Journal of Internet Services and Applications, 2025, 16:1, doi: 10.5753/jisa.2025.5152

Modeling Interest Networks in Urban Areas: A Comparative Study of Google Places and Foursquare Across Countries

Gustavo H. Santos [**Univ. Tecnológica Federal do Paraná** | ***gustavohenriquesantos@alunos.utfpr.edu.br***]
Fernanda R. Gubert [**Univ. Tecnológica Federal do Paraná** | ***fernandagubert@alunos.utfpr.edu.br***]
Myriam Delgado [**Universidade Tecnológica Federal do Paraná** | ***myriamdelg@utfpr.edu.br***]
Thiago H. Silva [**Universidade Tecnológica Federal do Paraná** | ***thiagoh@utfpr.edu.br***]

DAINF, Universidade Tecnológica Federal do Paraná (UTFPR), Av. Sete de Setembro, 3165, Rebouças, Curitiba, PR, 80230-901, Brazil

Received: 03 November 2024 • **Accepted:** 25 February 2025 • **Published:** 17 March 2025

Abstract Location-Based Social Networks (LBSNs) are valuable for understanding urban behavior and providing useful data on user preferences. Modeling their data into graphs like interest networks (iNETs) offers important insights for urban area recommendations, mobility forecasting, and public policy development. This study uses check-ins and venue reviews to compare the iNETs resulting from two distinct LBSNs, Foursquare and Google Places. Although these two LBSNs differ in nature, with data varying in regularity and purpose, their resulting iNETs reveal similar urban behavior patterns. When analyzing the impact of socioeconomic, political, and geographic factors on iNET edges — each edge representing users' interests in a pair of regions — only geographic factors showed a significant influence. When studying the granularity of area sizes to model iNETs, we highlight important trade-offs between larger and smaller sizes. Additionally, we propose a methodology to identify clusters of geographically neighboring areas where user interest is strongest, which can be advantageous for understanding urban space usage.

Keywords: Location-Based Social Networks, Google Places, Foursquare, User Interest, Urban Areas

1 Introduction

Location-Based Social Networks (LBSNs) help us understand several issues in the context of urban computing [Ferreira *et al.*, 2015; Santala *et al.*, 2017; Silva *et al.*, 2017a; Ladeira *et al.*, 2019; Veiga *et al.*, 2019; Ferreira *et al.*, 2020; Senefonte *et al.*, 2022; Silver and Silva, 2023; Silva and Silver, 2024]. In particular, LBSNs offer urban data that inherently reflect social aspects, such as user preferences [Silva *et al.*, 2019].

Geolocated user activities on LBSNs provide useful urban data, such as reviews and check-ins at venues throughout the city being analyzed. These activities can be aggregated into an undirected graph in which nodes represent areas (e.g., neighborhoods) in the city where the data have been shared, and each edge connects a pair of areas visited by the same user. This type of modeling leads to what we refer to in this work as Interest Networks (iNETs).

Modeling data through iNETs could provide valuable insights to improve the understanding of user behavior in urban environments. For instance, it could provide a deeper understanding of users' interest in specific physical spaces within the city, facilitate geographic area recommendations, enhance mobility forecasting, and support the development of public policies to increase interest in certain urban areas.

In this context, the main objective of this work is to compare iNETs modeled using data from two LBSNs: Foursquare and Google Places. From Foursquare, we utilize check-in data, representing users sharing their location at a venue with friends. In Google Places, we use data from user reviews of venues. Both platforms capture the type of venue the user is at, for example, a restaurant or a bookstore. However, the data they provide are different: check-ins are more real-time, reflecting where the user is at a specific moment, whereas reviews may be posted after the user has left the venue. With this, we aim to investigate whether different LBSNs capture similar information when modeling data through iNETs, considering various variables and scenarios.

In our previous work [Santos *et al.*, 2024], we compare the two LBSNs in the context of Curitiba, Brazil, to determine whether they model urban behavior similarly in capturing users' interest in neighborhoods. The present work significantly expands the previous similarity analysis between LBSNs by considering areas in different countries. Additionally, it assesses the impact on results and practical applications when varying the granularity of urban areas. Besides, it shows how the iNETs can be used to better understand urban phenomena, examining the influence of socioeconomic, political, and geographic factors on the interest for the neighborhoods in Curitiba and aggregates analyses regarding Urban Preference Zones (UPZones), and their corresponding networks (UPZ-iNETs) in London. These case studies also show how iNETs can be used for urban analysis across countries.

The main contributions of this work can be summarized as follows.

- We propose a new approach, Interest Networks (iNETs), to studying urban phenomena by analyzing user activities on Location-Based Social Networks (LBSNs). iNETs offer innovative ways to explore how people engage with different areas of a city.

- An impact analysis of urban area granularity on iNETs: In the present work, we introduce the tool h3-cities[1], which enables the division of any city, with available OpenStreetMap data, into hexagons of varying sizes. Our findings show that when modeling iNETs at larger granularities, the iNETs derived from the two LBSNs are more similar, whereas smaller granularities reveal greater differences. However, we should not only aim for large granularity because we could miss important nuances regarding user behavior – we present trade-offs to guide further studies.
- An influence analysis of socioeconomic, political, and geographic factors on iNETs: We observe that average monthly income, racial composition, and political polarization do not significantly impact people's interests in urban areas. However, geographic distance moderately correlates with urban behavior, as people often visit nearby areas.
- A methodology for identifying urban preference zones (UPZones) and their resulting networks: This methodology explores iNETs to identify clusters of geographically neighboring areas where user interest is strongest. The results are consistent across LBSNs, even at smaller granularities. This new methodology can be valuable for providing UPZone networks, referred to UPZ-iNETs, and for future analyses in understanding urban user interests.

The remainder of the article is organized as follows. Section 2 reviews related work, while Section 3 describes the used datasets and their characteristics. Section 4 details the methods applied and developed for analyzing iNETs. Section 5 presents the results. Finally, Section 6 concludes the study and presents directions for future work.

2 Related Works

This section reviews relevant literature across four key topics: comparison of LBSN data, people's movement patterns, understanding an urban context through different granularities, and the identification of similar urban zones.

2.1 LBSN Data Comparison

This section presents some works that use and compare data from LBSNs in their research. For example, Silva *et al.* [2013] investigated the possibility of using two different LBSNs, Instagram and Foursquare, to collect data for which location was shared. The study sought to understand whether the information obtained could be complementary and/or similar for both bases regarding city dynamics and urban behavior patterns. From this, they concluded that the two datasets provided compatible and complementary information, in which, for example, a check-in from Foursquare could bring information about the category of a venue commented on in an Instagram post and also capture urban aspects in a similar way, such as the most popular areas of cities.

Martí *et al.* [2019] explore the potential of using data from LBSNs such as Foursquare, Twitter, Google Places, Instagram, and Airbnb for research into urban phenomena, recognizing not only the benefits but also the challenges that the use of these data sources entails. The study presents research that uses data from LBSNs to analyze the city's dynamics and proposes a methodology for data retrieval, selection, classification, and analysis. It also identifies the main thematic lines of investigation based on the data provided by these platforms to offer a framework for studying urban phenomena through LBSN data.

Nolasco-Cirugeda and García-Mayor [2022] show how data from Foursquare, Twitter, and Google Places are crucial for analyzing the use of urban space and social dynamics. By detailing pioneering research and case studies over the last decade, the article shows how LBSN data offer important information about urban life, helping to understand social dynamics and specific urban interventions. The research employs multiple analytical approaches at scale to address diverse urban issues, from social neighborhood dynamism to tourism and green infrastructure preferences, highlighting the comprehensive analytical potential of LBSN data.

Skora *et al.* [2022] examined whether information extracted from Foursquare data could resemble information released by the WTO (World Tourism Organization). The study found the potential of using LBSNs to facilitate the understanding of tourist movements on larger scales and in more detail than traditional sources, despite limitations associated with LBSNs, such as the predominant use of young people with internet access.

