A Micro-Scale Walkability Metric for Pleasant Pedestrian Route Planning

George Taylor, Kaloyan Karamitov, Dessislava Petrova-Antonova

GATE Institute, Sofia University "St. Kliment Ohridski", 1164 Sofia, Bulgaria – george.taylor@gate-ai.eu, kaloyan.karamitov@gate-ai.eu, dessislava.petrova@gate-ai.eu

Keywords: Walkability Metric, Pedestrian Route Planning, Micro-scale Analysis, Multi-criteria Decision Analysis, Streetscape Indicators.

Abstract

This paper proposes a micro-scale walkability metric based on harmonised indicators that supports pedestrian route planning, which prioritises pleasant environments alongside distance efficiency. The employed method quantifies street segments and crossings using geospatial indicators, including pavement width, slope, shade, adjacency to traffic, park context, and crossing type and width. Indicator values are transformed into percentile ranks to harmonise heterogeneous inputs and aggregated into a single edge-level walkability score on a 0-1 scale. The score is integrated into a routing cost function that reduces edge costs with higher walkability, favouring calmer, greener, and wider links while bounding detours relative to the shortest path. The method also accommodates the incorporation of street-level perceptions through a structured survey instrument and a confidence-weighted fusion scheme. The results show various spatial patterns. Central areas and park-adjacent segments exhibit higher scores, while steep, narrow, and traffic-exposed links score lower, and several suburban and foothill districts display reduced walkability. The comparison with a distance-only baseline shows selection of quieter alignments with modest length increases, indicating potential gains in perceived pleasantness.

1. Introduction

In urban planning, there has been an increasing focus on discouraging car use and promoting more sustainable modes of transportation, such as walking, cycling, and public transportation. The health benefits of walking and cycling (Avila-Palencia et al., 2018; Pucher et al., 2010), as well as the reduction in health and environmental harms of motor vehicles (Manisalidis et al., 2020; Piras et al., 2024), make such a mode shift desirable. There has been extensive research into quantifying the extent to which urban environments facilitate walking through the measurement of 'walkability'. Walkability metrics can be used to target streetscape improvements where they will have the greatest positive impact, making them a valuable tool for increasing the attractiveness of walking as an urban transport option. The focus of this paper has been on developing a micro-scale walkability metric for Sofia, which is then used to find pleasant walking routes.

There has been extensive research into the negative health effects of inactivity, one of which is an increased risk of cancer (Friedenreich et al., 2021). Since cities with higher rates of walking and cycling tend to have a greater proportion of their population meeting recommended activity levels, policies that increase walking rates are likely to yield public health benefits (Pucher et al., 2010). Active transport modes have also been linked to mental health benefits, namely improvements in perceived health, stress, mental well-being, and quality of life (Avila-Palencia et al., 2018).

The prevalence of automobile use has health implications beyond the effects of inactivity. Air pollution can cause multiple adverse health effects, such as damage to the respiratory system and an increased risk of cancer (Manisalidis et al., 2020). Particulate matter (PM) is a significant source of pollution that can have severe health effects. While a shift from internal combustion engine (ICE) vehicles to electric vehicles eliminates exhaust emissions, most PM emissions come from non-exhaust factors, such as tyre and road wear (Piras et al., 2024), so even with a large-scale shift to electric vehicles, PM pollution remains a problem. Therefore, a shift to active travel modes would be one of the most effective methods to tackle PM pollution (Piras et al., 2024).

In Bulgaria, the transport sector accounts for 22.5% of total CO2 emissions (International Energy Agency, n.d.), so reducing motor vehicle use is crucial to reducing overall emissions. Additionally, Bulgaria has the highest rate of premature deaths from air pollution in all of Europe (European Environment Agency, 2025). Therefore, it is vital to reduce transportation-related pollution. Air pollution has both economic and health impacts, with an estimated 13.4% of Sofia's local GDP lost in 2019 (Deloitte, 2021).

This study aims to develop a micro-scale walkability metric for Sofia based on geospatial and quantitative survey data. The main contributions can be defined as follows:

- Proposing a comprehensive walkability metric for finding walking routes that lead to greater resident satisfaction, compared with the shortest route.
- Combining geospatial and survey data in calculating the walkability metric to achieve greater fidelity than a purely GIS-based approach and greater coverage than a metric based solely on observational audits.

The rest of the paper is organised as follows. Section 2 reviews related work on macro- and micro-scale walkability metrics and audit instruments. Section 3 introduces the study area, describes the datasets, and outlines the data preparation and processing pipeline. Section 4 presents the methods and algorithms for computing the edge-level walkability score and integrating it into the routing cost model. Section 5 reports the empirical results, including city-wide patterns and illustrative route comparisons against a distance-only baseline. Section 6 concludes with limitations and directions for future work.

2. Related work

Most existing walkability assessment tools fall into two categories – macro-scale GIS-based metrics and micro-scale observational audits. Macro-scale metrics typically depend on data such as land use patterns and population density (Huang et al., 2025). Micro-scale audits typically consist of observational surveys carried out by trained auditors or members of the public, focusing on factors such as pavement quality and perceived safety (Huang et al., 2025).

