

Automated Identification of Existing and Potential Urban Central Places Based on Open Data and Public Interest

Anna Pavlova(✉), Aleksandr Katynsus, Maksim Natykin, and Sergey Mityagin

ITMO University, Birzhevaya line, 14, Saint Petersburg, Russia
370012@niuitmo.ru

Abstract. Urban central places are scarcely identified and utilized in modern research. There are many interpretations of this concept and not so many methods for identifying, categorising and classifying such places. Previous works and studies are based on Christaller's Central Place Theory (CPT), which was introduced many years ago. All works are based on the concepts of CPT and they only extend or clarify these concepts, some works add new criteria and indexes for more accurate identification. This study presents a novel approach for automated identification of existing and potential urban central places based on open data and public interest, offering a data-driven and comprehensive perspective on centrality patterns that extends and complements the classical Central Place Theory (CPT) formulated by Christaller. In this paper, we describe the main types of central places to work with, and using data ingestion tasks, we analyse, scrape and extract centres from open data. We enrich service data from other sources such as social media networks and cartographic services. Points of interest, potential and existing services are defined and extracted using modern technologies such as Docker, Selenium, Airflow, Jupyter and Postgres DB. Our findings explain what central places are and how they can be found in modern cities. We also discovered potential and quasi-centres based on public interest and transport accessibility. The identification of central places is done through six different profiles - gastronomic, cultural, retail, multi-profile, potential central places and public interest hotspots - allowing a comprehensive understanding of centrality patterns driven by different urban functions and amenities that cater to the diverse needs and interests of residents and visitors. These results can be used in future exploratory city analyses.

Keywords: Open Data · Transport Accessibility · Centrality of urban blocks · Urban environment modeling · NLP · Data Engineering · Points of Interest · Central Places

1 Introduction

Every city is a complex mechanism consisting of a variety of interconnected parts. One of the most important parts of a city is its central places, which result from

© The Author(s), under exclusive license to Springer Nature Switzerland AG 2024
O. Gervasi et al. (Eds.): ICCSA 2024, LNCS 14813, pp. 408–421, 2024.
https://doi.org/10.1007/978-3-031-64605-8_28

variations in natural resources, spatial inequality, different transport accessibility, collocation of urban services and citizen activities [1]. Neglecting such an important part of the city can lead to spontaneous migration, reduced quality of public places, limited access to services and a blurred urban environment [2].

This paper focuses on developing a new approach to identifying central places of cities, which takes into account modern technologies, open data, public interest and best practices in urban studies. Central places are underdeveloped and there are many interpretations and meanings of this concept, therefore there is a lack of methodological approaches that use all the data generated by people every day. A central place is understood as a territorial unit that attracts people who want to satisfy different needs, for example gastronomic, cultural or shopping. It is characterised by a wide range of urban services to satisfy needs, recognition and well-developed transport accessibility. Based on CPT [3] we decided to define the main types of central places to existing and potential. Existing centres can be divided into gastronomic, cultural, market or multi-profile, i.e. having several equally expressed profiles. Potential: based on accessibility and points of interest (POIs) and quasi based on POIs only. The developed methodology will provide opportunities for future insights, increased accuracy of urban research and effective urban planning for sustainable development.

The article is divided into the following parts: related works with the reviews of existing approaches, method description with a brief summary of each stage, results, experimental studies with results, and finally a discussion part with the conclusion.

2 Related Works

All methods of identification of urban central places refer to the main theory - Central Place Theory by Christaller [4] and similar one by Lösch. Christaller introduced the main concept of a central place and types of such places. Moreover, he defined some laws according to which urban central places are formed, identified a range of their functions, their hierarchy and location patterns. The main difference between Christaller's and Lösch's theories was that Christaller defined the hierarchy and then applied functions to each level, whereas Lösch started with the basic single economic activity. The production of goods, services and markets were the logical sequence in Lösch's theory. Numerous studies have been based on these theories, applied to many cities and generalised, extended and modernised. The main drawback of the two methods described above is the amount of data analysed, due to the time in which they were defined. Big data and modern technologies can improve and develop these methods, making them more accurate for use in the modern world.

