

## Example data

### Neighbouring of polygons with different land use in the city

The area of a city consists of polygons with different land use. There are in the small extract (Fig. 1) of landuse (only 7 polygons) with these types of landuse:

C – commercial area, B – buildings blocks, W – water, G – green area (park).

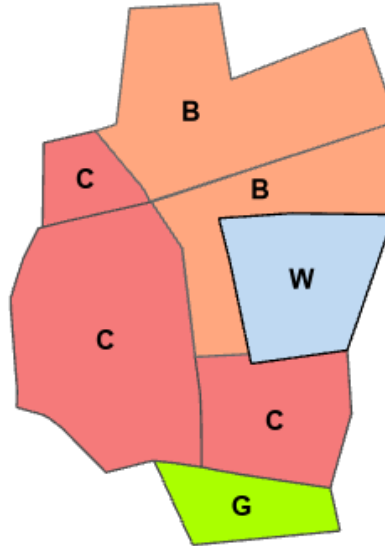


Fig. 1 Example data of polygon in the city with different landuse

### Neighbours

Each polygon neighbours with various numbers of polygons with different landuse. The neighbour is defined as existence of any landuse in defined distance around the polygon (buffer zone) – Fig. 2.

The neighbouring landuses are coded as unique existence of types in the set of landuse. It corresponds to transactions in item sets (like Market Basket Analysis). It means that repetition of the same landuse is considered as one time appearance.

**Situation - polygon C** – neighbours are only C, B, G. There is not important that type C is two times, B is two times, G only ones. Also, the order of neighbours is not important in the output data.

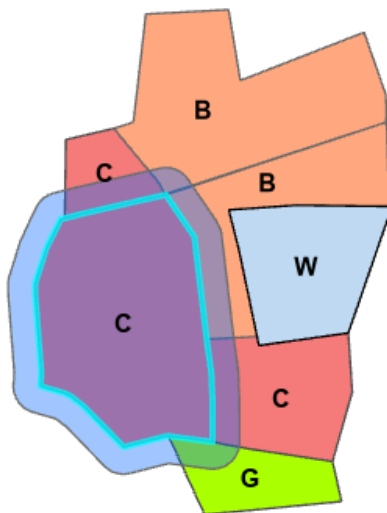


Figure 2 Neighbours of polygon C

Like the selected polygon C, it is possible to describe all small part of city (only 7 polygons).

For all polygons from the example, the data Table 1 contains the neighbours (in the depicted buffer zone in Fig. 2). The table has form of transactional data. The table with item set will be source data for finding frequent itemsets (e.g. by Orange software).

Table 1

Polygon No.	Input polygon	Neighbours
1	C	C B G
2	W	C B
3	C	C D
4	C	C G B W
5	B	B C
6	B	B C W
7	G	C

### QUESTION:

It is correct description of the situation of neighbours? The data maybe contain some **hidden** “repetition”!

2. row - W (water) contains information about water’s neighbour’s – **C** and **B**.

4. row - C (commercial) has neighbour W (beside C, G, B) .

*Is could be considered something like symmetry?*

*Are the input data correct for discovering frequent item sets? Will be there any bias by the “repetition”? Can we do not mind this “repetition”?*

NOTE: For me, it is not problem to calculate (generate) this type of item sets by GIS software for real source data to have data about the neighbours. I consider it as the first part of data processing.

BTW, there is influence of distance to the number (or types) of neighbours:

a) neighbours of C landuse in small distance: **C, B, G**

b) neighbours in bigger distance: **C, B, P, W (moreover W)**

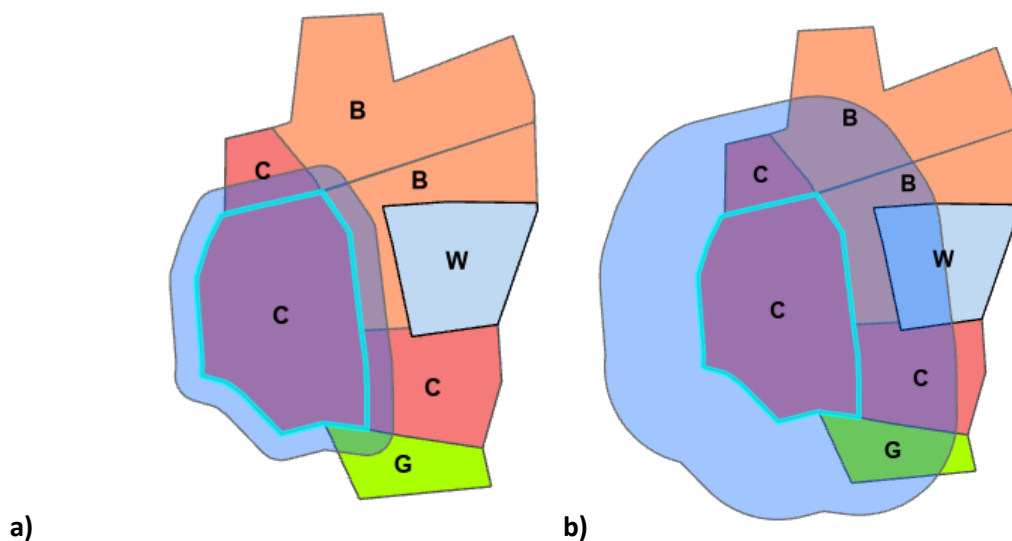


Figure 3. Influence of distance to the number of neighbours

It could be solved as two variants with different outputs (it is not problem for me). It is interesting idea in the stage of interrtaation.

