The Similarity of European Cities Based on Image Analysis

Zdena Dobesova(✉)

Department of Geoinformatics, Faculty of Science, Palacky University, 17. listopadu 50, 779 00 Olomouc, Czech Republic
zdena.dobesova@upol.cz

Abstract. This article presents the finding of similar cities in Europe from data set Urban Atlas. Basic categories of landuse describe each city. One hundred cites were selected as a basic data set according to size. For finding the similarity, the trained neural network was used. A neural network is part of embedded add-ins Image Analytics in Orange software. One embedder in Orange was selected for the presented purpose. Finally, the hierarchical clustering was used for image descriptors received form neural networks. As a result, the couples of most similar cities is presented in the article. The cities are similar according to the patterns of urban fabrics or green areas patterns or shapes of some areas.

Keywords: Data mining · Neural network · Similarity · City structure · Geoinformatics · Urban Atlas · Orange

1 Introduction

The finding of similarity belongs to the data mining task. The history and evolution of cites are different, and they are under the influence of the political, industry and importance drivers. Moreover, the terrain arrangement, soil and surrounding influence the shape of the city. The existence of the river, mountains are determining on extension of the city during the years.

Some research tries to compare cities according to size and the structure determined for living, industry, commercial zones and leisure. Also, current urban comparisons often focus primarily on urban street networks. Work called A typology of street patterns [1] published in 2014 at the Institute of Theoretical Physics in France Compares 131 European and American cities based on street network metrics on which they apply a hierarchical clustering method. The resulting groups of cities subsequently characterizes and compares. They were faced with the problem of too high a similarity of street networks in most cities. To solve this problem, they focused on the city block geometry which street networks create. As a result of this work, cities were divided into four groups according to the shape and size of the city blocks.

Another similar work published in 2018 works with a street network for 27,000 US cities available through Open Street Map [2], relying on graph theory. Network metrics such as average edge length, crossover density, average node connectivity is calculated for multi-scale urban street networks [3]. The author compared individual territories by

R. Silhavy et al. (Eds.): CoMeSySo 2019, AISC 1046, pp. 341–348, 2019.
https://doi.org/10.1007/978-3-030-30329-7_31

these calculated metrics. None of the presented works did assume the landuse and function of urban parts in the cities or surroundings. Our research tried to consider city landuse.

The data set is described in Sect. 2. Section 3 is a description of used software Orange and steps of processing. The findings of similarities are described in Sect. 4.

2 European Urban Atlas as a Source Data

Copernicus Land Monitoring Service offers the dataset European Urban Atlas. Urban Atlas provides pan-European comparable landuse and land cover data for Functional Urban Areas (FUA). The European Urban Atlas was designed to compare landuse patterns amongst major European cities, and hence to benchmarking cities in Europe. It uses images from satellites to create reliable and comparable high-resolution vector maps of urban land in a cost-efficient manner [4]. There older data set is from 2006, and the newer dataset is from 2012. Some data are newer, e.g. from the 2016 year. Urban Atlas offers over 800 Functional Urban Area and their surroundings (more than 50.000 inhabitants) for the 2012 reference year. The data are freely downloadable in vector format together with metadata, map and color legend for categories of landuse (Fig. 1).