In the research 'Investigating the central place theory using trajectory big data' [5] by Pengjun Zhao, Haoyu Hu and Zhao Yu, a new method of central place identification has been proposed. Their study aims to apply CPT combined with mobile phone trajectory big data in rural China. They investigated the benefits

and insights of this hypothesis and created a new methodology. Using trajectory big data, the researchers revealed the dynamic connection between settlements, which reflects the main functions of services described by Christaller. To describe and identify city centres, the authors used spatial autocorrelation and geographically weighted regression. A major problem is that trajectory data is not open data and this method cannot be automated and reproduced. Furthermore, mobile phone data cannot describe emotions and intentions.

The next article 'Rethinking the Identification of Urban Centres from the Perspective of Function Distribution: A Framework Based on Point-of-Interest Data' [6] is based on the extraction of points of interest with their future classification and clustering. In this research, the authors proposed to identify urban centres using only POIs. They defined a model with three main steps: data classification, data computation and data visualisation. In the first step, all points ingested from social media networks and electronic maps were classified according to their contribution to urban functions. In the second step, the researchers used the DBSCAN algorithm and the CVNN method for clustering and validation. Finally, the last step was to visualise the distribution of these points and to reveal the implications. Based on the number of points, their quality and distribution, the students identified urban centres. Like the previous methods, this one was not automated and it had no characteristics of the points and centres included.

Another relevant work related to this study is 'Identifying city centre using human travel flows generated from location-based social networking data' [7] by Yeran Sun, Hongchao Fan, Ming Li and Alexander Zipf. The researchers proposed to use location-based social networking (LBSN) data to extract the main centres of cities. LBSN is represented as location-tagged data in social media networks and services such as Foursquare. This data was extracted and analysed, and then the main clusters were detected using the local Getis-Ord statistical method, which includes hotspots of travel flow detection and the clustering itself. The next steps were to apply DBSCAN and the Grivan-Newman (GN) algorithm to the extracted data. The last step was to determine the city centres by the rate defined in the work and to delineate them. The final result was areas represented by Voronoi polygons with precise boundaries. This work does not identify emotions and characteristics, so there are only central places and we can't really say why it is a centre and what emotions people feel there.

One of the approaches to identifying central places is described in the paper 'Urban Centrality: A Simple Index' [8]. The authors define a new centrality index that can be applied to urban areas. This index is based on variables such as employment and population density or trip generation rates. This new index was tested on the following metropolitan areas: Pittsburgh, Los Angeles, São Paulo, Brazil and Paris. The results were compared with other indices such as global and local Moran's I and density gradient estimates. This research has a major drawback - the data used is not open data, so the approach cannot be reproduced and automated.

3 Method

To test the developed method it was decided to choose the city of Saint Petersburg in Russia with a population of about 4.5 million people. The city was divided into districts and for each district all the central urban places were extracted. The Moscow district of St Petersburg was used for demonstration purposes. The Moscow district of St Petersburg was chosen for a number of reasons. Firstly, it is a well-established and densely populated area within the city. It represents a typical urban environment with a wide range of services and amenities. Secondly, it is known for its significant cultural and historical landmarks as well as a vibrant gastronomic and commercial scene. Thus, the study of this district allows for a comprehensive analysis of central places with different profiles. In addition, this district is industrially developed and accounts for 40% of the production of industrial products. By focusing on an area within a large city such as Saint Petersburg, the results and findings of the study can have direct relevance and applicability to real urban contexts.

It is implemented in a dedicated environment with all the necessary Python libraries and services such as Selenium, Airflow, Postgres DB and Jupyter Server. The business process modelling diagram of the method is shown in Fig. 1.