A micro-scale metric is best suited for generating pleasant walking routes, as it can provide ratings per street. Macro-scale metrics typically consider only factors that indicate a propensity to walk, rather than those that indicate a pleasant walking experience (Huang et al., 2025). Micro-scale metrics can better capture this, as they can include aesthetic factors (Huang et al., 2025). A major issue with existing micro-scale audit approaches is that they require trained auditors and can take considerable time to survey an entire area, making them prohibitively expensive for assessing a city's walkability. To address this issue while still providing micro-scale data, a hybrid approach combining GIS data and observational methods could be employed. A study into this approach by Lee and Talen concluded that such a combination of methods "can be an effective way of obtaining physical environment data that is very comparable to in-person observation” (Lee & Talen, 2014).

Existing metrics include a range of factors, with some audits including over 200 individual items (Day et al., 2006). To facilitate data gathering, some auditing tools, such as the Microscale Audit of Pedestrian Streetscapes (MAPS), have developed shortened versions that focus on their most essential factors (Sallis et al., 2019). Because many auditing tools lack studies on their predictive validity, which rely instead on face and construct validity, some factors may not have a significant effect on observed walking rates. In a study by Boarnet et al. on the predictive validity of the popular Irvine-Minnesota Inventory (IMI) tool, it was found that a significant portion of the measured factors had no noticeable effect on walking rates in the USA (Boarnet et al., 2011). They concluded that a version of the IMI cut down to its most important factors could perform similarly well to the full tool in predicting walking rates.

In a review of the literature on quality-aware pedestrian route finding, Siriaraya et al. found that a common approach was to use Dijkstra’s shortest-path algorithm with a modified cost function (Siriaraya et al., 2020). Data items representing factors such as safety and aesthetics are often mapped onto streets or areas (Siriaraya et al., 2020). They can then be used to adjust the cost of edges in the routing network. An alternative approach was taken by Quercia et al. for finding pleasant walking routes in London, UK (Quercia et al., 2014). This approach used Eppstein’s k-shortest path algorithm to generate multiple candidate paths, which are then iteratively ranked based on beauty, quietness, or happiness.

3. Study Area and Data

3.1 Study Area

Sofia, the capital of Bulgaria, is situated in a basin at the foot of Vitosha Mountain. Terrain conditions range from flat central plains to steeper southern and southwestern foothills, which influence pedestrian comfort through slope. The historical core features fine-grained blocks, dense intersections, and an abundance of services and transit, all of which support higher walkability. Large arterial corridors and superblocks in postwar housing complexes increase exposure to traffic and reduce pedestrian comfort, while major parks and linear green corridors create localised high-scoring segments.

The analysis covers the entire Sofia Municipality, which includes the continuous urbanised city, suburban settlements, and rural areas where pedestrian networks are sparser and formal sidewalks are less common. Elevation data enables municipality-wide slope estimation, while enriched streetscape attributes, such as shade, are more complete in the urbanised area. This approach supports comparable edge-level metrics despite differences in morphology, infrastructure, and data completeness.

3.2 Data Collection and Preprocessing

The pipeline outlined in the following sections converts heterogeneous city datasets into a routable pedestrian graph with edge-level walkability attributes. Coordinate reference systems and schemas are standardised, municipal layers are conflated with OpenStreetMap, and a topologically sound network is constructed. Geometry corrections include splitting rings and enforcing node–edge connectivity. Attribute enrichment attaches elevation and slope, joins road context such as speed and lanes, assigns park and shade context, and derives width and crossing properties, followed by normalisation to percentile scores. Quality control applies topology checks, coverage diagnostics, and missing-data flags to avoid bias. The workflow is reproducible and versioned, implemented using PostGIS and pgRouting for graph operations, as well as Python tools such as GeoPandas and Rasterio. Datasets either contain data for the whole municipality or for the main urbanised area of Sofia.

The pedestrian network dataset includes MultiLineString geometry representing pedestrian infrastructure in Sofia. It went through six steps of pre-processing as follows:

- Three lines composed of more than one section were identified and manually corrected. This involved removing unconnected or duplicated sections and combining continuous sections.
- The geometry type for every line was changed from MultiLineString to LineString.
- All rings where a line begins and ends at the same point were split into separate lines. Rings that did not intersect with other lines were split into two equal halves. Rings that intersected with other lines at a single point were split into two equal halves such that both halves joined at that point. Rings that intersected with other lines at more than one point were split into sections at those points.
- Lines that were not fully contained within the boundary of Sofia Municipality were removed.
- A graph was constructed by taking lines as edges and their endpoints as nodes. Where the endpoints of multiple lines were within 0.1 metres of each other, they were merged into a single node.
- All edges that were not part of the largest connected component were removed. This removed 318 lines, constituting 0.25% of the dataset. Disconnected components were largely due to poor data quality on the city's outskirts.

The preprocessing results in each entry in the cleaned dataset representing a single non-self-intersecting edge. Furthermore, it is guaranteed that a path can be found between any two arbitrary points in the graph.

Additional datasets used in the study are as follows. All datasets use the BGS2005 (EPSG:7801) coordinate reference system, except where otherwise stated:

- Parks and gardens - MultiPolygon geometry for every park in Sofia Municipality.

https://doi.org/10.5194/isprs-annals-XI-4-2026-369-2026 | © Author(s) 2026. CC BY 4.0 License.