## Real data

Copernicus Urban Atlas data about land use of European cities.

<https://land.copernicus.eu/local/urban-atlas>

Salzburg city:

Legend – types of landuse

- 11100: Continuous Urban fabric (S.L. > 80%)
- 11210: Discontinuous Dense Urban Fabric (S.L.: 50% - 80%)
- 11220: Discontinuous Medium Density Urban Fabric (S.L.: 30% - 50%)
- 11230: Discontinuous Low Density Urban Fabric (S.L.: 10% - 30%)
- 11240: Discontinuous very low density urban fabric (S.L. < 10%)
- 11300: Isolated Structures
- 12100: Industrial, commercial, public, military and private units
- 12210: Fast transit roads and associated land
- 12220: Other roads and associated land
- 12230: Railways and associated land
- 12300: Port areas
- 12400: Airports
- 13100: Mineral extraction and dump sites
- 13300: Construction sites
- 13400: Land without current use
- 14100: Green urban areas
- 14200: Sports and leisure facilities
- 21000: Arable land (annual crops)
- 22000: Permanent crops
- 23000: Pastures
- 24000: Complex and mixed cultivation patterns
- 25000: Orchards
- 31000: Forests
- 32000: Herbaceous vegetation associations
- 33000: Open spaces with little or no vegetations
- 40000: Wetlands
- 50000: Water

Fig. 4 Legend of Urban Atlas data – categories of landuse

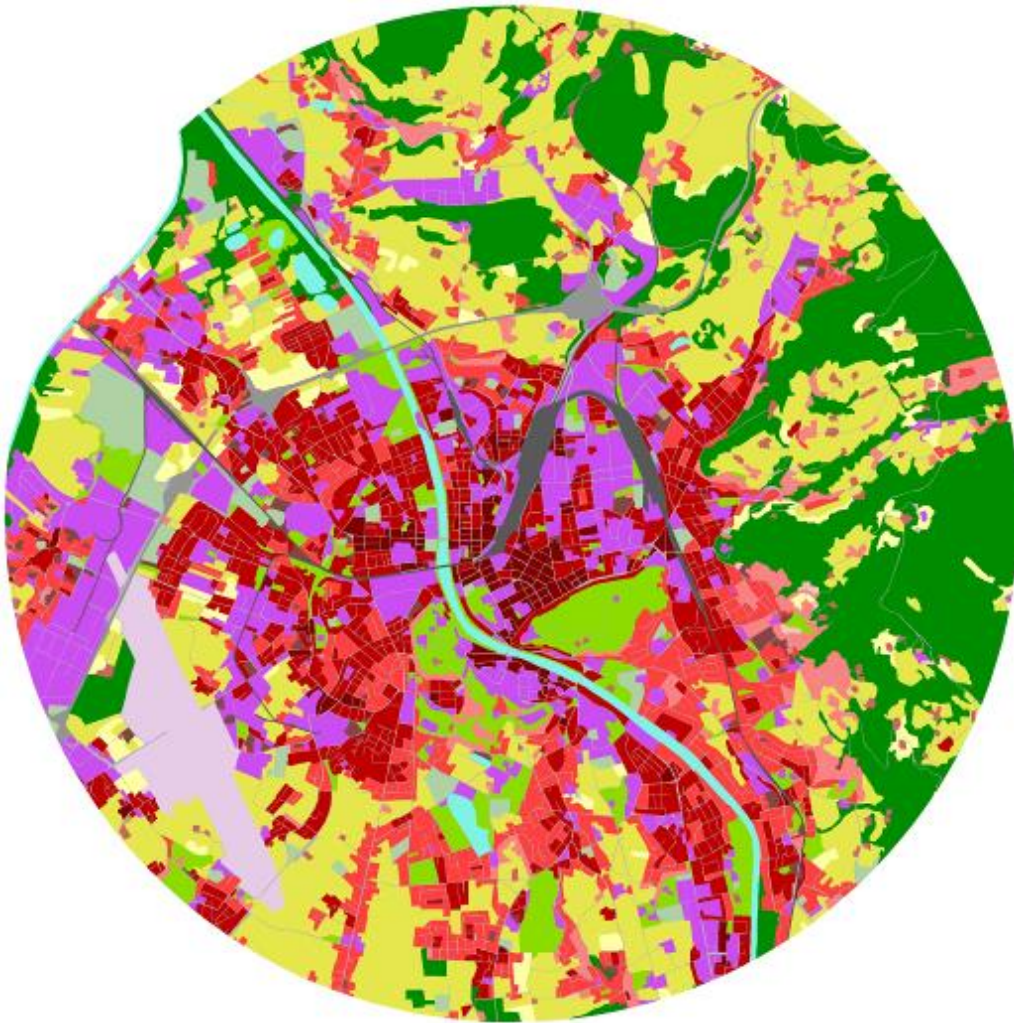


Fig . 5 – Landuse of Salzburg city (colors according to legend in Fig. 4)

## My hypotheses are:

*Frequent item sets with high support:*

Buildings *and* Commercial areas

Building *and* Green areas

Building (urban fabric) with high density *and* Bulding (urban fabric) with medium density

*Frequent item sets with minimal support (low fregeuency) – very seldom (nuggets from the point of Data mining):*

Water *and* Airport

Swamp *and* Sports areas

**The results, task or research question:**

**Find European cities that are similar in frequent item sets of landuse.**