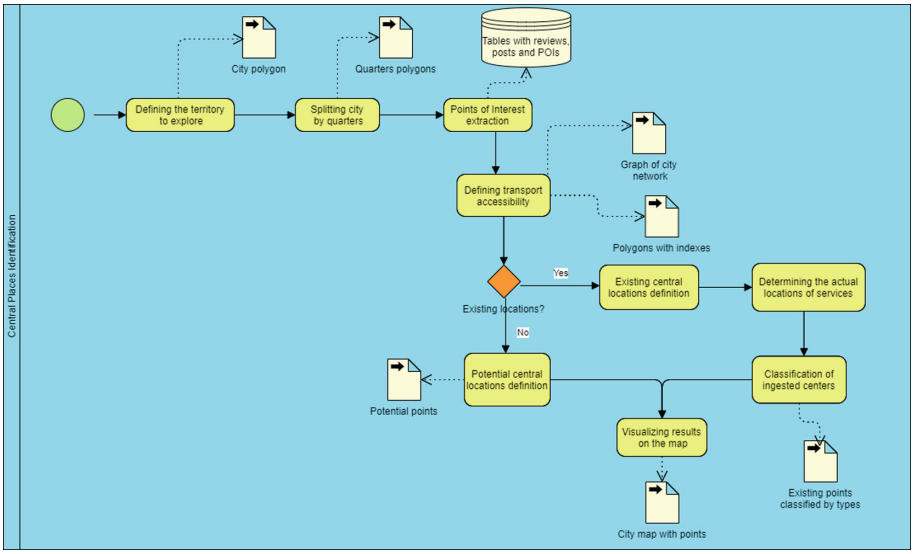


Fig. 1. Method Diagram

The new method developed in this article consists of six main stages:

1. Identification of the boundaries of the study area and dividing it into blocks;
2. Points of Interest extraction. On this stage we used data engineering techniques with natural language processing methods and models;

3. Transport accessibility definition. This stage uses a city network graph and quarters splitting to apply to every quarter an index;
4. Identification of existing urban central places via services and points of interest;
5. Identification of potential urban central places via transport accessibility and public interest;
6. Result visualisation on the map

3.1 Identification of Boundaries and Blocks of the Study Area

At the beginning of the method, the study area should be defined. Once the area has been selected, its boundaries are extracted from open sources such as OpenStreetMap. As the method takes transport accessibility into account, it is recommended to select an area of no more than 150 square kilometres. For larger cities, such as St. Petersburg or Moscow, the study is carried out within a single district.

Once the boundaries of the area have been determined, it is necessary to divide it into the smallest elements of the city - blocks. The division into blocks helps to evaluate each section of the territory in terms of points of public interest, transport accessibility and concentration of city services. In order to identify neighbourhoods, the researcher must divide the territory by roads, reservoirs and railway lines.

The result of this stage is a document containing information on the location of each block in the study area.

3.2 Points of Interest Extraction

The second step of the proposed method is to identify POIs in a city. POIs are objects on the map that contain information about important public places connected to the social infrastructure. They have various characteristics such as coordinates, names, degrees of public interest and purposes. The concept of POIs is accompanied by another one - the infrastructure of society, which can be understood as industries, sub-sectors indirectly related to the production process in the sense of ensuring the conditions of work and social activity of the population, activities in the field of culture, everyday life, social and interpersonal communication.

It is proposed to add the public interest for the identification of existing and potential centres due to the amount of data generated by people. This phase is based on open data, social media parsing, data engineering and natural language processing (NLP) [9]. The idea of POI extraction is to ingest data from cartographic services, social media networks and geoinformation services, enrich this data with context and process it using NLP methods and models to extract named entities associated with POIs. This type of data is open and easily ingested. It also represents many people who use the Internet and have an active civic position. There are many neighbourhood groups where people post problems and ideas on how to develop areas. We can easily find and use this

data. The only limitation is that not all categories of citizens are represented by this data. This limitation cannot be solved by the idea of open data. We can't use data from surveys or questionnaires. More and more people use social media and leave reviews on maps, so very soon this will not be a problem.

The whole method works in a predefined environment with Airflow, PostGis, Selenium and Jupyter Server deployed in a docker container. This stage of the method is split into two parts: Data Ingestion and Data Processing.