- Neighbourhoods - MultiPolygon geometry for every neighbourhood in Sofia Municipality.
- Street lights - MultiPoint geometry to represent every light fitting in Sofia Municipality.
- Street polygons - contain all roads and paths in Sofia Municipality with Polygon geometry that represents the road surface.
- OSM roads – LineString geometry for all roads in Sofia Municipality containing information about the maximum speed limit in km/h and the number of lanes. This dataset uses the WGS84 (EPSG:4326) coordinate reference system.
- Shade - raster dataset containing information about the level of shade on the street polygons. The data is a snapshot recorded on 22 August 2018, so it may be inaccurate for other periods of the year.
- Map of heights - raster dataset containing heights above sea level with a resolution of 1m^2 per pixel.

4. Methods and Algorithms

The walkability metric calculation method described in this section yields a single 'walkability score' for every edge in the cleaned pedestrian network dataset. It was chosen to yield a single score rather than multiple components to facilitate easier planning of walkable routes. The final score is in the range 0 to 1, where 0 denotes the 'worst' walkability and 1 denotes the 'best'. The score is a weighted mean between separately calculated geospatial and survey percentile scores.

The pedestrian network dataset is divided into 'segments' and 'crossings', and the score for each edge type is calculated separately using the same method, based on different factors. The dataset is split based on the 'type' property on the edges. Edges with a type of 'Нерегулирано' (unregulated), 'Светофар' (traffic light), or 'Пешеходна пътека' (pedestrian crossing) are classified as crossings, while all other edges are classified as segments. The dataset was split this way because at-grade crossings where pedestrian flows conflict with car movements have different relevant factors than other edges.

The metric incorporates factors commonly used in previous walkability studies (Day et al., 2006; Sallis et al., 2019; Al Shammas & Escobar, 2019). Only the survey-based factors deemed most relevant were used to allow the survey to be completed quickly by untrained evaluators. This can perform similarly well to a longer survey (Boarnet et al., 2011).

Each factor in the metric is awarded a score for one (or more) criteria. The three criteria for segments are comfort, accessibility, and safety (see Table 1 and Table 2).

Factor	Comfort	Accessibility	Safety
Steepness	-	1	-
Speed limit of nearby road	0.5	-	0.5
Number of lanes on nearby road	0.5	-	0.5
Whether segment is within a park	1	-	-
Shade	1	-	-
Pavement width	0.5	1	0.5
Lighting	-	-	1

Table 1. Geospatial factors of segments

Factor	Comfort	Accessibility	Safety
Pavement condition	-	2	-
Obstacles	-	2	-
Presence of benches	0.5	0.5	-
Presence of greenery	1	-	-
Cleanliness	1	-	-
Crowdedness	1	-	1
Condition of surrounding buildings	0.5	-	-
Volume of vehicle traffic on nearby road	0.5	-	0.5
Separation between pavement and road	0.5	-	0.5

Table 2. Survey factors of segments

The two criteria for crossings are accessibility and safety (see Table 3 and Table 4). Comfort is not considered separately from safety because comfort at crossings tends to be closely linked to perceived safety (Gill et al., 2022), so it is less beneficial to have it as a separate criterion. The factors that comprise the metric, along with the maximum scores that can be awarded to each criterion, are presented in the following tables.

Factor	Accessibility	Safety
Crossing type	2	2
Crossing width	1	1

Table 3. Geospatial factors of crossings

Factor	Accessibility	Safety
Markings	-	1
Obstructions	1	1
Adaptations for people with reduced mobility or disabilities	2	-
Wait time	1	-
Volume of passing vehicle traffic	-	2

Table 4. Survey factors of crossings

4.1 Geospatial Score Calculation

Each edge is first given a score in the range 0 to 1 for each factor. The following section outlines the calculation of scores for each factor, concluding with details on how the final geospatial score is determined.

4.1.1 Steepness score calculation: Using the heightmap raster dataset and the Rasterio Python package, heights above sea level are calculated for the endpoints of every segment. The grade of a segment k is then calculated using the equation:

$$grade_k = \frac{start\_elevation_k - end\_elevation_k}{length_k} \quad (1)$$

The absolute value is then taken for every grade value. This is done so that each edge has a single steepness score. For future work, it would be beneficial to assign different scores depending on whether the user is walking up or down an edge, since pedestrians typically find walking downhill easier (Meeder et al., 2017). The final 0 to 1 steepness score for a segment k is calculated by the equation:

$$steepness_k = 1 - (0.01 \times percentile\_rank(G, g_k)) \quad (2)$$

where g_k denotes the grade of segment k
G denotes the set of all absolute segment grade values.

percentile_rank(G,g_k) gives the percentage of elements in G that are less than or equal to g_k

4.1.2 Speed limit & road lane score: First, the road dataset is transformed to the BGS2005 (EPSG:7801) coordinate reference system to match the pedestrian network dataset. Using the ‘sjoin_nearest’ function from the GeoPandas Python package, segments are matched with the nearest line in the OSM road dataset that has the same ‘name’ attribute. Road name is considered, as well as distance, to avoid a segment parallel to a major roadway getting joined with a side street that appears closer. Scores are calculated based on the road information using a pre-defined mapping. The mappings are given below, where *{k: v}* means that the score of v is assigned to value *k*.