Our work does not seek to investigate the complementary character of LBSNs, which could be beneficial for, for example, data integration [Silva and Fox, 2024] nor compare it with official sources; instead, it aims to seek to what extent distinct LBSNs will model urban behavior in a similar way, understanding when to expect different results when using these tools.

2.2 Understanding Urban Displacement

Other studies use large-scale data, including LBSN data, and data mining techniques to understand which factors may be associated with people's movement patterns. For example, Cheng *et al.* [2021] used geolocated data from Twitter to understand user movements. The authors associated this spatial information with the economic characteristics of users, the geographic aspects of the areas frequented, as well as their positioning within the social network and the language used in their *check-ins*. Thus, they identified the Lévy Flight model in mobility patterns, in which short distances are traveled more frequently, and longer distances occur more rarely. They also presented the influence of population density and popularity on the social network on the distances traveled by users.

Huang and Butts [2023] investigated several socioeconomic characteristics and sought to understand which im-

[1] `https://h3-cities.streamlit.app/`

pacted migration between counties in the United States. This work examined the hypothesis that migrations occurred between similar areas, proposing a theory of segregation and social immobility about these movements. For their analyses, the authors use a temporal graph model. In this type of model, specific parameter settings allow an analysis of people's behavior in simulations where, for example, political segregation is disregarded.

Santin *et al*. [2020] analyze public transit mobility patterns of different economic classes in Curitiba, Brazil, using smart card data. The authors find that higher-income classes delay morning travel by about two hours and have more localized trips than lower-income groups. A transit mobility network reveals distinct spatial and temporal patterns across classes. The approach is validated by comparing it with household travel surveys, offering a cost-effective method for urban mobility studies and providing insights into the socioeconomic factors influencing urban transit usage. These findings are key to urban planning and sustainable development.

Senefonte *et al*. [2022] focus on understanding and exploring what drives international tourists' mobility patterns using data from LBSNs. The authors construct mobility descriptors, grouping users with similar behaviors based on interests captured to places visits. The approach identifies mobility patterns critical for tourism, revealing insights into how tourists explore new countries based on their previous travel profiles. The proposed approach enhances predictions compared to traditional models, making it valuable for urban planners and tourism service providers in optimizing services and improving tourist experiences through data-driven insights.

Silva and Silver [2024] reveal how people's movement patterns between places such as restaurants, parks, and shops, observed through Yelp, correlate with cultural attributes. The authors model the interactions between locations and individuals to study the cultural influences associated with urban movement and social behavior. They also explore the potential of using Graph Neural Networks (GNNs) to predict cultural traits by analyzing Yelp data.

Our research aims to understand people's interests in different city areas. Unlike these previous works, we do not capture and analyze users' mobility patterns; instead, we map users' preferences across urban areas and seek to understand these preferences through external factors.

2.3 Influence of Area Granularities

In this section, we discuss some of the most important previous studies that explore the need to understand the urban context by analyzing different granularities. For example, Rogov and Rozenblat [2018] noticed the absence of studies exploring cities' resilience in their multi-scalar entirety. While some works explored the ability to adapt to local impacts (micro level), such as natural disasters, and their consequences in the city (meso level), others investigated how shocks in the city system (macro level) influenced the city (meso level). They propose a framework for analyzing the city across these three scales to ensure a comprehensive understanding. In this context, they discuss the role of interactions between individuals (micro) that affect the urban characteristics of the city (meso) and the impact on the overall city system (macro). They also illustrate how, for example, economic difficulties within this system (macro) influence cities (meso) and alter individuals' daily lives (micro). Thus, they emphasize the importance of conducting studies at various levels of granularity.

Wu *et al*. [2020] employ the 2D discrete wavelet transformation method to analyze the spatial structure of Beijing, China, in a multi-scalar manner. To this end, they illustrate how analyzing smaller areas captures the details of specific locations but fails to describe the broader characteristics of the city. They argue that analyzing different granularities provides a holistic understanding of urban aspects, from area definitions to their roles in the city's overall structure.

Pafka [2022] shows how an urban analysis using statistical and/or administrative divisions led to a series of studies that are highly biased and difficult to compare. Therefore, they use a grid system of squares of different sizes to analyze city characteristics, such as the density of pedestrian-friendly areas and population density, in a standardized way in several cities worldwide. In this way, they defend the importance of consistently studying urban phenomena.

Based on those studies, we see the need to investigate the urban context through different granularities. However, differently from previous studies, we seek to understand the importance of urban areas' granularities in understanding urban behavior in the context of interest per area.

2.4 Identification of Similar Urban Zones

Some studies have explored new methods for mapping urban zones by integrating multiple LBSN data sources to identify and classify urban spatial patterns and functional areas [Cranshaw *et al*., 2012; Gao *et al*., 2017; Miao *et al*., 2021; Shouji Du and Zheng, 2020; Ye *et al*., 2021].

For instance, Cranshaw *et al*. [2012] suggest a new way of defining urban areas, the so-called Livehoods, dynamic areas of activity, recognizing that the arbitrary delimitations that form neighborhoods, for example, may not adequately reflect the reality of a city to urban planning. Therefore, they use data from Foursquare and a spectral clustering method to define these new areas based on the venues' similarity, the proximity of check-ins and venues, and the behavior of LBSN users. Furthermore, they validate these results in Pittsburgh through interviews with demographically diverse residents living in several parts of the city. They also used interviews with city hall professionals responsible for managing public resources and professionals in the real estate market.

Gao *et al*. [2017] developed a method capable of extracting topics from Foursquare data. These topics characterize areas based on their functionalities and activities carried out there. For example, a university and its surroundings can be considered an educational region. They use a probabilistic technique that classifies areas, considering the functions extracted directly from the data. As a result, the quantity and specificity of topics/functionalities analyzed may vary.

By combining physical features from remote sensing images with social attributes from Point of Interest data, researchers have achieved high accuracy in large-scale urban zone mappings [Shouji Du and Zheng, 2020]. Additionally, integrating social media data and street-level imagery has proven effective in recognizing urban functions, with verbs extracted from social media posts serving as proxies for human activities [Ye *et al.*, 2021]. These methods exemplify the potential value of insights for urban planning, management, and sustainability efforts regarding cities.

Unlike these studies, our proposal identifies urban zones through a grid-based system, where user behavior determines which cells are aggregated to form a zone, ensuring uniformity in their construction.

3 iNET Datasources

This section explains how data can be obtained for modeling iNETs, using information gathered from users located in different neighborhoods within the city of Curitiba, Brazil, London, United Kingdom, as well as for 20 of the main American counties and 20 of the main American cities. The section also details the characteristics of the resulting datasets and information about the collection of socioeconomic characteristics of Curitiba neighborhoods. Curitiba has been chosen because the authors are most familiar with the city, allowing for a thorough investigation of the factors related to users' interests. London was chosen because it is a city with a significant amount of data in both LBSNs. The data from both the LBSNs are more abundant in the USA, which is why they were used across different cities.