The first part is an Airflow DAG (directed acyclic graph) with the following tasks:

1. Getting the polygon of the processed territory. It can be a city, a district or even a street;
2. Social media network parsing (VK). Receiving all the posts, photos and group posts where the territory was mentioned [10];
3. Buildings ingestion from the open data;
4. Reverse geocoding. On this stage all the buildings' polygons are converted to points, then the reverse geocoding is applied to every building without address;
5. Scrapping reviews from the cartographic service by the addresses.

The method has a stage where all the buildings of the city are ingested from open street map (OSM). It is a well-known fact that this service can have not accurate and not actual data. This problem is solved via using maps with reviews. All the buildins on the map are processed and connected to the buildings from OSM. While scrapping the data from maps all the posts are preprocessed to remove graphic information for the better results on the next stage.

All the scrapped and parsed data is saved into two Postgres tables: Reviews and Posts. Reviews table contain the name of the service in the building and the date when the reviews was made. Posts table contains group name, post itself, and date in UNIX timestamp. To ingest all the data it was decided to use Python programming language and special libraries such as vk (to parse data from VK social media network), selenium and web driver to scrape data from cartographic services, ArcGis to get existing points of interest, airflow to build DAG, and sqlalchemy to work with database.

When all the data is stored in tables, comes the second part - data processing. It is the second DAG with three main tasks:

1. NLP models initialization. Here main models for sentiment analysis, named entity recognition, and keywords extraction are defined. Stanza, RuBert, and KeyBert were chosen for this task. Stanza is used to extract named entities, RuBert is for emotions predictions, KeyBert is for keywords extraction;
2. Posts and reviews ingested before are processed with the three models to extract locations and organisations, emotions and keywords;
3. All the extracted locations are geocoded to receive points on the map;
4. Data analysis.

There are some limitations in the use of NLP models for sentiment analysis. Sometimes emotions are not well defined in texts, and models can't identify if the post contains positive or negative emotion. This can be fixed by transfer learning on the ingested data all the time we get new data.