Speed limit mapping – {20: 0.8, 30: 0.7, 40: 0.6, 50: 0.5, 60: 0.3, 70: 0.1, 80: 0, 90: 0. 140: 0}

Lanes mapping – {1: 0.8, 2: 0.7, 3: 0.6, 4: 0.5, 5: 0.3, 6: 0.1, 7: 0}

Road information is fetched only for segments on or near roads. This is determined based on having a ‘type’ attribute such as ‘тротоар’ (pavement) or ‘Пътно платно’ (roadway). Segment types which are typically not near roads, such as ‘Алея‘, are given a score of 1 for both speed limit and lanes. Segments without a defined type are assigned a default score of 0.5 for both factors.

Due to incomplete OpenStreetMap data, some segments near roads may lack the required road information. In these cases, the standard speed limits of 50 km/h in urban areas and 90 km/h in non-urban areas (Your Europe, 2024) apply, except for segments of type ‘паркинг‘ (parking), which default to 30 km/h. Parking lots have a reduced default speed because it is assumed that cars will generally be travelling slower than on a roadway. Whether a segment is within an urban area is determined using the neighbourhoods dataset. This dataset contains a polygon representing all areas outside a neighbourhood, so any edge that intersects it is deemed not to be within an urban area. For road lanes, a default value of 2 is used because it is assumed that most roads are 2 lanes wide, and that most roads wider than 2 lanes are main roads, which are more likely to have complete information on OpenStreetMap.

4.1.3 Park score: The parks and gardens dataset contains polygons representing all parks in Sofia. A segment is deemed to be within a park if it intersects with one or more of the polygons in the parks dataset. Segments within parks are assigned a score of 1, while those outside parks are assigned a score of 0.

4.1.4 Shade score: The shade score calculation utilises the street polygon dataset and the shade raster dataset. First, polygons are assigned a ‘polygon_shade’ value, which is calculated by taking the mean of every pixel in the shade dataset that a polygon overlaps. This is done using the Rasterio Python package. Polygons that do not overlap any pixels in the dataset are assigned a null value, which is handled later in the process.
Next, shade information from the polygon dataset must be mapped onto the segments. Segments are categorised into four types: pavements, parking lots, roadway segments, and others. A 5-metre buffer is drawn around every polygon in the street polygon dataset. For pavements, parking lots, and segments on the roadway, each segment’s intersections with the buffered polygons are found. A buffer is used because some segments in the pedestrian network dataset are offset from the street polygon they serve and therefore do not directly overlap it. The intersections are found using the ‘overlay’ function from GeoPandas, which produces new line geometry representing the portion of the line that overlaps a polygon. The length of this new geometry is compared against the full length of each segment, and only intersections which are greater than half the length of the full segment are retained. This is done to ensure that segments only take data from polygons that are parallel to them.

Once the intersections between lines and polygons have been found, the intersection geometries are grouped based on their segment *id*. In cases where a segment has intersected with multiple polygons, the mean of all the intersecting polygons’ shade values is taken. For pavement segments, if they have intersected with any polygons that are tagged with a codename including ‘тророар‘ (pavement), only those polygons are used in calculating the mean. The same is done for parking segments for polygons tagged with ‘паркинг’ (parking). This ensures that information is extracted from the most relevant polygons where available, but that a value can still be calculated even if they are not present. For example, some pedestrian segments tagged ‘тротоар’ (pavement) do not have a corresponding pavement polygon in the polygon dataset, and instead only have a parallel roadway polygon.

Now that shade values have been added to segments, the separate categories are recombined. Any segment without shade data is given a null value. For segments with a non-null shade value, the shade score is calculated as the percentile rank relative to all other segments with a non-null shade value, divided by 100. Segments with a null shade value, such as segments that were in the ‘other’ category, are given a default score of 0.5. The ‘other’ category included segments which are not near any roads and therefore do not have data in the shade dataset, since it only includes values for roads in Sofia.

4.1.5 Pavement width score: The width score is calculated based on the ‘width_cl’ attribute in the pedestrian network dataset. For segments with a non-null ‘width_cl’ value, their width score is calculated by finding their percentile rank compared to all other segments with a non-null ‘width_cl’ value and dividing it by 100. If a segment has ‘type’ attribute ‘Тротоар’ (pavement), but has a null value for ‘width_cl’, it gets assigned a default score of 0.5. Segments with a ‘type’ attribute that indicates they are away from roads, such as ‘Алея’, get assigned a default score of 1. Finally, any other segment gets assigned a score of 0. This default score is low because most segments that get assigned it are on roads without pavements. These segments can be identified by having a ‘type’ attribute, such as ‘Пътно платно’ (roadway). The aim of this factor is to penalise segments with narrow or no pavements, while rewarding segments which are away from roads or have wide pavements.

4.1.6 Lighting score: The light fittings dataset contains points representing streetlights in Sofia. A segment is deemed to be lit if there is at least one streetlight within a 10m buffer zone around the segment. Lit segments are given a score of 1, while unlit segments are given a score of 0.

This score provides only a rough estimate, as no data are available on the brightness of the streetlights. There is no way to know how well a segment is lit, since one streetlight could be brighter than multiple different streetlights together, so merely counting the number of streetlights is insufficient. The method given above is somewhat lenient on what qualifies as ‘lit’, so we can be fairly confident that a segment marked as ‘unlit’ has been correctly labelled.

https://doi.org/10.5194/isprs-annals-XI-4-2026-369-2026 | © Author(s) 2026. CC BY 4.0 License.