3.1 *LBSN Datasets*

Google Places: The Google Places dataset is built by extracting reviews carried out by users of the Google Plus social network in venues registered with the Google Maps service. The data are made available by the authors of He *et al.* [2017] and Pasricha and McAuley [2018] for academic use. From this dataset, we extract data from the city of **Curitiba**, Brazil, **London**, United Kingdom, and **American counties** (name, followed by state): New York, New York; Los Angeles, California; Cook, Illinois; Clark, Nevada; Maricopa, Arizona; District of Columbia, District of Columbia; San Francisco, California; Harris, Texas; San Diego, California; Orange, Florida; Fulton, Georgia; Miami-Dade, Florida; Philadelphia, Pennsylvania; Milwaukee, Wisconsin; Orange, California; King, Washington; Kings, New York; Suffolk, Massachusetts; Dallas, Texas; Travis, Texas; and **American cities**: New York, New York; Chicago, Illinois; Los Angeles, California; Washington, District of Columbia; San Francisco, California; Philadelphia, Pennsylvania; Houston, Texas; Boston, Massachusetts; Atlanta, Georgia; Paradise, Nevada; Austin, Texas; San Diego, California; Milwaukee, Wisconsin; San Antonio, Texas; Dallas, Texas; Seattle, Washington; Indianapolis, Indiana; Phoenix, Arizona; Charlotte, North Carolina; Nashville, Tennessee. The Google Places (G.P.) dataset includes specific information such as the user's name, education level, employment details, review text (in multiple languages), review score, the time the review has been posted, and a unique user identifier. This dataset also contains specific information about the evaluated venue, such as its name, category, opening hours, contact phone number, address, latitude, and longitude.

Foursquare: We also use publicly available data, extracted from Foursquare via check-ins shared on the social network Twitter. Our Foursquare dataset is made up of user check-ins in the cities of **Curitiba**, Brazil, and **London**, United Kingdom, as well as the **aforementioned American regions**. It includes details such as the check-in date, the name and category of the venue where it occurred, and the user's unique identifier. This dataset was explored and made available by Silva *et al.* [2017b].

The amount and characteristics of available data for the analyzed regions are presented in Table 1, for Google Places, and Table 2 for Foursquare. These tables highlight that Google Places has more reviews than Foursquare check-ins, except in Curitiba, which shows fewer reviews than check-ins. Additionally, Foursquare exhibits fewer distinct categories than Google Places. This is because Foursquare uses more general categories (e.g., "Food"), while Google Places offers more specific ones (e.g., "Italian Restaurant"). Furthermore, Google Places data span a longer period than Foursquare's. It is important to note that although some cities analyzed for Google Places contain data before 2010, the volume is significantly lower compared to the 2010–2014 period. For this reason, we focus on the latter period to align with the available Foursquare data.

Table 1. Description of Google Places dataset

	Curitiba	London	USA	
			cities	counties
Reviews	8,372	178,231	1,191,934	1,632,165
Users	4,909	75,897	394,588	486,393
Venues	2,213	31,075	186,639	286,075
Categories	685	1,81	3,439	3,793
Period	from 2010 to 2014			

Table 2. Description of Foursquare data

	Curitiba	London	USA	
			cities	counties
Reviews	53,253	27,088	398,805	495,698
Users	5,116	9,128	83,414	85,946
Venues	8,523	11,104	122,808	164,655
Categories	368	427	562	567
Period	2014*	from 2012 to 2014		

* from April to June

When comparing the most evaluated categories in Curitiba for both LBSNs, see Figure 1, the differences between Google Places and Foursquare data are clear. For instance, Foursquare includes information about users' homes and workplaces, which is absent in Google Places. This raises an interesting question: can the modeled iNETs provide similar insights despite these differences?

A comparison between the number of Google Places reviews, and the number of Foursquare check-ins for each county/city in the USA can be seen in Figure 2 for American counties and in Figure 3, for American cities. It is evident that Google Places contains a larger volume of data, but Foursquare still provides a substantial amount of information.

(a) Google Places **(b)** Foursquare

Figure 1. Word Cloud of the 50 most reviewed or visited categories in Curitiba

Since this work focuses on modeling through iNETs the areas people frequent, we also examine how often users post a message using their unique identifiers on each platform. Additionally, we analyze the intervals between the reviews/check-ins made by users. We then examine the number of users who made at least X publications (reviews/check-ins) and the intervals between publications that occurred within specific periods. These results are presented in Table 3 for Google Places data and in Table 4 for Foursquare. The tables illustrate user engagement with each platform and present the intervals between reviews/check-ins in the analyzed areas.

It can be observed that Google Places users have a lower frequency of platform usage than Foursquare users. Additionally, there is a noticeable difference in the regularity of use between the two platforms. Google Places users tend to post reviews either in quick succession or at intervals greater than a week. In contrast, Foursquare users check in more frequently and rarely go more than a week between check-ins, particularly in Curitiba. Therefore, in addition to the differences in the categories, we identify variations in the data collected from the LBSNs concerning user engagement, particularly regarding the regular use of the tools provided by each platform.

The previous analyses underscore the importance of this study. Although the platforms Google Places and Foursquare exhibit significant differences, if they model urban phenomena similarly, they could reveal more general patterns about the urban environment.

3.2 Socioeconomic and Electoral data

To understand the factors associated with people's interests, information on the socioeconomic aspects of each neighborhood in Curitiba was obtained from the 2010 Brazilian Demographic Census conducted by the Brazilian Institute of Geography and Statistics (IBGE)[2]. This national survey aimed to portray the Brazilian population with its socioeconomic characteristics and provide a basis for public and private planning for the decade between 2000-2010. Aiming to understand the differences between the neighborhoods of Curitiba, we extract data on average monthly income and racial composition, the latter categorized as follows: White, Black, Brown, Yellow, and Indigenous. This information has been collected for all neighborhoods in Curitiba.

To gather electoral data in Curitiba, we utilize the information provided by the Regional Electoral Court (TRE)[3] regarding the second round of presidential elections in 2014. The selection of these data is motivated by the works of Liu *et al.* [2019] and Huang and Butts [2023], in which the discrepancy in the percentages of votes during federal elections from 2004 to 2020, particularly in 2008 and 2020, was utilized to assess political polarization between areas. For example, a high percentage of votes for Democrats in one region and a low percentage in another resulted in greater polarization.

The TRE data includes information on the number of voters who voted for each candidate at the polling place level, organized into ten electoral zones. To extract data at the neighborhood level, we tried a source from Curitiba city hall that links voting locations to neighborhoods; however, we only found this information for 2012 and 2016. Consequently, to determine the percentage of voters who supported a particular candidate in each neighborhood, we adopted the Google Maps API to associate the addresses of voting locations (from a total of 418 voting locations) to their respective neighborhoods, ensuring that the API results were specific to the city of Curitiba. Through this method, we successfully linked voting locations to their corresponding neighborhoods. Nonetheless, some neighborhoods lacked electoral data, specifically: Centro Cívico, Campina do Siqueira, Alto da Rua XV, Riviera, São Miguel, Caximba, Lamenha Pequena, São João, and Cascatinha. These are shown hatched in Figure 4c.

The socioeconomic data analyzed in the present work and the electoral data collected for the experiments are shown in Figure 4. The distribution of average income in the neighborhoods of Curitiba is illustrated in Figure 4a. To represent the racial composition, Figure 4b shows the percentage of people self-declared *Black* for each neighborhood of Curitiba. Using the electoral data for the neighborhoods, we counted the number of voters in each area and calculated the percentage that voted for the government candidate and those that voted for the opposition party. Political polarization can be visualized in Figure 4c. It is possible to see in Figure 4c that some of Curitiba's neighborhoods are in peripheral areas, and, for those who did not present data, their residents probably voted in nearby neighborhoods that belong to the same electoral zones. Although located in central areas, smaller neighborhoods like Centro Cívico lack electoral information in this study. One possible explanation is the same as that presented for other areas - residents of these areas also voted in nearby neighborhoods. Alternatively, the issue could stem from inaccuracies in the method used to associate voting locations with their respective neighborhoods.

In 2010, the income variation between the lowest and highest income neighborhoods was approximately sevenfold, and the self-declared black population did not exceed 7.2% of the total. In the 2014 elections, the percentage of voters for the government candidate did not reach 45% in any of the neighborhoods analyzed. Neighborhoods closer to the city center exhibited a higher concentration of income and a predominantly white population, along with a lower percentage of votes for the government candidate. Conversely, the highest percentage of government votes occurred in the more remote areas, particularly in the southern region, where neighborhoods have lower purchasing power and a greater proportion of self-declared Black residents.