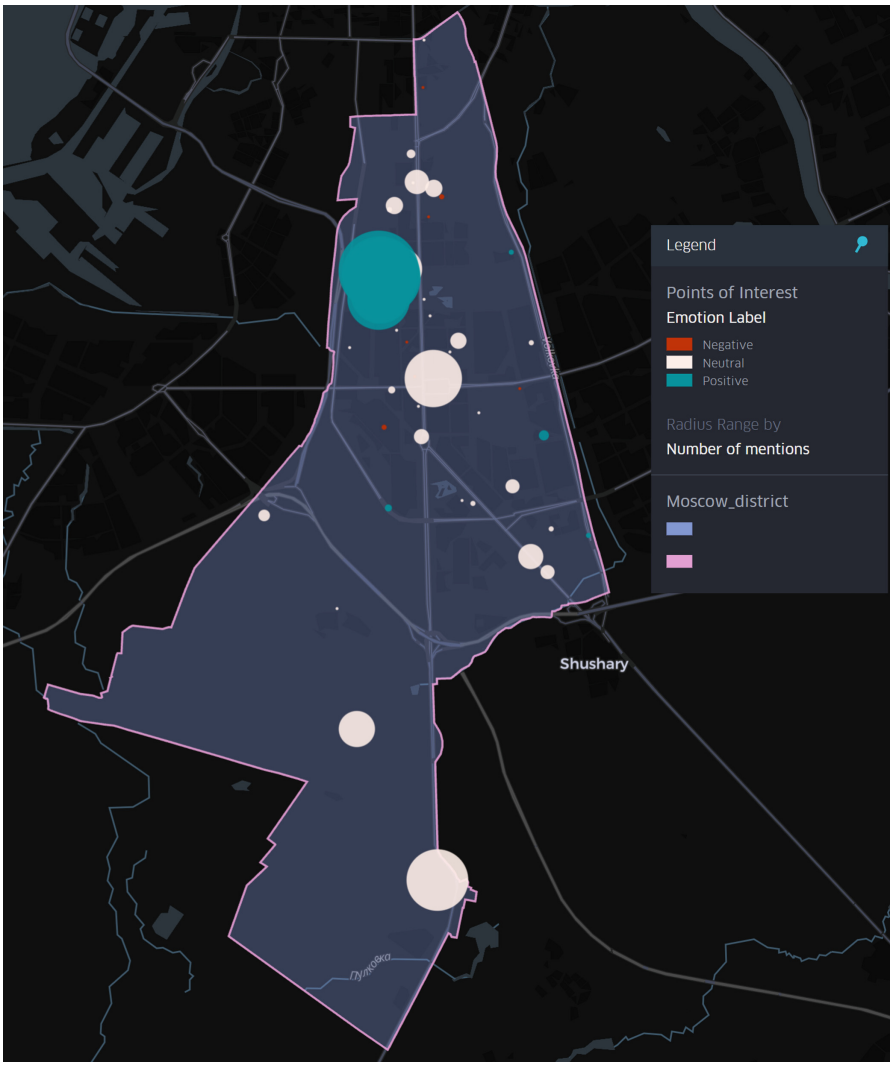


Fig. 2. Points of Interest

The extracted data is saved in a new table with the name Processed posts. Then with the help of aggregate functions all the points are split in popular and not popular. The emotion degree of a point is an average of all the emotions (0 - Neutral, 1 - Positive, 2 - Negative). The second stage result of the defined method is represented as a map with the polygon of the explored territory and points of interest. To show the results it was decided to use Moscow district, Saint Petersburg, Russia, Fig. 2.

In the picture shown below a map of a district can be seen. Extracted points are colored according to their labels: Neutral - white, Negative - red, Positive - green. Radius of each point is a number of mentions of the point in social media network and map services.

As a result of this stage, each block is assigned a mention score (1 - frequently mentioned place, 2 - averagely mentioned place, 3 - rarely mentioned place), which means the concentration of mentions for the block.

3.3 Transport Accessibility Definition

In the third step of the method, the transport accessibility of each identified block is assessed. This is done by loading a road network from open sources. It contains information on pedestrian routes, public transport routes and cycle routes. In this study, we deliberately excluded the road network for cars, as the travel time by car depends on various situational factors that are difficult to access in open sources. Such factors may include traffic congestion during rush hours, traffic restrictions during holidays, and others.

Transport accessibility is calculated for each block based on the time it takes to travel from one block to all others. Finally, a block with the best transport accessibility and one with the worst accessibility are identified.

The result of this stage is a transport accessibility score for each block (1 - good transport accessibility, 0 - poor transport accessibility), which corresponds to the travel time required to get from each block to all others by walking, cycling or public transport.

3.4 Identification of Existing Urban Central Places

At this stage, the location of urban services is downloaded from open sources. Urban service refers to a service provided within the urban environment, aimed at meeting the needs of residents and visitors of the city. Data is uploaded across three profiles, corresponding to the highlighted profiles of existing central places:

- The gastronomic profile includes cafes, restaurants, pubs, canteens, snack bars and other catering establishments;
- The cultural profile consists of museums, exhibitions, art galleries, attractions and libraries;
- The trade profile includes shopping centers.

Table 1. Tags of city services by profiles of existing central places

Profile	Service Tags
Cultural	amenity = theatre, tourism = museum, tourism = gallery, amenity = library, amenity = arts_centre, tourism = attraction
Gastronomic	amenity = pub, amenity = bar, amenity = hookah_lounge, shop = bakery, amenity = cafe, amenity = restaurant, amenity = fast_food
Retail	amenity = marketplace, shop = supermarket, shop = mall

To retrieve the actual locations of urban services by profiles, the tags presented in Table 1 are used. The output consists of three files with point geometries.

After all services for three profiles have been unloaded, the concentration of its services for each quarter is determined separately for each profile.

That is, the result of this process will be three assessments for each quarter (1 - the most concentrated, 2 - moderately concentrated and 3 - not concentrated) for gastronomic, cultural and trade profiles.

This method also identifies other types of central places:

- The dominant central place profile is one in which a particular neighborhood scores higher than others. If the scores of several profiles are equal, the place is considered to be multi-profile.
- Further, places that have high public interest and good transport accessibility, but absolutely no developed services of any of the profiles, are considered potential.

