Draft Report
GIS Analysis techniques 2:
Mid-Term Project

1 Original study context

Green spaces are a vital component for the overall health of the population in cities. People are exposed to multiple pressures like air pollution, water pollution, high noise levels, thermal and visual stress. Therefore, the Accessible Natural Greenspace Standards (ANGSt) - a list of rules for the accessibility of green spaces in European cities - was established in the 1990s and revised in 2008. The standards for cities would be that nobody should live more than 300m away from a green space with at least 2 ha as well as further instructions for the accessibility of larger green spaces.

The paper A GIS-based analysis of the urban green space accessibility in Craiova city, Romania by Vilcea C. and Sosea C. (2020) analysed the accessibility of public green spaces in Craiova, Romania using distance bands of 0-300m, 300-500m, 500-700m and >700m, using both Euclidian Distance and Network Analysis. The paper asks how accessible public urban green spaces are to the residents and how this accessibility varies spatially across the city and its neighbourhoods.

The urban green spaces were derived from the land cover vector data from the European Corina Land Cover datasets as well as from data published by the Urban Atlas. Green areas that were missing in these datasets were then added by digitalization. Open Street Map was used for the street network. Finally, the census-based demographic data was provided by Romanian statistical authorities. The study uses ArcGIS for all of their analysis.

The main results of the paper show that large parts of Craiova have good access to public green spaces within 300-500m, but there are still significant parts of the city without access within 700m - especially within peripheral or newly built residential areas. The network-based distances reveal lower effective accessibility than the Euclidian buffers, because the street layout and barriers lengthen the walking routes. Taking everything into consideration, the authors evaluate that targeted urban planning interventions are needed to improve green space accessibility in under-served parts of Craiova (Vilcea & Sosea, 2020).

2 Reproducibility Challenges

The first challenge was choosing the best option for the dataset depending on availability - especially considering the high-resolution population dataset. The highest resolution we could get were district-based, while the paper had higher resolution and therefore the results for the population based research approach were more detailed. The original paper uses Corine Landcover data for their public green spaces but we decided to rely mainly on Open Street Map Data. Therefore we used OSM tags and rules for our study and there could arise a reproducibility challenge: small changes in the tags for OSM filters or additional manual exclusion will change the results for public green spaces. Additionally, as Open Street Map is edited over time, repeating our workflow with our code in the future may lead to different green spaces or streets.

Another key reproducibility challenge is the limited technical detail in the original paper’s methods. The study outlines the general idea (park entrances, walking network, distance bands) but not the exact algorithms or parameters for critical steps like deriving entrances, simplifying the network, or constructing accessibility zones. This forced us to choose our own implementation: automatically taking every intersection between park boundaries and streets as an entrance, inserting virtual nodes and splitting edges using distance thresholds, and building 50 m grid-based accessibility zones from interpolated network distances. These choices are clearly scripted and transparent, but they are still interpretations of the original concept rather than an exact replication. As a result, another team making reasonable but different assumptions could obtain slightly different accessibility patterns and statistics. One main difference here is that the original paper uses ArcGIS whilst we conducted the entire workflow using Python.

Despite these challenges, the core accessibility patterns described in the original study could be reproduced at a conceptual level, particularly the systematic difference between Euclidean and network-based distances.

3 Methodological Adaptations

Whilst the original paper uses land-use classes form the CORINE land cover - so the public green spaces are defined by local authorities - we decided to use mainly Open Street Map data for this mid-term. This adaptation was implemented to keep the python script more simple for us. But therefore, there is of course a bias for the category “public green spaces” since we made our own decisions on what areas to include and which not to include. However, we tried to keep it as reproducible as possible by including a clear overview on what areas were included and which not.

Another methodological adaptation is how accessibility is summarised and visualised for Graz. The original paper presents accessibility primarily through maps and statistics for administrative units based on its GIS tools, but does not describe a fully scripted, reproducible workflow. In our project, all steps—from distance computation to aggregation and cartographic layout—are implemented in Python. This moves from a more manual GIS workflow to a programmatic one, which changes implementation details but improves transparency and makes every figure directly reproducible from the code.

