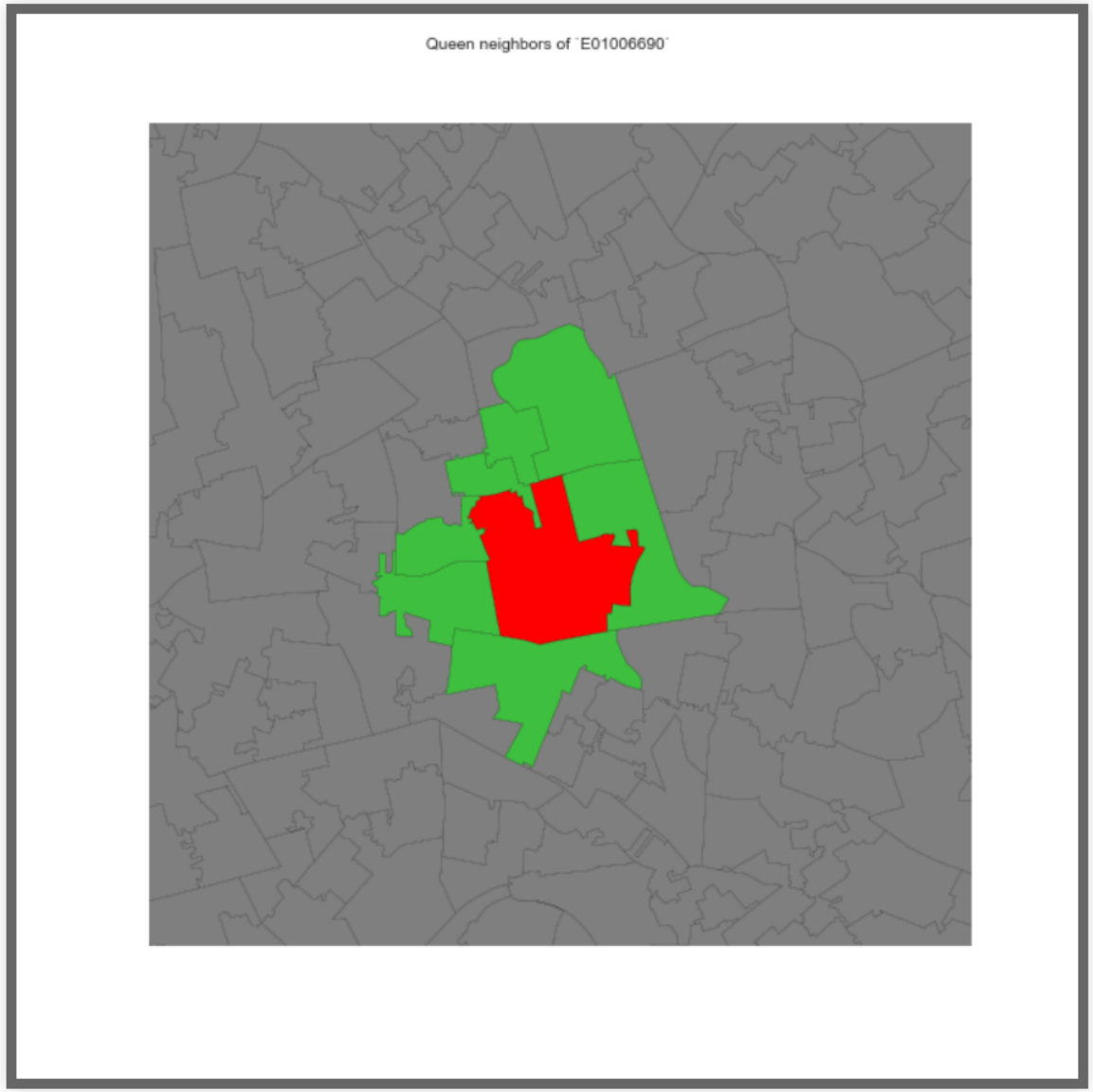


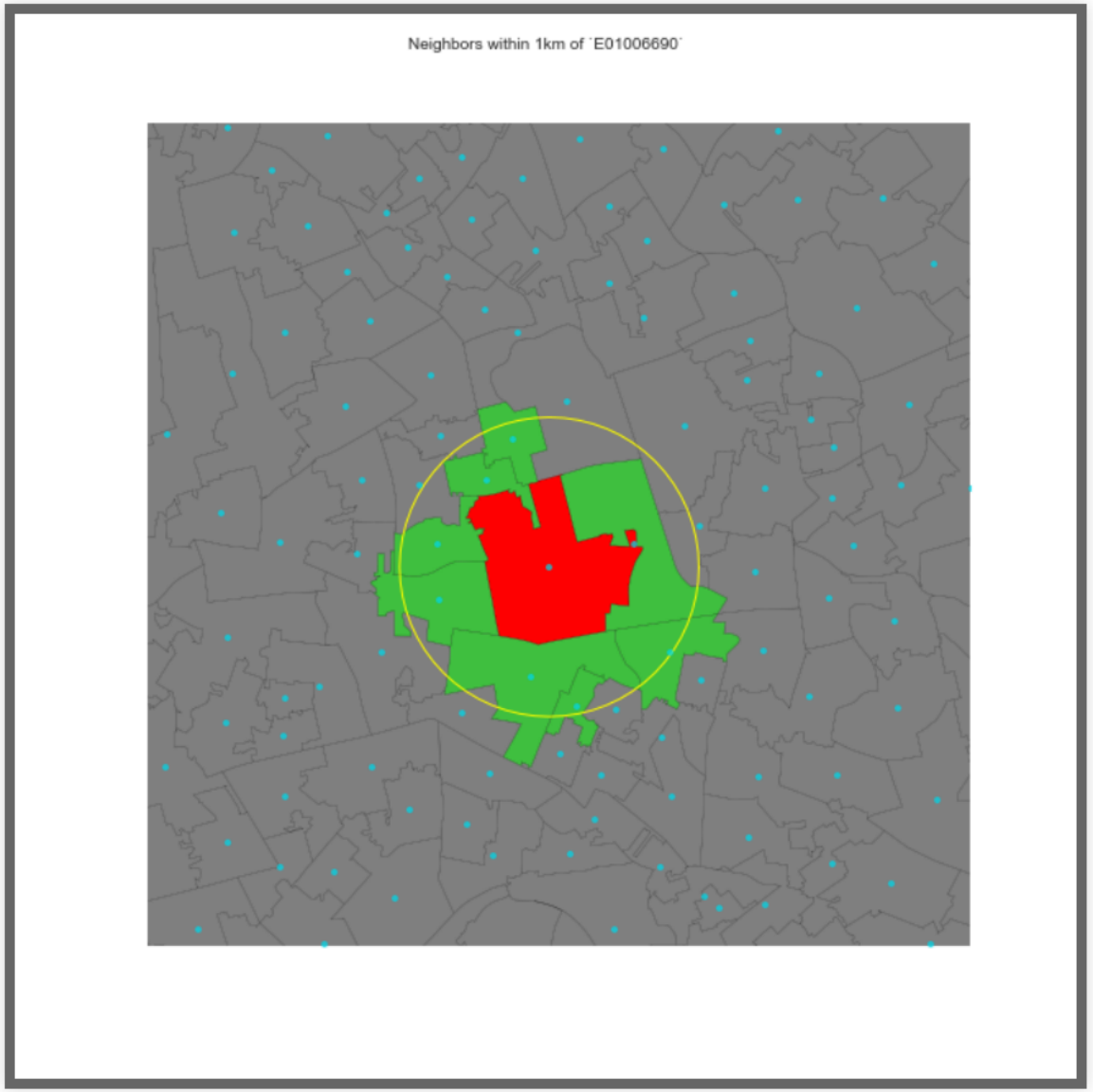
Distance-based weights

Weight is (inversely) proportional to distance between observations Rook

- Inverse distance (threshold)
- KNN (fixed number of neighbors)
- …




Queen neighbors of `E01006690`


Neighbors within 1km of `E01006690`

Block weights

Weights are assigned based on discretionary rules loosely related to geography

For example:

- Buurts into Wijks
- Post-codes within city boundaries
- Counties within states
- …

How much of a neighbour?

No neighbours receive zero weight: $w_{ij} = 0$

Neighbours, it depends, w_{ij} can be:

- One: $w_{ij} = 1$ → Binary
- Some proportion ($0 < w_{ij} < 1$, continuous) which can be a function of:
  - Distance
  - Strength of interaction (e.g. commuting flows, trade, etc.)

Choice of W

Should be based on and reflect the **underlying channels of interaction** for the question at hand.

Examples:

- Processes propagated by immediate contact (e.g. disease contagion) → Contiguity weights
- Accessibility → Distance weights
- Effects of county differences in laws → Block weights

Standardisation

In some applications (e.g. spatial autocorrelation) it is common to *standardize* W

The most widely used standardization is row-based: divide every element by the sum of the row:

$$w'_{ij} = \frac{w_{ij}}{w_{i.}}$$

where $w_{i.}$ is the sum of a row

Spatial Lag