[2] `https://sidra.ibge.gov.br`
[3] `https://www.tre-pr.jus.br/eleicoes/eleicoes-anteriores/eleicoes-2014`

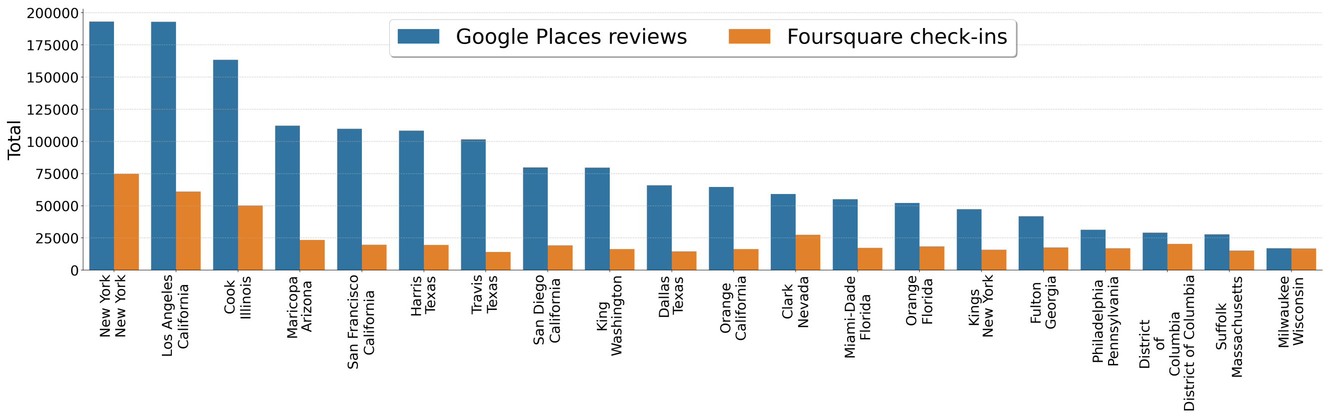


Figure 2. Google Places reviews (Blue) *vs.* FourSquare check-ins (Orange) in American **counties**

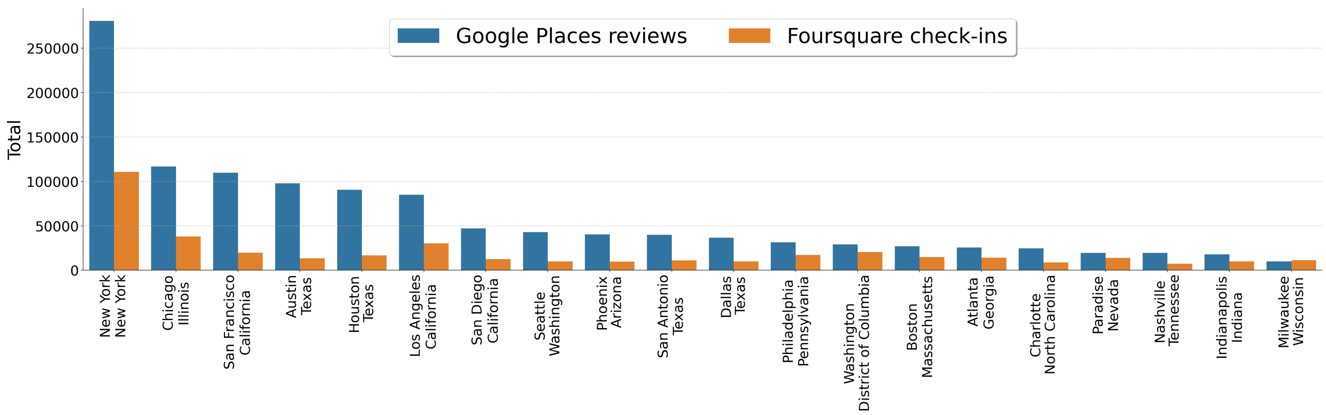


Figure 3. Google Places reviews (Blue) *vs.* FourSquare check-ins (Orange) in American **cities**

Table 3. Description of User Interaction Patterns on Google Places

	Recurrence in Reviews			Interval Between Reviews			
	at least 2 reviews	at least 5 reviews	at least 10 reviews	less than 6 hours	between 6 and 24 hours	between 1 day and 1 week	greater than 1 week
Curitiba	23%	4.7%	1.49%	50.3%	2.93%	9.45%	37.3%
London	27.1%	7.8%	3%	56.9%	3.3%	9.42%	30.3%
USA cities	25.8%	7.18%	2.72%	57.9%	2.8%	7.53%	31.7%
USA counties	25.5%	6.8%	2.5%	55.1%	2.76%	7.71%	34.2%

Table 4. Description of User Interaction Patterns on Foursquare

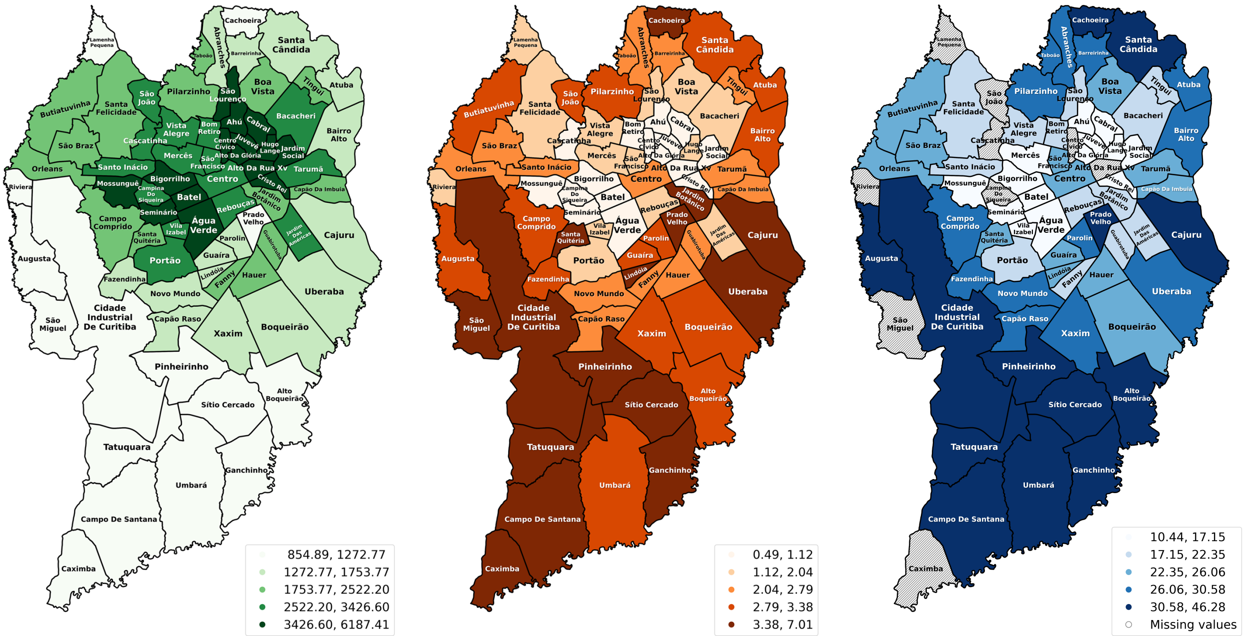
	Recurrence in Check-ins			Interval Between Check-ins			
	at least 2 *check-ins*	at least 5 *check-ins*	at least 10 *check-ins*	less than 6 hours	between 6 and 24 hours	between 1 day and 1 week	greater than 1 week
Curitiba	76.9%	47.6%	28.1%	36.8%	25.5%	28.7%	9%
London	52.3%	15.7%	4.6%	55%	24.7%	20.3%	0%
USA cities	59.8%	23.5%	8.56 %	52.8%	27.8%	19.4%	0%
USA counties	61.2%	25.1%	9.4%	53.7%	27.7%	18.6%	0%

4 Methods

This section is organized into six topics. The first subsection introduces iNETs. The second discusses methods for evaluating the granularity of iNETs. The third subsection examines common patterns across different iNETs. The fourth explores the relationship between external influences and user interests. The fifth subsection describes how these zones are defined based on user preferences. Finally, the last subsection analyzes the common characteristics among these urban zones.