4.1.7 Crossing type score: Crossing type is determined based on the ‘type’ attribute on the crossings. The score is assigned as follows:

- ‘Нерегулирано’ (unregulated) – 0
- ‘Пешеходна пътека’ (pedestrian crossing) – 0.5
- ‘Светофар’ (traffic light) - 1

4.1.8 Crossing width score: Here, ‘crossing width’ refers to the distance a pedestrian must walk to cross the road, so a crossing’s width is determined based on its ‘length_m’ attribute. The score is determined by taking the percentile rank of a crossing’s width and dividing it by 100 to get it in the range 0 to 1.

4.1.9 Final geospatial score: Once the individual factor scores have been calculated, an edge’s ‘raw geospatial score’ is calculated as follows:

$$raw\_geospatial\_score = \sum_{k=1}^{n_f} \sum_{c=1}^{n_c} w_{k,c} \times s_k \tag{3}$$

where n_f gives the number of factors
n_c gives the number of criteria
$w_{k,c}$ denotes the maximum score for criteria c for factor k
s_k denotes the score for factor k.

Finally, an edge’s ‘percentile geospatial score’ is found by calculating the percentile rank of its raw score compared against the raw scores of all other edges of its type (segments or crossings) and dividing it by 100 to get the value in the range 0 to 1.

4.2 Survey Score Calculation

First, a score must be calculated for each factor. Each survey answer is assigned a specific score. There are two questions in the survey which are only asked for segments next to roads. For segments that are not adjacent to roads, the scores for those two factors default to the maximum score. There are also two generic questions which are not included as factors in the score calculation but are asked to gather extra data on the usefulness of the metric. The score for each factor is calculated as follows:

$$s_f = \frac{\sum_{k=1}^{n_a} \sum_{c=1}^{n_c} a_{k,c}}{n_a} \tag{4}$$

where s_f denotes the score for factor f
n_a gives the number of answers
n_c gives the number of criteria
$a_{k,c}$ denotes the score for criterion c given by answer k.

The ‘raw survey score’ for an edge is then calculated with the equation:

$$raw\_survey\_score = \sum_{f=1}^{n_f} s_f \tag{5}$$

where n_f gives the number of factors.

A raw survey score can be calculated only for edges with sufficient survey responses. Constant value ‘min_survey_answers’ gives the minimum number of responses to each survey question an edge needs to have for it to have a raw score calculated. Edges that do not meet this threshold are assigned a null value.

The ‘survey percentile score’ is calculated similarly to the geospatial percentile score, the percentile rank is compared only against edges with a non-null raw score. Edges which have a raw score of ‘null’ are assigned a percentile score of ‘null’.

4.3 Final Score Calculation

The final walkability score is calculated using the following set of equations:

$$walkability = (geospatial\_weight \times geospatial\_percentile) + (survey\_weight \times survey\_percentile) \tag{6}$$

$$survey\_weight = \min\{1, survey\_confidence\} \times survey\_max\_weighting \tag{7}$$

$$survey\_confidence = \frac{min\_answer}{survey\_confidence\_threshold} \tag{8}$$

$$geospatial\_weight = 1 - survey\_weight \tag{9}$$

where *min_answer* denotes the smallest number of answers to any of the survey questions
survey_confidence_threshold is an integer constant
survey_max_weighting is a constant in the range 0 – 1.

If the survey percentile score for an edge is ‘null’, then its walkability score is set equal to the geospatial percentile score. Conceptually, the walkability score is a weighted mean of the geospatial and survey percentile scores, with the survey score's weight increasing as an edge receives more survey responses. Increasing the constant ‘s*urvey_confidence_threshold*’ increases the number of survey responses required for the survey component to have full weighting. Increasing the constant ‘*survey_max_weighting*’ increases the survey component's weight in the final score. Therefore, by changing the two constants, the balance between the geospatial and survey scores in the metric can be adjusted.

4.4 Route Finding

Both route-finding approaches mentioned in Section 2 were considered. Due to the size of the pedestrian network dataset, the approach using Eppstein’s algorithm had a prohibitively high run-time. Therefore, the approach using Dijkstra’s algorithm and a modified cost function was chosen. The cost function was designed to balance walkability and route length, and is defined as follows:

$$edge\_cost = edge\_length \times (1 - (walkability \times cost\_reduction\_factor)) \tag{10}$$

The ‘cost reduction factor’ is a constant in the range 0 to 1 which controls the magnitude of the effect that walkability has on edge cost. For example, with a cost reduction factor of 0.3, the lowest an edge’s cost can be reduced to is 70% of its length. This means that any route found will be no more than around 42% longer than the shortest path by distance.

The routing method was implemented by developing an API endpoint using the FastAPI Python package. The developed API also supports submitting survey data. Walkability scores and edge weights are recalculated periodically, allowing newly received survey responses to be incorporated into the values. A walking route is provided as a list of longitudes and a list of latitudes, which specify the points along the desired route. After validating the provided points, *pgRouting* function *pgr_withPoints* is used to find a route between each pair of points. These routes are combined into one route. The endpoint returns a GeoJSON Feature with the route as its geometry. The function *pgr_withPoints* is used rather than Dijkstra’s algorithm (pgr_dijkstra) because it allows routes to be found between arbitrary points on edges, rather than only between graph nodes.