Then the researcher needs to collect the data into a single table, which will contain the following data:

- Polygons. Information reflecting the border of the quarter in coordinate format;
- Mention score. Information reflecting the number of points of public interest;
- Transport accessibility. Information reflecting the extent to which the block is connected to other blocks by pedestrian and transport paths;
- Gastronomic concentration. Information reflecting the number of services related to the gastronomic profile;
- Cultural concentration. Information reflecting the number of services related to the cultural profile;
- Trade concentration. Information reflecting the number of services related to the trading profile;
- Centrality Score. A summary assessment of all indicators, allowing to assess existing central places and the influence of the central place on the studied area;
- Central place profile. Highlighting the dominant profile of an existing central location.

The degree of centrality (1 - the highest degree of centrality, 2 - a high degree of centrality, 3 - an average degree of centrality, 4 - a weak degree of centrality, 5 - a very weak degree of centrality) is determined by estimates of mentions, good transport accessibility and concentration of services. Neighborhoods with poor transport accessibility are not assessed because they do not belong to existing central locations.

- If both the mention score and the score of one of the profiles are equal to 1, then the researcher is facing a central place with the highest degree of centrality.
- If one of the scores is 1 and the other is 2, then the researcher is facing a central place with a high degree of centrality.
- If both the mention score and the score of one of the profiles are equal to 2, then the researcher is facing a central place with a medium degree of centrality.
- If one of the scores is 2 and the other is 3, the researcher is facing a central place with a low degree of centrality.
- If both values are equal to 3, then the researcher is facing a central place with a very weak degree of centrality, which does not stand out as an existing central place.
- Furthermore, places that have one score of 1 and another score of 3 are not considered central. This is because it is assumed that either the place is crowded with services that are not of interest to the population, or that the population is interested in a central place that has not yet been established in terms of services.

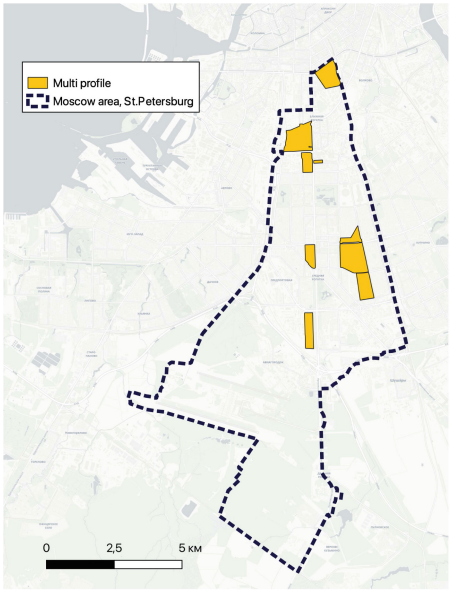
Based only on the assessment of the number of points of public interest per block, quasi-places are also identified, representing information about the popularity of the place among the population.

4 Results

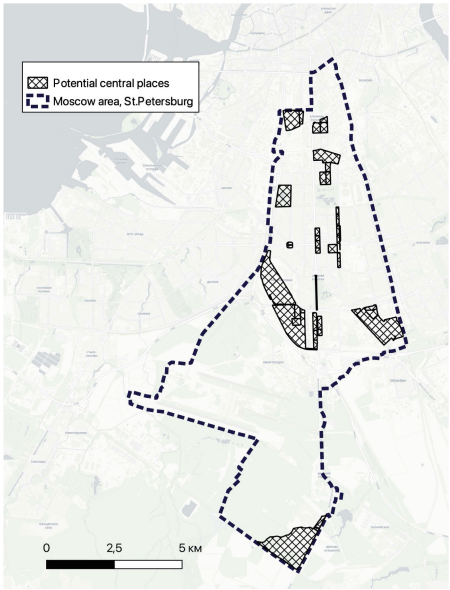
As it was described in the previous sections, this method is focused on identifying urban central places by 5 categories. Each category is dependent of the city's infrastructure and should be extracted separately to receive more accurate results. Hence, the following central places in different profiles were identified:

1. Gastronomic centers shown on the Fig. 3(a).
2. Cultural centers shown on the Fig. 3(b).
3. Commercial center shown on the Fig. 3(c).
4. Multi-profile centers what can consist of the previously described types shown on the Fig. 4(a).
5. Potential central places presented in Fig. 4(b).