4.1 Interest Networks (iNETs)

An iNET can help explain users' interest in urban areas [Santos *et al.*, 2024]. It is represented as a weighted, undirected



(a) Average Monthly Income (R$) **(b)** Self-Identified Black People (%) **(c)** Votes for the Government Party Candidate (%)

Figure 4. Distribution of Socioeconomic and Electoral Characteristics in Curitiba

graph $G = (V, E)$, where the set V of nodes represents the city's urban areas (such as grid cells, census tracts, zip codes, neighborhoods, or boroughs). An edge $e_{i,j} \in E$ connects urban area $v_i \in V$ to urban area $v_j \in V$, with the weight $w_{i,j} \in \mathbb{N}$ representing the number of users who have reviewed venues in both areas. These edges, therefore, represent users' interest in the two distinct urban areas. The constructed network also includes *self-loops*, meaning an edge $e_{i,i} \in E$ that connects a region $v_i \in V$ to itself. The weight $w_{i,i} \in \mathbb{N}$ on this edge represents the number of users who have reviewed venues in the same area at least twice. As this construction analyzes iNETs, it focuses only on edges with positive weights. To construct iNETs, we use data from users who reviewed at least twice, regardless of the venue

4.2 Testing iNETs Granularity

One of the objectives of this work is to investigate how users' interest in urban areas differs between the two LBSNs (Google Places and Foursquare) and how the size and shape of these areas affect this comparison. We developed a tool, called h3-cities[4], which utilizes OpenStreetMap and Uber's Hexagonal Hierarchical Geospatial Indexing System to subdivide a particular region into hexagons of various sizes. This allows for the division of any city with available OpenStreetMap data, providing a consistent, multi-scalar framework for urban analysis, as emphasized by Rogov and Rozenblat [2018] and Pafka [2022] in their review of the literature. For our analyses in Curitiba, Londres and USA using h3-cities, we tested four different hexagonal grid resolutions (hr): $h6$ with an average area of $36.12km^2$, $h7$ with $5.16km^2$, $h8$ with $0.74km^2$, and $h9$ with $0.11km^2$.

In addition to these areas delimitations, we consider some extra boundaries in specific analysis. In USA regions, we also use Census Tracts and Zip Codes. Figure 5 shows the region of Chicago with the highest amount of data on Google Places across all six levels of granularities used for that country. For Curitiba, we also used neighborhood delimitations. A comparison between the neighborhoods and the hexagonal system at resolution $h8$ is shown in Figure 6. We also used London's Boroughs for visualization purposes as will be show in Figure 15.

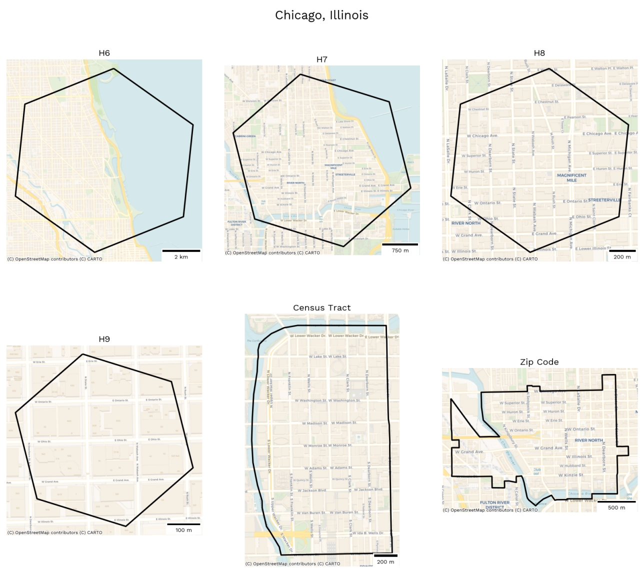


Figure 5. Comparison of different granularity levels in Chicago

Census tract and zip code delimitations have been obtained from the U.S. Census Bureau's Tiger geographic database[5]; and the neighborhood polygons from the Curitiba city hall

[4] `https://h3-cities.streamlit.app`

[5] `https://www.census.gov/cgi-bin/geo/shapefiles/index.php`

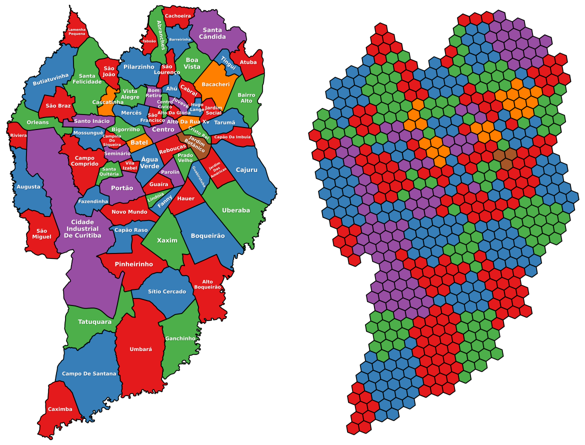


Figure 6. Comparison of the segmentation of Curitiba through neighborhoods *vs.* $h8$ granularity level - a subdivision via h3-cities

website[6]. London's boroughs were collected from the London Datastore[7].

4.3 Similarities Between the iNETs

To understand the similarity between the iNETs modeled by the different LBSNs, we employed Pearson correlation, which evaluates the presence and intensity of a linear relationship between two variables. Values close to 1 indicate a strong positive linear relationship, whereas values close to -1 indicate a strong negative linear relationship. When comparing the networks, we associate the weight of an edge in the Google Places network with its respective edge in the Foursquare network. Then, we calculate the Pearson correlation using the entire set of edges in both graphs. We also use the Spearman correlation for the edge weights, similar to the Pearson correlation.

Furthermore, we verified the similarity to the eigenvector centrality of both networks. This method indicates the importance of a node in the network based on the importance of the nodes with which it is connected. With it, we built a ranking in the analyzed granularities, and to investigate the similarity, we used Kendall's Tau correlation coefficient. This coefficient is used to evaluate the similarity between the orders of two datasets. Values close to 1 indicate a similar ordering, while values close to -1 suggest very different orderings. With these three methods, we have different ways to assess the similarity of the iNETs modeled by the different LBSNs, getting a more robust analysis. This way, we can check whether the interest (edges) are modeled similarly, as well as whether the most important areas of the city (nodes) are captured similarly.

4.4 Correlating External Factors and Interest

We also aim to understand whether additional factors—particularly socioeconomic, political, and geographic affect the interest of users from the analyzed areas in the two platforms studied. For each region, we considered the following factors: average monthly income, racial composition, political polarization, and geographic position. For brevity, we show the results for Curitiba, but this method can be applied to any city.

To achieve this goal, we correlate the edge weights with the distances between a given factor. For average income, we consider the absolute difference in average monthly income between two neighborhoods. For political polarization, we consider the results of the presidential election by region, similar to the work of Huang and Butts [2023] and Liu *et al.* [2019]. That is, we calculate the absolute difference between the percentages that voted for the presidential candidate for the areas. To compute the difference in racial composition, we use the same method as Huang and Butts [2023], in which the difference between neighborhoods A and B is defined as

$$R_{A,B} = \frac{1}{2}\sum_{i=1}^{n}\left|\frac{P_i(A)}{P(A)} - \frac{P_i(B)}{P(B)}\right|$$

where $R_{A,B}$ is the difference in racial compositions between neighborhoods A and B, $P(A)$ represents the population size of neighborhood A and $P_i(A)$ the size of the population that belongs to the i-th racial category in neighborhood A. We use the categories defined by the 2010 Brazilian Demographic Census: White, Black, Brown, Yellow, and Indigenous.