5. Results

The results shown in this section are calculated solely from geospatial data, as it was not possible to collect a large volume of survey data.

5.1 Pedestrian Network Score

Figure 1 illustrates how parks are highly scored by the metric. It can also be seen that walkability is not equally distributed among the suburbs. Suburbs such as Lyulin and Nadezhda are rated quite highly, while suburbs in the southwest, such as Gorna Banya, are rated poorly. The neighbourhoods in the foothills of Vitosha Mountain, such as Simeonovo and Dragalevtsi, are also rated poorly. The mentioned neighbourhoods are shown in Figure 2.

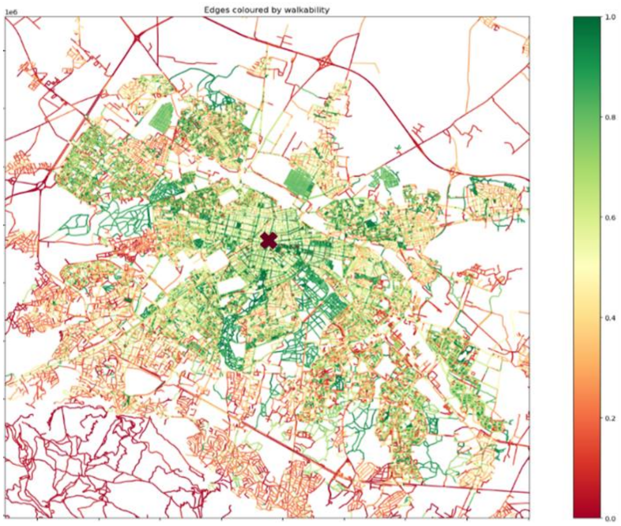

Figure 2. Scores of pedestrian network edges for the whole city. The cross marks the Serdica metro station as a reference point.

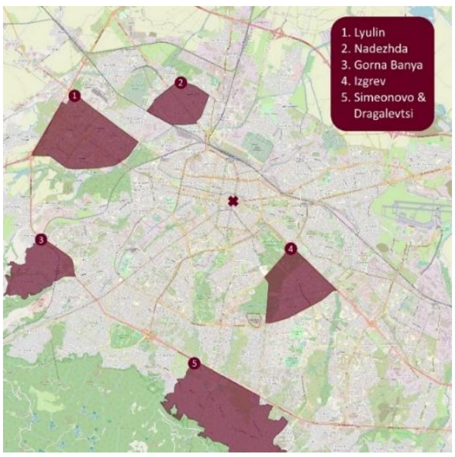

Figure 2. Notable neighbourhoods in Sofia. The cross marks the Serdica metro station as a reference point.

It can be observed that walkability is highest in the city centre, generally lower in the suburbs, and lowest in rural areas. This pattern is largely expected and has been found in cities throughout Europe (Patel et al., 2025). However, it must be noted that rural areas may have an unfairly low score under the developed metric, as some datasets – such as the shade dataset – cover only the urbanised area of Sofia rather than the whole Sofia municipality. The paths around Vitosha Mountain, shown in the bottom left of Figure 1, are rated poorly by the metric, despite being popular walking routes. This emphasises that the metric focuses on utilitarian walking rather than recreational walking. The scores for similar walking paths may increase with survey data, as their appeal largely depends on aesthetic factors not captured in the geospatial data.

The coloured map of the Izgrev district, shown in Figure 3, provides a closer look at the walkability results. Notably, the edges around the main road in the centre of the figure are rated poorly. This was because these edges are mostly quite narrow, lack shade, are on a slope, and are near a wide main road. Some neighbourhood streets are rated poorly because of narrow pavements, while others are rated highly because of good shade coverage and low speed limits on adjacent roads.

Figure 3. Scores of pedestrian network edges for District Izgrev.

It is helpful to compare these maps against previous walkability studies in Sofia. By comparing Figure 1 with a map from a previous walkability study by the GATE Institute (Petrova-Antonova et al., 2025), similar patterns emerge. The patterns observed in the previous paragraphs are mostly replicated between the two maps. This is despite the fact that the previous study was focused on access to points of interest (POIs), a factor not included in the metric described in this paper. This is in line with expectations, as macro- and micro-level walkability factors have been shown to be correlated (Koo et al., 2022). Although both walkability measures exhibit similar aggregate trends, they differ at the micro scale.

5.2 Route Finding

Although time restrictions prevented a thorough evaluation of the walkability route-finding functionality, some example routes were generated to provide a rough comparison between the route-finding algorithm outlined in this paper and the shortest path by distance calculated by Dijkstra’s algorithm. Most routes generated with a cost reduction factor of 0.3 were very similar across the two approaches. This indicated that this constant may have been set too low, since paths did not vary much from the shortest path.

To investigate the effect of the cost reduction factor on the resulting routes, a Monte Carlo simulation was run. This involved randomly selecting two points in the routing network, calculating a path using both Dijkstra's algorithm and the algorithm defined in this paper, and then finding the percentage increase in length of the 'walkable' path relative to the shortest path. This was repeated for 500 iterations at multiple values of the cost-reduction factor. The results are summarised in Figure 4.