Figures 3 and 4 discover a real picture of services in the area. As it can be seen, almost all the services are concentrated on the north of the selected district. Such results can be explained with the closeness to the city's center, where all the goods are mostly placed.

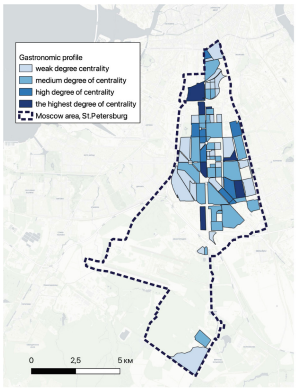


(a) Multi-profile central places

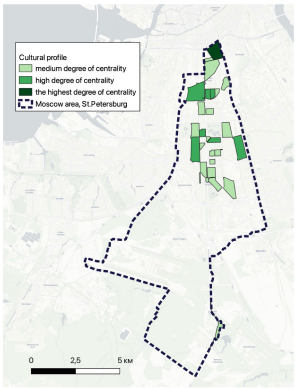


(b) Potential central places

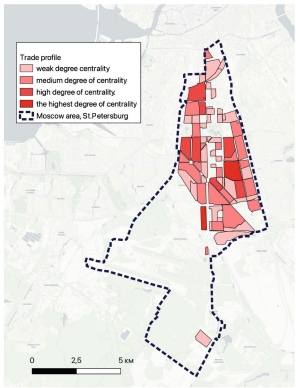
Fig. 3. Other types of central places



(a) Gastronomic profile



(b) Cultural profile



(c) Trade profile

Fig. 4. Central places of different profiles

The summary map, including existing central places of all degree of centrality and all profiles, as well as potential and quasi places are shown on the Fig. 5. These visual representations provide a comprehensive view of the central places in different profiles and highlight their respective degrees of centrality. The results contribute to a better understanding of the urban dynamics and support decision-making processes for urban planning and development in the Moscow District of Saint Petersburg.