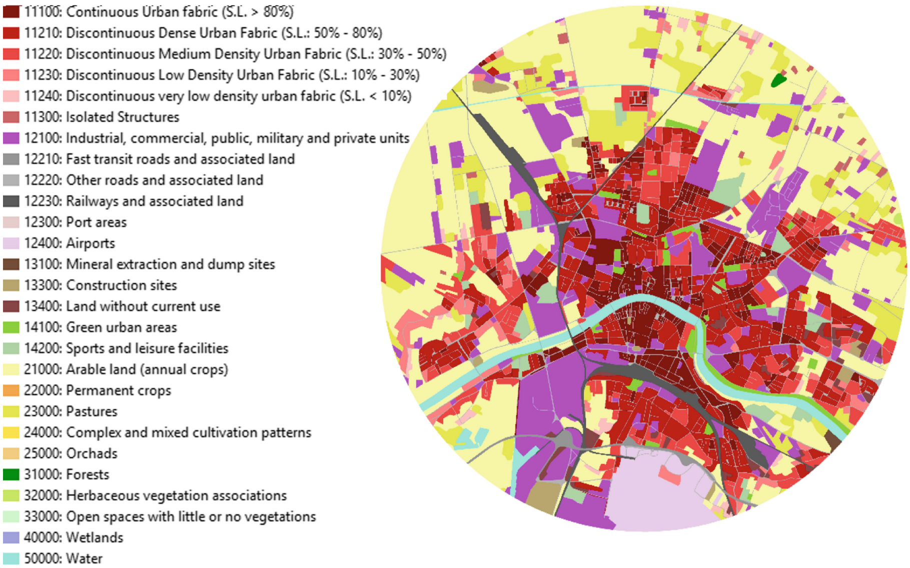


Fig. 1. Categories of landuse in Urban Atlas and city Pisa in Italy.

The land is divided to nearly 30 hierarchical classes: *Urban fabric with different density, Industrial and commercial areas, Transit and railway areas, Water, Green and sports and leisure facilities, and other green areas like forests* [5]. Each category has its color on the map. The tones of red colors are for urban fabric; blue tones are for water and wetlands; yellow tones are for arable land (pastures) and green tones for green urban areas (forests, vegetation, sports areas); grey tones are for transit roads and railways. Moreover, specific areas like construction sites, ports and airports are present (Fig. 1).

We selected 100 cities from the Urban Atlas. The selection was based on the number of the inhabitant in the interval between 50 to 200 thousand that has the largest representation in Urban Atlas. The number of inhabitants was taken from Eurostat statistics [6]. Our selection covers all countries in Europe equally. Because the source data contain not only cities but also the surroundings with other suburbs, cities and villages (whole FUA) we selected the central city manually by a circle that covers only the main city in FUA. The central point of the city was set primarily to the historical center or the buildings with a concentration of government, commercial activities or traffics [7]. The process of data preparation – the setting of a central point, definition of a circle for clipping of the source vector data were processed in GIS software Arc Map for Desktop v. 10.6 [8]. Finally, the export from GIS to the color images was made for all 100 cities. To automate the preprocessing of data, the Python script was utilized. All images were a source data set for similarity findings.

3 Orange and Neural Networks

The *Orange* is an open source machine learning software with good data visualization for novice and expert. Orange offers interactive data analysis workflows with a large toolbox [9]. The add-ons *Image Analytics* simplifies the loading of images and through deep network-based embedding enables their analysis. Embedding represents images with a feature vector, allowing the use of Orange's standard widgets for clustering, classification or any other kind of feature-based analysis [10].

The *Image Embedding* widget reads images and uploads them to a remote server. This widget offers seven embedders (neural networks), each trained for a specific task. For the map processing, the most suitable is *Painters embedder*, which is trained to predict painters from artwork images. The second suitable is VGG-19 embedder: 19-layer image recognition model trained on ImageNet. Deep learning models calculate a feature vector - image descriptors in the form of an enhanced data table with additional columns. We experimented with all embedders, and finally, the Painters embedder was chosen.

The workflow in Orange depicts the whole processing of data about cities (Fig. 2). After the calculation of image descriptors, the node *Distances* calculates distances between image descriptors by Cosine metric. The last node *Hierarchical Clustering* produces a dendrogram. For the linkage in hierarchical clustering, the Ward method was set. The levels in dendrogram reveal the group of most similar cities. The workflow also contains the *Data Table* with image descriptors and *Image Grid* to display all 100 cities in the grid to express proximity. The dendrogram was used as the result for the evaluation of similar cities. The lowest level displays the most similar cities in join branches (Appendix 1).