For the geographic distance, we use the centroids of each neighborhood and calculate the distance using the Python library *geopandas* [Jordahl *et al.*, 2020]. First, we transform the geographic coordinates of latitude and longitude from the WGS84 coordinate system to the UTM22S flat projection, which provides the most accurate projection of the region where Curitiba is located. Then, we calculate the geographic distance between the two neighborhoods in meters, looking at their centroids. The library uses the Euclidean distance for the points in the flat projection and, based on the projection, returns the geographic distance between the centroids of the neighborhoods in meters.

4.5 Formation of Urban Preference Zones

With the iNETs of Google Places and Foursquare modeled, we propose a new approach to defining urban preference zones. When analyzing the strongest interests within the iNETs—represented by the weight of the edges—we observe that much of the interaction occurs between neighboring areas. This suggests significant potential for using a grid system to model zones according to users' interests rather than relying on arbitrary boundaries. Our proposal offers a method for constructing zones of interest defined by the users of the analyzed platforms, based on the places they frequent and their spatial proximity.

To achieve this, we use iNET nodes to model densely connected geographically adjacent zones. These zones consist of cells with strong connections between them and spatial proximity. Specifically, we select from the iNETs only the edges between nodes whose hexagons share a vertex or edge, ensuring that the captured interests reflect interactions between nearby areas. In this work, we use the $h9$ resolution to represent the cells (nodes in the graph) and compare the results

[6] `https://ippuc.org.br/geodownloads/geo.htm`
[7] `https://data.london.gov.uk/dataset/statistical-gis-boundary-files-london`

obtained with this resolution against those using $h8$ in London. In this way, we can better understand the zones formed in a detailed manner, but the approach is universal.

With this subgraph identified for each iNET, the Leiden method [Traag *et al.*, 2019] is applied to detect communities within the graph, effectively capturing densely connected nearby areas. This method ensures the identification of internally connected areas, and through multiple iterations, it optimizes the communities found for maximum density. The algorithm continues iterating on the graph until no further improvements are achieved. This process is illustrated in Figure 7. We also use the resolution parameter $\gamma = 1$, following the value the method's authors applied in their applications [Traag *et al.*, 2019]. Lowering this value would result in larger areas while increasing it would produce smaller ones.

We use word clouds to improve the understanding of urban preference zones. Similar to Hu *et al.* [2015] for semantic differentiation of modeled areas, the natural language method TF-IDF (Term Frequency - Inverse Document Frequency) is used. This method emphasizes the importance of categories that are frequently evaluated by people within an analyzed area. At the same time, it reduces the relevance of categories that are common in several areas. Thus, high TF-IDF values are given to categories that are more frequent in an area but are also rarely mentioned in other areas. The goal is to understand the differences between these areas by observing the interests of users.

4.6 Similarity Between Urban Preference Zones

We also aim to understand the similarities between the constructions of urban zones of interest among the studied LBSNs. After constructing the zones by clustering grid cells, i.e., each cell is assigned to a community according to the Leiden method, an alignment is performed between these clusters using the Hungarian algorithm [Kuhn, 1955]. This method seeks the optimal alignment of clusters to facilitate subsequent analyses, as cluster A in the Google Places model may correspond to cluster B in Foursquare. Once this alignment is completed, the Normalized Mutual Information (NMI) metric is applied, where values close to 1 indicate significant mutual information shared between the clusters, and values near 0 indicate almost no mutual information. Additionally, the Rand Index [Hubert and Arabie, 1985] measures the similarity between two clusters by evaluating all pairs of samples and counting those that belong to the same clusters.

5 Results

This section is organized into four subsections. It covers the iNETs modeled from Google Places and Foursquare data and also the effect of different granularities on results in cities worldwide. Then, it analyzes the influence of socioeconomic, political, and geographic factors on edge strengths, showing results for Curitiba. Finally, it provides and discusses the identification of urban preference zones based on user behaviors and their resulting networks (UPZ-iNET), showing these results for London.

5.1 Modeled iNETs

To create the iNET for Google Places users in Curitiba, we considered 1,127 users – recall that we consider users with at least two reviews – and 4,590 reviews carried out by them. We construct Foursquare's iNET, considering 3,933 users and their 52,033 check-ins. We present in Figure 8 the iNETs generated for the city of Curitiba considering neighborhoods to illustrate how the interest networks are formed. For this case, we obtained a graph for Google Places comprising 69 neighborhoods (with available data) out of 75, containing 1,287 edges, 53 of which were self-loops. For Foursquare, we modeled a graph with 75 nodes and 1,618 edges, 58 of which are self-loops.

In the displayed graphs, the size of each node is proportional to its weighted degree, i.e. the sum of the edge weights connected to it, and the width of each edge is proportional to its weight. The most noticeable visual difference is that the network derived from Google Places shows interest in areas slightly farther from the city center, such as the Alto Boqueirão and Xaxim neighborhoods. Such a phenomenon is not observed in the network modeled with Foursquare data. Despite this, there is a considerable visual similarity between the iNETs constructed from the two LBSNs. Both graphs reveal a high concentration of activities in the central region, in neighborhoods such as Centro, Batel, Água Verde, Rebouças, and Alto da Rua XV, where retail, restaurant, and office sectors predominate. Other neighborhoods, such as Santa Felicidade and São Francisco, are known for their numerous typical restaurants, bars, burger joints, and casual pubs with live music. Meanwhile, neighborhoods like Bigorrilho, Centro Cívico, Portão, Seminário, Jardim Botânico, Bacacheri, and Cabral offer extensive green spaces with parks and squares, as well as commercial and leisure infrastructure, including shopping centers.

Regarding the other studied cities/counties, we modeled the iNETs using the same methodology applied to Curitiba, except for using other granularities instead of neighborhoods. For London considering $h9$ cells, for example, the Google Places iNET comprises 123,292 reviews from 67,276 users, forming 396,876 edges, including 1,533 self-loops, across 5,654 of the 6,600 cells covering the city. In contrast, the Foursquare network comprises 22,738 check-ins from 8,178 users, forming a graph with 34,075 edges, including 1,209 self-loops, across 2,899 cells.

Regarding the United States, in total, we modeled 240 iNETs for each LBSN – one for each of the 20 cities and 20 counties considered, at six different granularity levels. These iNETs are very diverse. The iNET in the United States with the lowest nodes is from New York county using $h6$ cells and Google Places data, with 7 nodes and the 28 edges between them, while the iNET with the most nodes is from Los Angeles county using $h9$ cells and Google Places data, with 10,363 nodes and the 455,976 edges between them.

5.2 Impact of Different Granularity Levels

To illustrate the differences between the studied LBSNs for different granularity levels, Figure 9 shows the areas of strongest user presence for each LBSN in the city of Chicago.