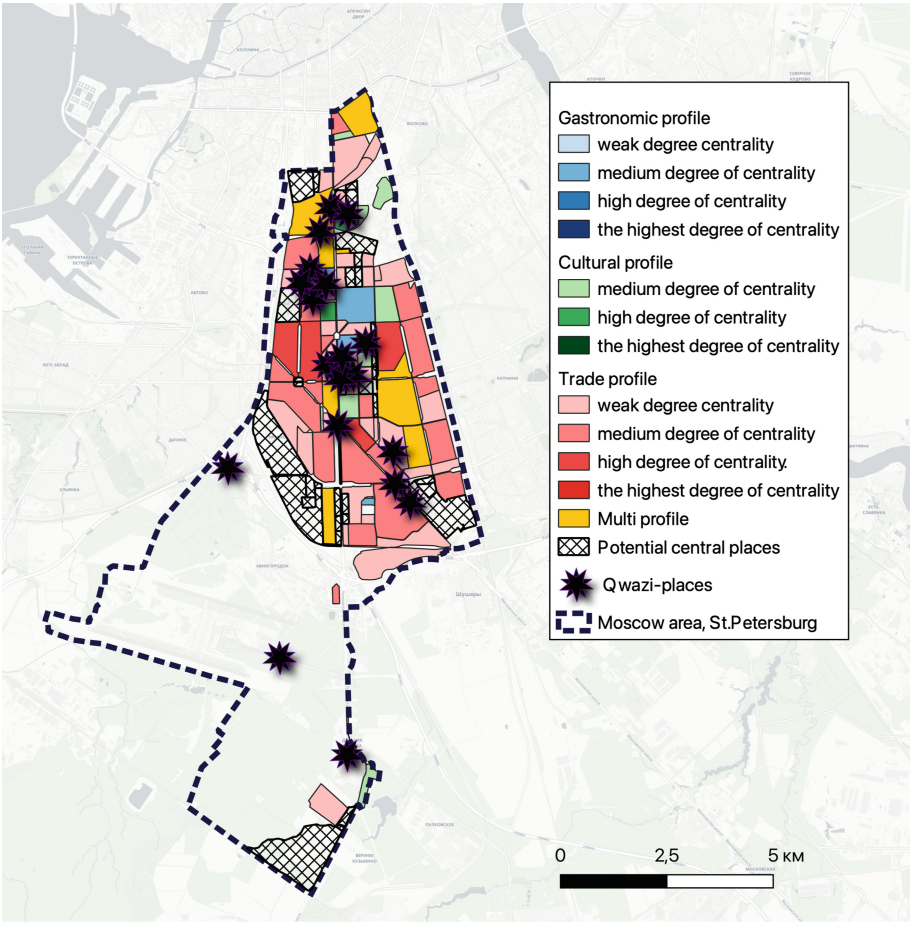


Fig. 5. Summary map

5 Discussion

In this study, an experimental investigation of the method for automatic identification of central places in a city was carried out, with the ability to identify

existing central places and to determine places with the potential to become central.

Based on the results of this research, it can be concluded that this method is a valuable tool for both companies and city administrations in the planning and optimisation of urban areas. The data obtained through this method can be used to make decisions regarding the development of urban infrastructure, the identification of centres of attraction for residents and visitors, and the improvement of the urban environment and quality of life.

The identification of new central places helps to reduce the concentration of population and traffic in existing central zones, contributing to a more balanced development of urban spaces and an increased level of comfort for city dwellers. In addition, the identification of new central places can act as a stimulus for attracting investment and promoting tourism, which in turn can contribute to the economic development of the city and increase employment opportunities.

6 Conclusion

In conclusion, this study has demonstrated the applicability and effectiveness of the method for automatic identification of central places, confirming its significance for practical use in various areas of urban planning and development. The proposed method for automatic identification of central places based on open data and public interest is not limited to the case study of St. Petersburg, but can be easily extended and applied to cities in different countries and continents, providing a versatile and transferable approach to unravel centrality patterns in different urban contexts. The method provides valuable insights for stakeholders involved in urban development, enabling them to make informed decisions and optimise the growth and liveability of cities. Further research and refinement of the method can further enhance its utility and contribute to the sustainable development of urban environments.

References

1. Wong, Z., Li, R., Zhang, Y., Kong, Q., Cai, M.: Financial services, spatial agglomeration, and the quality of urban economic growth-based on an empirical analysis of 268 cities in China. Financ. Res. Lett. **43**, 101993 (2021)
2. Khan, M.F., Aftab, S., et al.: Quality of urban environment: a critical review of approaches and methodologies. Curr. Urban Stud. **3**(04), 368 (2015)
3. Christaller, W., Baskin, C.W.: Central places in Southern Germany. Prentice Hall, Englewood Cliffs, NJ, English translation of the German original: Christaller, Walter (1933). Die Zentralen Orte in Suddeutscland, Fischer, Jena (1966)
4. King, L.J.: Central place theory (2020)
5. Zhao, P., Hu, H., Yu, Z.: Investigating the central place theory using trajectory big data. Fundam. Res. (2023)
6. Yu, L., Yu, T., Wu, Y., Wu, G.: Rethinking the identification of urban centers from the perspective of function distribution: a framework based on point-of-interest data. Sustainability **12**(4), 1543 (2020)

7. Sun, Y., Fan, H., Li, M., Zipf, A.: Identifying the city center using human travel flows generated from location-based social networking data. Environ. Plan. **43**(3), 480–498 (2016)
8. Pereira, R.H.M., Nadalin, V., Monasterio, L., Albuquerque, P.H.: Urban centrality: a simple index. Geogr. Anal. **45**(1), 77–89 (2013)
9. Cai, M.: Natural language processing for urban research: a systematic review. Heliyon **7**(3) (2021)
10. Chen, N.C., Zhang, Y., Stephens, M., Nagakura, T., Larson, K.: Urban data mining with natural language processing: social media as complementary tool for urban decision making. In: The 17th International Conference on Computer Aided Architectural Design Futures, pp. 101–109 (2017)