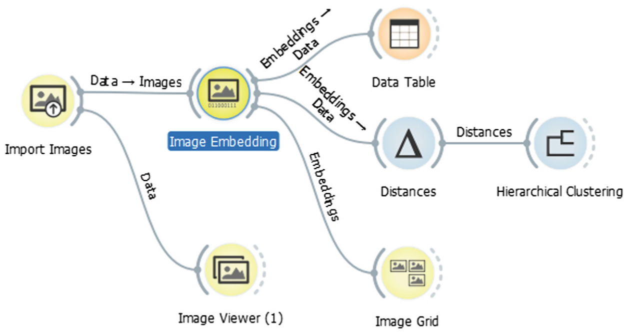


Fig. 2. The workflow in Orange.

4 Results

The hierarchical clustering identifies the four basic groups of similar cities. The first blue group (Appendix 1) are cities that only with a partial cover of circular extent. They are namely ports like Livorno in Italy and Cadiz and Almeria in Spain. All those cities are specific, and the similarity is not high.

Interesting couple of two similar is cities Bern in Switzerland and Maribor in Slovenia at the lowest level in the second red group in the dendrogram. The selected couple is in Fig. 3. These two cities are similar, namely by the curve shape of surrounding categories *arable lands*, *pastures* and *green areas*.

The second couple is from the third green group. They are Cambridge and Warwick from the United Kingdom. The rectangular shape of the urban fabric and landuse like a mosaic of small areas is typical for both cities. Also, the discontinuous urban fabric is dominant than continuous urban fabric in centers for cities in this group. Category *pastures* are dominant in the surroundings of both cities. The third couple in Fig. 3 are from the same green group. They are Le Mans from France and Enschede from the Netherlands. The typical is the equal spreading of small areas of category *Industrial, commercial, public, military and private units* (violet color) for these two cities. Also, the category C*ontinuous urban fabric* (with high density >80%) is dominant than D*iscontinuous medium density urban fabric*.

The last orange group is the biggest in the dendrogram (Appendix 1). For comparison, the couple of two similar Czech towns was selected. The first is České Budějovice and the second is Hradec Králové. For these two towns are typical the dispersion of *Continuous urban fabric with high density* into numerous part of the city that is touched by *Industrial, commercial, public, military and private units.* The category of the *Continuous urban fabric* is typical for the old historical part that is dominant in city centres (like in Bern). In these towns, the continuous fabric is distributed to small kernels over the whole area of the city.

The presented comparison of couples of towns shows that some towns are similar according to the structure of urban fabric or according to the types of surroundings. Also, the shape of areas is influential to the findings of similarity. Some structure of cities is typical for some countries (Cambridge and Warwick from the United