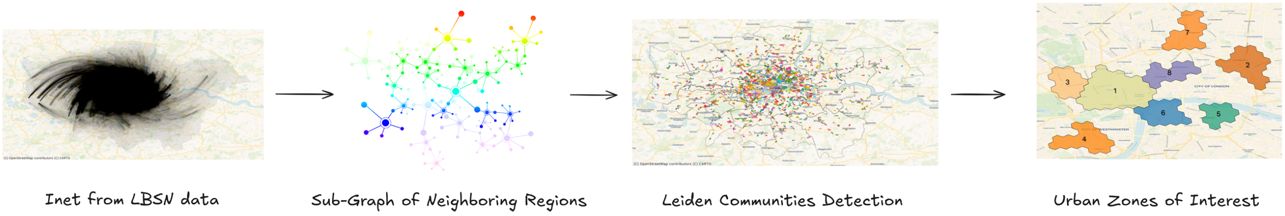
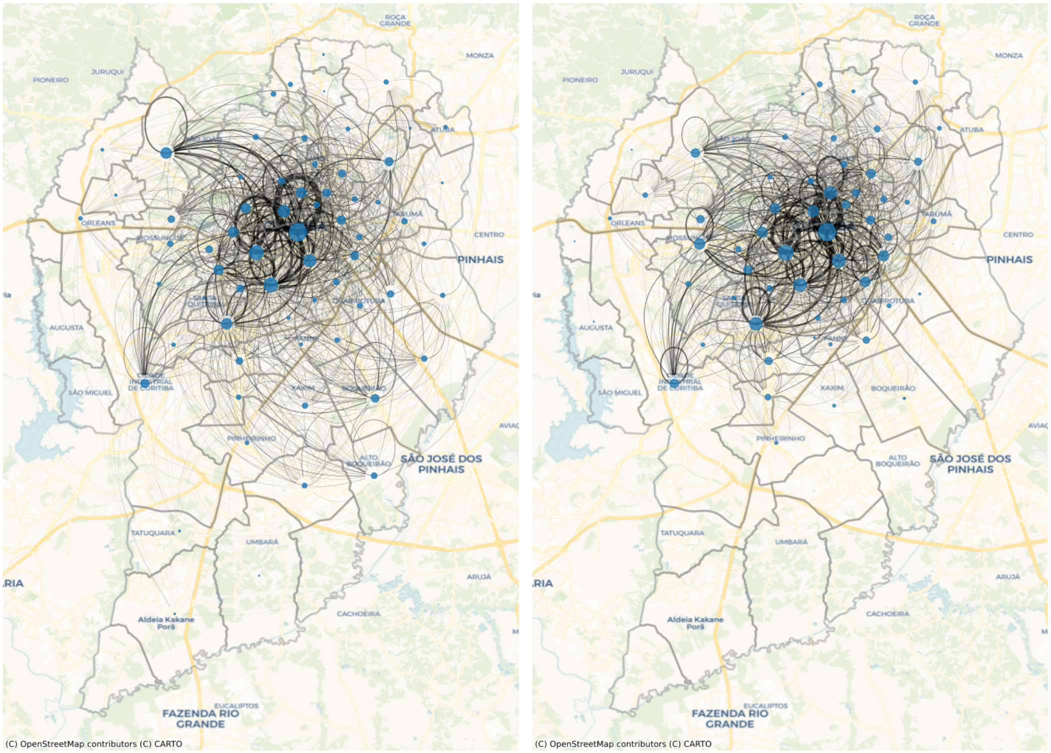


Figure 7. Formation of Urban Preference Zones (UPZones)



(a) The iNET from Google Places data **(b)** The iNET from Foursquare data

Figure 8. LBSN iNETs in the City of Curitiba

In this figure, darker colors indicate a higher number of reviews/check-ins, whereas white represents the absence of data. In addition to the spatial distribution of this information, the figure also presents a cumulative distribution function (CDF) that provides a general overview of the amount of data throughout the city. The figure illustrates the differences across several analyzed granularity levels, indicating a considerable tendency for larger areas to be represented similarly. Conversely, when smaller areas are used, there is a tendency for greater variability among the LBSNs.

With the 240 iNETs, for each LBSN, modeled in the USA (six granularities for each one of 20 cities and 20 counties), we conducted similarity analyses between them, as explained in Section 4.3 – one of them considers ranking. Before presenting the comparison results, to better understand how the ranking of urban areas is formed and compared, Figure 10 illustrates the process. It presents the 20 most central neighborhoods in Curitiba, ranked by eigenvector centrality in Google Places' iNET, along with their respective rankings in Foursquare's iNET. Visually, small variations in the order of the neighborhoods can be observed when comparing the two datasets. However, the rankings remain relatively close, indicating a moderate relationship between the centralities.

The results of the mentioned comparison are presented in Figure 11 for three different correlation metrics: Pearson, Spearmen, and Kendaltau. In this figure, the y-axis indicates the correlation value analyzed for each granularity level, while the x-axis presents, on a logarithmic scale, the number of nodes in both iNETs (Google Places and Foursquare).

Observing Figure 11, there is an evident pattern that the more nodes an iNET has, the greater the differences between the iNETs modeled by the studied LBSNs. It is also important to highlight that counties vary in area from $87km^2$ to $23,895km^2$ while cities vary from $177km^2$ to $1,740km^2$. As a result, some regions have more nodes at the $h7$ granularity level than other regions at the $h8$ level, even if the $h8$ nodes are smaller than the $h7$.

Figure 12 shows the results of the similarity correlations between iNETs from Google Places and Foursquare in Curitiba, with divisions into neighborhoods and $h6$ to $h9$ granularity levels. It is worth noting that Curitiba's neighborhoods vary in size between $h7$ and $h8$ levels. The results for the iN-

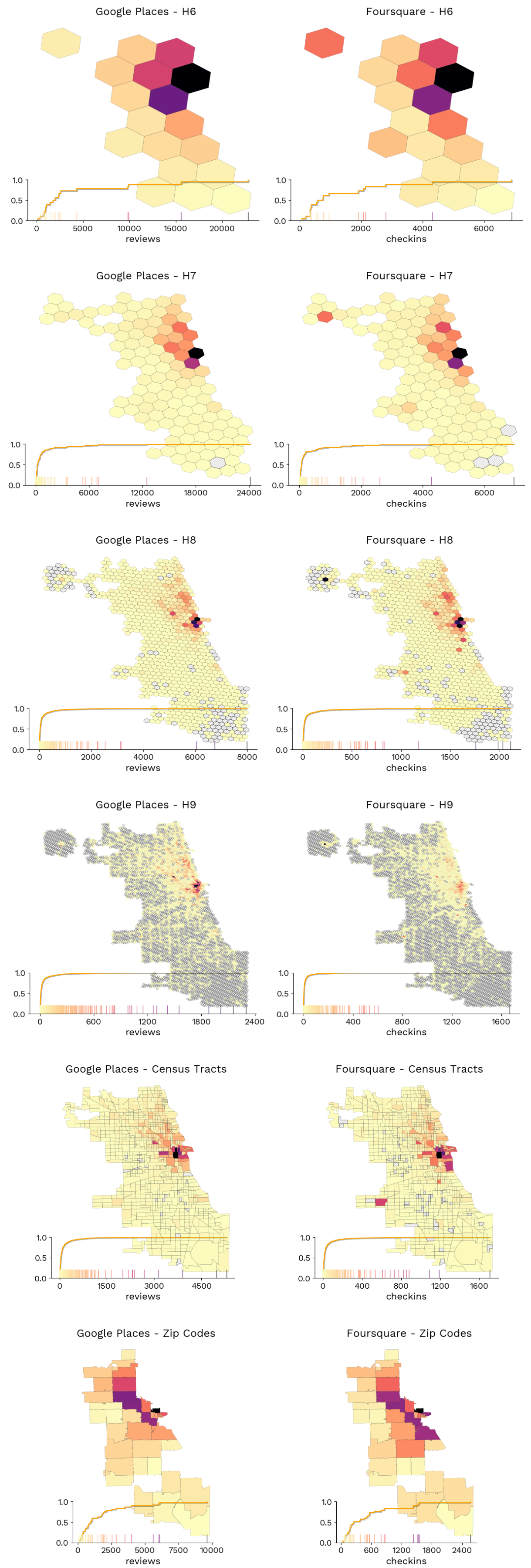


Figure 9. The subdivisions from Google Places and Foursquare data in the city of Chicago: analyses for six different granularity levels

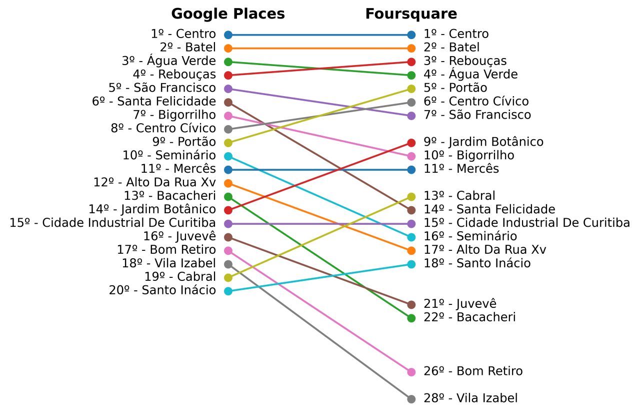


Figure 10. Comparison of the most central neighborhoods of Curitiba concerning centrality by eigenvector

ETs in London are shown in Figure 13 based on the analyzed granularity levels. These findings reinforce the notion that the smaller the urban areas analyzed, the greater the differences among the iNETs.