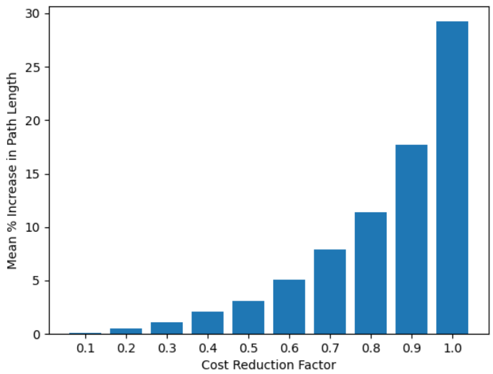

Figure 4. Estimated effect of the cost reduction factor on path length.

As expected, the routes calculated with a higher cost reduction factor were longer and varied more from the shortest path than those calculated with a smaller cost reduction factor. Since users have varying priorities for distance and comfort, it would be beneficial to make the cost reduction factor user-adjustable. One route that showed a difference when using a cost reduction factor of 0.3 is shown in Figure 5 and Figure 6.

Figure 5. Route calculated based on the proposed method.

In the second half of the route, the shortest path routes the user via the main road, whereas the proposed algorithm instead routes the user through the parallel park. Although 'pleasantness' is subjective, it is generally agreed that paths through parks are more pleasant than paths near major roads (Sundling & Jakobsson, 2023). While concrete conclusions cannot be drawn from a single example, this suggests that using a walkability metric can enhance the pleasantness of generated routes.

Figure 6. Shortest path by distance calculated with Dijkstra's algorithm.

6. Conclusion and Future Work

This paper proposes a micro-scale walkability metric for Sofia, aiming to identify pleasant walking routes. The results from its walkability analysis show that the south-western suburbs of Sofia are generally the least walkable part of its main urban area, while the centre is the most walkable. The effectiveness of combining geospatial and survey data could not be evaluated due to insufficient survey data. However, based on geospatial data alone, there is preliminary evidence that using a walkability metric can be effective in finding pleasant walking routes. This gives strong motivation for future work in this area.

Future work considers several key points. Further analysis would benefit from the collection of survey data. This would allow for a full evaluation of the method for combining geospatial and survey data. This includes validating weights and validating the confidence-weighted scheme for small-sample survey data. Additionally, inter-observer reliability could be assessed. A study into perceptions of routes calculated using the metric would be beneficial to assess the hypothesis that generated routes will be more pleasant than the shortest path by distance. Since the weights of factors in the metric are fixed, it would be beneficial to develop multiple user profiles with different weighting schemes to better fit the heterogeneous needs of different user groups. The metric's accuracy is limited by the age of some of the datasets. For example, the pedestrian network dates back to 2020, while the shade dataset dates back to 2018. This is especially relevant for the analysis of rapidly developing areas of Sofia, such as the southern suburbs. Some of the metric's scoring criteria could be refined. For example, steepness scores would benefit from distinguishing between uphill and downhill. Additionally, seasonal variations should be included in shade scores. Previous work by Al Shammas and Escobar provides a potential approach for this (Al Shammas & Escobar, 2019). Since the metric was designed specifically for Sofia, based on available data, future work could adapt the framework and apply it to other cities with different data.

Acknowledgements

This research is part of the GATE project funded by the Horizon 2020 WIDESPREAD-2018-2020 TEAMING Phase 2 programme under grant agreement no. 857155 and the programme "Research, Innovation and Digitalization for Smart Transformation" 2021-2027 (PRIDST) under grant agreement no. BG16RFPR002-1.014-0010-C01, the deployEMDS project, funded by the Digital Europe Programme DIGITAL-2022-CLOUD-AI-03 under grant agreement no. 101123520 and the CU project, funded by the Digital Europe Programme DIGITAL-2023-CLOUD-AI-04 under grant agreement no. 101168005.

References

Al Shammas, T., & Escobar, F. (2019). Comfort and Time-Based Walkability Index Design: A GIS-Based Proposal. International Journal of Environmental Research and Public Health, 16(16), 2850. https://doi.org/10.3390/ijerph16162850

Avila-Palencia, I., Int Panis, L., Dons, E., Gaupp-Berghausen, M., Raser, E., Götschi, T., Gerike, R., Brand, C., de Nazelle, A., Orjuela, J. P., Anaya-Boig, E., Stigell, E., Kahlmeier, S., Iacorossi, F., & Nieuwenhuijsen, M. J. (2018). The effects of transport mode use on self-perceived health, mental health, and social contact measures: A cross-sectional and longitudinal study. Environment International, 120, 199–206. https://doi.org/10.1016/J.ENVINT.2018.08.002

Boarnet, M. G., Forsyth, A., Day, K., & Oakes, J. M. (2011). The Street Level Built Environment and Physical Activity and Walking. Environment and Behavior, 43(6), 735–775. https://doi.org/10.1177/0013916510379760

Day, K., Boarnet, M., Alfonzo, M., & Forsyth, A. (2006). The Irvine-Minnesota inventory to measure built environments: Development. American Journal of Preventive Medicine, 30(2), 144–152. https://doi.org/10.1016/j.amepre.2005.09.017

Deloitte. (2021). IMPACT OF AIR POLLUTION ON CAREER DECISIONS OF THE HIGHLY-SKILLED WORKFORCE IN SOFIA Commissioned.