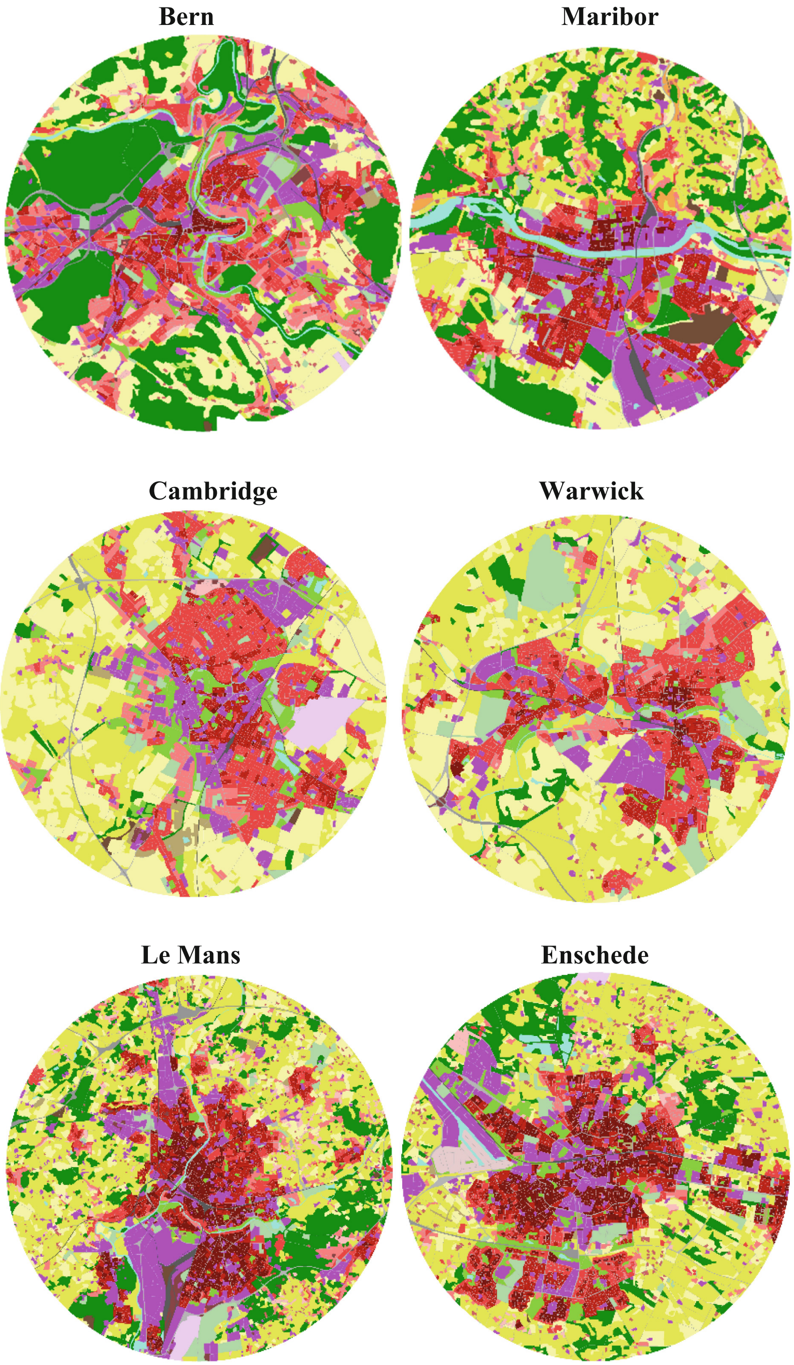


Fig. 3. Interesting similar couples of cities.

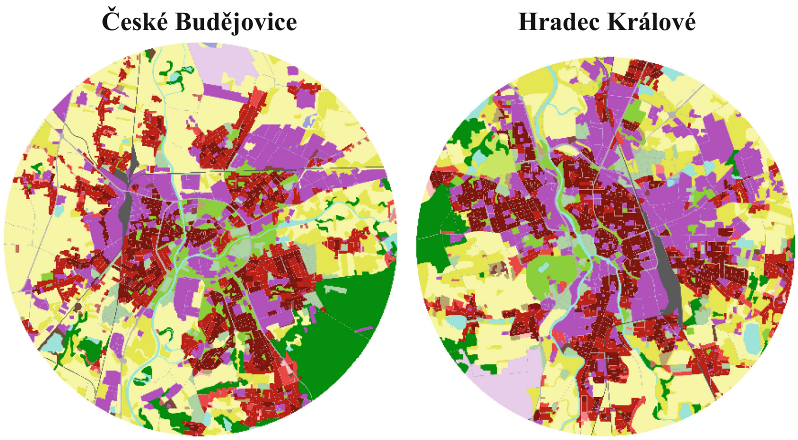


Fig. 3. (*continued*)