Due to differences in data sources, definitions of green spaces, population resolution, and implementation details, our results for Graz are not directly comparable to those of Vilcea & Șoșea (2020). However, our workflow addresses the same accessibility questions and reproduces the important qualitative pattern that network-based walkways have lower effective accessibility than simple Euclidean buffers.

4 Annexes

Data

Population:
https://www.graz.at/cms/beitrag/10034466/7772565/zahlen_fakten_bevoelkerung_bezirke_wirtschaft.html

Description of Work and Responsibilities

Generally, all group members participated in this project and there were no internal conflicts regarding responsibility and participation.

Duit, Vita / Hochreiter, Tobias

Data preparation, Data integration and Data management,
Implementation of the two analysis methods,
Visualization,
Additional organisational tasks (repository management, structuring the notebook)

Erdler, Miriam/ Fleischhacker, Yvonne

Extraction of the original workflow from Vilcea & Șoșea (2020)

Visualization,

Data cleaning, code review, removal of redundant cells, etc.
Refinement of code comments and documentation

Use of AI tools: Prompts & Reflection

Duit, Vita:

Erdler, Miriam:

For all AI prompts, Perplexity Pro was used, since it is provided by Uni Graz.

"I have the following code for mapping public green spaces in Graz by category. How can I change the colour of the three categories? [Include Code]"

→ This was a very quick and straight forward way to change the colour of the map, since the first code used automatic colours.

"Error Code: This gave the following error: ValueError: Invalid RGBA argument: nan [include my own code]."

Result of perplexity: "That error means at least one row in public_green_categories['category'] is NaN or not in your color_map, so .map() returns NaN which Matplotlib cannot use as a color. Use a safe mapping with a fallback color:"

→ helped me understand the error better and work through the bug.

"How can I make the legend box a bit bigger with bigger font size for this plot: [include my own code]".

→ very fast and straight forward answer with the new code.

Fleischhacker, Yvonne:

Hochreiter, Tobias:

AI Prompts: of Tobias and Vita

Perplexity:

i want to do an analysis of green space accessibility and also an analysis of how green a district / area is. so for the accessbility, i need all public parks and urban forests, and for how green the district is, i would also consider private gardens, meadows, grass, etc. can you write me a query to load all of these features?

→ worked very well instantly, working query to load features

what about graveyards and alleyways? are they relevant for the green area analysis?

→ gave us arguments for keeping graveyards and leaving alleyways out, we acted accordingly

Claude:

i have a gdf with green spaces in Graz dissolved into 1 mulitpolygon and a street network of graz. i want to find access points of the green spaces and was thinking of intersecting the green spaces with the edges of the network to get these? how to do this?

→ quite a lot of debugging but worked well

Claude:

hello! i have a network graph of the city of graz and the access points to green areas (intersections between edges of the network and green areas). now i want to do one to all routing to the access points to make buffers of walking distances of different distance steps. now i want to know how i can do a routing, because the access points are not really in the network... so i cant directly do the routing. i was thinking of finding nearest nodes in the network that are closest to the access points. but then the nearest node might not be representative to the real access point, because it might find a node behind a fence or something as the nearest node. can you think of any option how i can do a routing to the access points?

→ more complicated task but worked well with some debugging and more questions about details in the code

Claude for accessibility map (after deciding buffers are not suitable):

maybe buffers are not the best choice. is there another way to separate the area into parts that are within walking distance of 100, 200 and 300m? maybe by checking the value of access_dist for most of the area? so i could make a map that shows areas that are within these distances and those which are not and also to have some areal statistics results

→ quite a lot of debugging and asking follow-up questions necessary, but good result in the end that I wouldn't have come up with without the input

Claude for "general green-ness":

Now i also want to do an anlaysis of general greenness in the districts of graz. i have green_spaces from osmnx, which includes all green spaces in graz, and i have gdf_residential also from osmnx and gdf_districts with the districts of graz.

can you write me a short script with the following outputs:

* percentage of green area (green_spaces) in the whole of graz (gdf_graz)

* percentage of green area in each district

→ worked very well, no debugging necessary, made it a lot easier and faster than typing everything by ourselves

Perplexity:

now i want to calculate deficient areas with my output zones_gdf like i did it here:

[code deficient areas euclidean distances]

how to do this with my output zones_gdf?

→ worked instantly