European Environment Agency (2025). Harm to human health from air pollution in Europe: burden of disease status, 2025, https://www.eea.europa.eu/en/analysis/publications/harm-to-human-health-from-air-pollution-burden-of-disease-status-2025

Friedenreich, C. M., Ryder-Burbidge, C., & McNeil, J. (2021). Physical activity, obesity and sedentary behavior in cancer etiology: epidemiologic evidence and biologic mechanisms. Molecular Oncology, 15(3), 790–800. https://doi.org/10.1002/1878-0261.12772

Gill, G., Bigazzi, A., & Winters, M. (2022). Investigating relationships among perceptions of yielding, safety, and comfort for pedestrians in unsignalized crosswalks. Transportation Research Part F: Traffic Psychology and Behaviour, 85, 179–194. https://doi.org/10.1016/J.TRF.2022.01.007

Huang, X., Liang, H., Zeng, L., & White, M. (2025). Evaluating urban walkability: a comprehensive review of tools and techniques. Architectural Science Review, 68(4), 263–277. https://doi.org/10.1080/00038628.2024.2410731

International Energy Agency. (n.d.). Bulgaria - Countries & Regions - IEA. Retrieved 18 September 2025, from https://www.iea.org/countries/bulgaria/emissions

Koo, B. W., Guhathakurta, S., & Botchwey, N. (2022). How are Neighborhood and Street-Level Walkability Factors Associated with Walking Behaviors? A Big Data Approach Using Street View Images. Environment and Behavior, 54(1), 211–241. https://doi.org/10.1177/00139165211014609

Lee, S., & Talen, E. (2014). Measuring Walkability: A Note on Auditing Methods. Journal of Urban Design, 19(3), 368–388. https://doi.org/10.1080/13574809.2014.890040

Manisalidis, I., Stavropoulou, E., Stavropoulos, A., & Bezirtzoglou, E. (2020). Environmental and Health Impacts of Air Pollution: A Review. Frontiers in Public Health, 8, 505570. https://doi.org/10.3389/FPUBH.2020.00014/BIBTEX

Meeder, M., Aebi, T., & Weidmann, U. (2017). The influence of slope on walking activity and the pedestrian modal share. Transportation Research Procedia, 27, 141–147. https://doi.org/10.1016/J.TRPRO.2017.12.095

Patel, N., Nguyen, H.-H., van de Geest, J., Wagtendonk, A., Raju, M. J. S., Dadvand, P., de Hoogh, K., Cirach, M., Nieuwenhuijen, M., Minh Lam, T., & Lakerveld, J. (2025). A Walk across Europe: Development of a high-resolution walkability index. Health & Place, 96, 103544. https://doi.org/10.1016/J.HEALTHPLACE.2025.103544

Petrova-Antonova, D., Murgante, B., Malinov, S., Nikolova, S., & Ilieva, S. (2025). Walkability analysis of Sofia’s neighborhoods powered by 15-minute city concept. Cities, 165, 106171. https://doi.org/10.1016/j.cities.2025.106171

Piras, G., Pini, F., & Di Girolamo, P. (2024). PM10 emissions from tires: A disruptive estimate questioning present pollution mitigation strategies. Atmospheric Pollution Research, 15(1), 101939. https://doi.org/10.1016/J.APR.2023.101939

Pucher, J., Buehler, R., Bassett, D. R., & Dannenberg, A. L. (2010). Walking and cycling to health: a comparative analysis of city, state, and international data. American Journal of Public Health, 100(10), 1986–1992. https://doi.org/10.2105/AJPH.2009.189324

Quercia, D., Schifanella, R., & Aiello, L. M. (2014). The shortest path to happiness: Recommending beautiful, quiet, and happy routes in the city. Proc. 25th ACM conference on Hypertext and social media, Santiago, Chile, 1-4 September, pp. 116–125. https://doi.org/10.1145/2631775.2631799

Sallis, J. F., Cain, K. L., Conway, T. L., Gavand, K. A., Millstein, R. A., Geremia, C. M., Frank, L. D., Saelens, B. E., Glanz, K., & King, A. C. (2019). Is Your Neighborhood Designed to Support Physical Activity? A Brief Streetscape Audit Tool. Preventing Chronic Disease, 12(9). https://doi.org/10.5888/PCD12.150098

Siriaraya, P., Wang, Y., Zhang, Y., Wakamiya, S., Jeszenszky, P., & Kawai, Y. (2020). Beyond the Shortest Route: A Survey on Quality-Aware Route Navigation for Pedestrians. IEEE Access, 8, 135569-135590. https://doi.org/10.1109/ACCESS.2020.3011924

Sundling C, Jakobsson M. (2023) How Do Urban Walking Environments Impact Pedestrians’ Experience and Psychological Health? A Systematic Review. Sustainability, 15(14):10817. https://doi.org/10.3390/su151410817

World Health Organisation. (2023). Air pollution mortality rate per 100 000 population - European Health Information Gateway. https://gateway.euro.who.int/en/indicators/epw_2-air-pollution-mortality-rate-per-100-000-population/#id=36843

Your Europe. (2024). Road rules and safety|Bulgaria. https://europa.eu/youreurope/citizens/travel/driving-abroad/road-rules-and-safety/bulgaria/index_en.htm#speed_limits