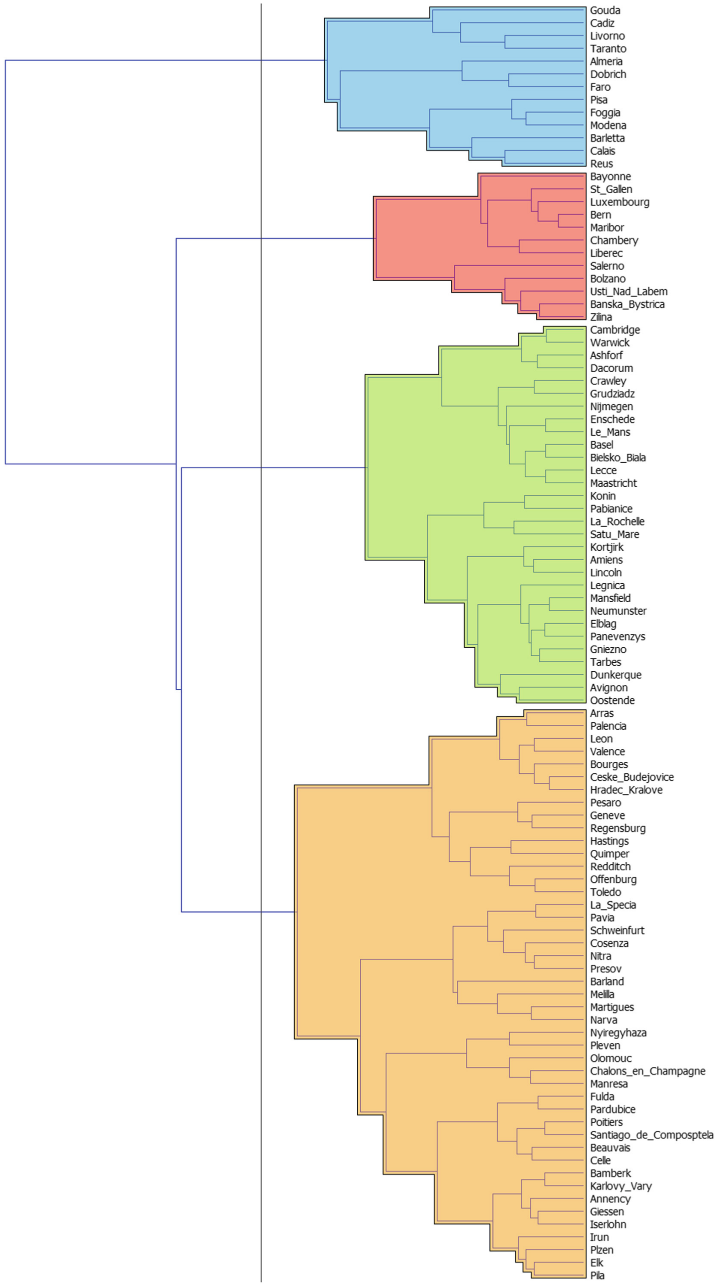
Kingdom, České Budějovice and Hradec Králové from the Czech Republic). In those cases is more influential the regional history of town evolution than the morphology of the terrain. Interesting are also presented couples of cities from different countries and parts of Europe.

The presented results show interesting couples where the similarity is produced by the same historical and economical evolution in the same country: Cambridge and Warwick in the United Kingdom, České Budějovice a Hradec Králové in the Czech Republic. In those cases, the evolution of urban grew is not limited like in some specific morphological condition of the terrain. Surprisingly it is possible also found similar couples from different part of Europe, from different countries like couple Bern and Maribor or Le Mans and Enschede. In the case of one hundred cities from Europe, it is possible to find more similar cities that are presented in the article. Furthermore, there is an opportunity to process all 800 FUA and their cities from European Urban Atlas in future.

This type of unsupervised processing data does not consider the number of inhabitants or the area size of landuse categories. Only the arrangements and categories of landuse are assumed. The using of deep learning by neural networks contribute to the assessment of the urban structure of cities. The presented results are promising for father research in the area of urban structures. Beside them, the presented steps of investigation can be used as a lecture for students at Palacky University for study branch Geoinformatics like other practical examples presented in the literature [11, 12]. Students are familiar with Urban Atlas dataset, and they attend the course Data Mining at the master level of study which is granted by the author of the article. This case study suits to be a lecture about the application of deep learning neural network as a method of data mining in the area of urban landuse.