Spatial Lag

Weighted average of neighbouring values

- Neighbour definition comes from spatial weights w_{ij}

$$Y_{iL} = w_{i1}Y_1 + w_{i2}Y_2 + w_{i3}Y_3 + \ldots w_{in}Y_n$$

Spatial Lag variable has a *smaller* variance than Y because it is a smoother function

Spatial Lag

- Measure that captures the behaviour of a variable in the neighborhood of a given observation i.
- If W is standardized, the spatial lag is the average value of the variable in the neighborhood (good for comparison and scaling)

Spatial Lag

- Common way to introduce space formally in a statistical framework
- Heavily used in both **ESDA** and spatial regression to delineate neighborhoods.
- Examples:
  - Moran's I
  - LISAs
  - Spatial models (lag, error…)

Recapitulation

- Everything is connected and must be considered so
- Spatial Weights matrices: matrix encapsulation of space
- Different types for different cases
- Useful in many contexts, like the spatial lag and Moran plot, but also many other things!

For next class..

Finish Lab 05 to practice programming

Submit Homework 05 for peer review on Peer

Submit Assignment 2 – due in **Week 5** on Friday at **2330**

See "To do before class" for every lecture (~ 1 hour of self study)

Read paper for **Discussion** session before every Friday

Post questions on the **Discussion** forum on Brightspace (especially on **Networks and Space** for this week)