Acknowledgement. This article has been created with the support of the Operational Program Education for Competitiveness – European Social Fund, project CZ.1.07/2.3.00/20.0170 Ministry of Education, Youth, and Sports of the Czech Republic.

Appendix 1


Gouda
Cadiz
Livorno
Taranto
Almeria
Dobrich
Faro
Pisa
Foggia
Modena
Barletta
Calais
Reus
Bayonne
St_Gallen
Luxembourg
Bern
Maribor
Chambery
Liberec
Salerno
Bolzano
Usti_Nad_Labem
Banska_Bystrica
Zilina
Cambridge
Warwick
Ashforf
Dacorum
Crawley
Grudziadz
Nijmegen
Enschede
Le_Mans
Basel
Bielsko_Biala
Lecce
Maastricht
Konin
Pabianice
La_Rochelle
Satu_Mare
Kortjirk
Amiens
Lincoln
Legnica
Mansfield
Neumunster
Elblag
Panevenzys
Gniezno
Tarbes
Dunkerque
Avignon
Oostende
Arras
Palencia
Leon
Valence
Bourges
Ceske_Budejovice
Hradec_Kralove
Pesaro
Geneve
Regensburg
Hastings
Quimper
Redditch
Offenburg
Toledo
La_Specia
Pavia
Schweinfurt
Cosenza
Nitra
Presov
Barland
Melilla
Martigues
Narva
Nyiregyhaza
Pleven
Olomouc
Chalons_en_Champagne
Manresa
Fulda
Pardubice
Poitiers
Santiago_de_Composptela
Beauvais
Celle
Bamberk
Karlovy_Vary
Annency
Giessen
Iserlohn
Irun
Plzen
Elk
Pila

References

1. Louf, R., Barthelemy, M.: A typology of street patterns. J. Roy. Soc. Interface. arXiv:1410.2094v1 (2014)
2. Open street map. https://www.openstreetmap.org. Accessed 03 May 2019
3. Boeing, G.: A multi-scale analysis of 27,000 urban street networks: every US city, town, urbanized area, and Zillow neighborhood. Environ. Plann. B: Urban Anal. City Sci. arXiv:1705.02198 (2018)
4. Urban Atlas. https://land.copernicus.eu/. Accessed 05 May 2019
5. European Environment Agency: Updated CLC illustrated nomenclature guidelines. https://land.copernicus.eu/user-corner/technical-library/corine-land-cover-nomenclature-guidelines/docs/pdf/CLC2018_Nomenclature_illustrated_guide_20170930.pdf. Accessed 13 Jan 2019
6. Eurostat database. https://ec.europa.eu/eurostat/data/database. Accessed 08 Jan 2019
7. Janousek, M.: Comparison of an urban area by circular sectors, p. 75. Diploma theses, Department of Geoinformatics, Palacky University, Olomouc (2019)
8. Esri: ArcMap for desktop. http://desktop.arcgis.com/en/arcmap/. Accessed 05 May 2019
9. University of Ljubljana: Orange. https://orange.biolab.si/. Accessed 05 May 2019
10. Orange3 image analytics documentation. https://buildmedia.readthedocs.org/media/pdf/orange3-imageanalytics/latest/orange3-imageanalytics.pdf. Accessed 05 May 2019
11. Dobesova, Z.: Teaching database systems using a practical example. Earth Sci. Inform. **9**, 215–224 (2016)
12. Dobesova, Z.: Discovering association rules of information dissemination about Geoinformatics university study. In: Silhavy, R. (ed.) Artificial Intelligence and Algorithms in Intelligent Systems, vol. 764, pp. 1–10. Springer, Cham (2019)