In the quest to understand users' interests in different urban areas, these results provide crucial information. For instance, when investigating the central areas of the city and their relationships with other urban areas from a broader perspective, the granularity level $h6$, with areas of $36.12km^2$, can be particularly important. This level tends to yield similar results regardless of the chosen LBSN. In contrast, analyzing the urban landscape at the $h7$ and $h8$ levels results in smaller urban areas, which can provide more detailed insights than larger areas obtained with $h6$. However, these finer granularity levels may lead to the emergence of more diverse iNETs, potentially resulting in a loss of relevant information and capturing only the behaviors of users specific to each LBSN.

The $h9$ granularity level exhibits the most significant differences in the LBSNs. This occurs because $h9$ areas are much smaller, allowing for better differentiation between the LBSNs. However, due to the limited data and a much larger number of nodes in the formed iNETs, noise presence can be more pronounced. When examining census tracts, zip codes, as shown in figure 9, and neighborhoods, as illustrated in figure 8, we find that the census tracts and neighborhoods are similar in the formed iNETs comparable to the $h7$ and $h8$ granularity levels. In contrast, zip codes show a greater similarity to the $h6$ level in terms of homogeneity among iNETs associated with both LBSNs. These findings align closely with the results reported by Wu *et al.* [2020], as smaller areas capture different details compared to larger ones within a city. Thus, each granularity offers distinct and complementary insights into urban phenomena.

An advantage of analyzing divisions used by government entities is that their associated socioeconomic data help investigate factors related to users' interest in urban areas. However, using different granularity levels provides, for example, the flexibility to look at an urban landscape capturing overall patterns, as in $h6$ level, or examine the relationship between small areas, with $h9$ level.

Figure 11. Results of correlation between both iNETs in USA

5.3 Influence of Socioeconomic, Political and Geographic Factors

This section aims to determine whether additional factors – particularly socioeconomic, political, and geographic – affect the interests of people who frequent the analyzed areas

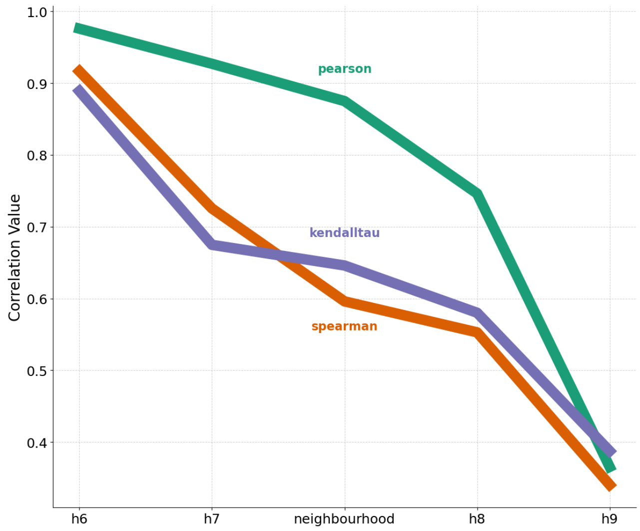


Figure 12. Results of correlation between both iNETs in Curitiba

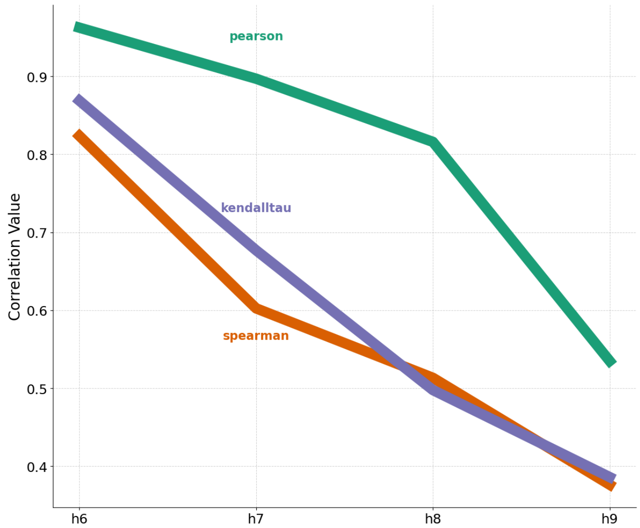


Figure 13. Results of correlation between both iNETs in London

through a study case in Curitiba. We exemplify this analysis for Curitiba using neighborhoods as graph nodes due to their association with socioeconomic data.

As explained in Section 4.3, we correlate the edge weights of the modeled networks with the distances associated with a given factor for each pair of nodes in the graph.

To gain further insight into how these distances apply to the city of Curitiba, Figure 14 presents a subgraph showing the distances for political and socioeconomic factors in some of the most central neighborhoods in Curitiba. The node color is proportional to the factor value being analyzed: darker colors represent higher values, while lighter colors correspond to lower values within the subset of the addressed neighborhoods. For political polarization, darker nodes indicate a higher percentage of votes for the government presidential candidate. For racial composition, the percentage of individuals identified as Black in each neighborhood is represented. For average monthly income, darker nodes correspond to higher income levels. The edge color reflects the difference between neighborhood factors: darker edges indicate greater distances, while lighter ones represent smaller differences. In Figures 14a, 14b and 14c, the edge differences account for the political, racial and income compositions, respectively, as previously explained.

Intending to gain deeper insight into the factors influencing visitors between pairs of nodes, we relate each edge weight to its distance across the analyzed characteristics (average monthly income, racial composition, political polarization, geographic position), for both iNETs. Using Spearman correlation, the results for both networks, considering information from both complete iNETs (shown in Figure 8) and filtered iNETs, are presented in Table 5, for Google Places networks, and in Table 6, for Foursquare networks.

Table 5. Results of Spearman Correlations with the Analyzed Factors

	Google Places Complete	Google Places Filtered
Average monthly income	-0.09	-0.10
Racial Composition	-0.25	-0.19
Political Polarization	-0.23	-0.14
Geographic Distance	-0.38	-0.30

Table 6. Results of Spearman Correlations with the Analyzed Factors

	Foursquare Complete	Foursquare Filtered
Average monthly income	0.06	-0.04
Racial Composition	-0.11	-0.20
Political Polarization	-0.09	-0.15
Geographic Distance	-0.55	-0.44

In complete iNETs, we observe the impact of all edges found. In the filtered iNET, we only consider edges with a weight greater than or equal to 5, focusing only on areas connected by higher interest. For the filtered iNET, the Google Places modeling shows only 418 edges out of the total $1,287$ (32.5%) edges in the full network, while the Foursquare data show $1{,}169$ edges out of the total $1,618$ (72.2%).

The results suggest that average monthly income, racial composition, and political polarization do not significantly explain visitors' interest in specific areas. In other words, people do not necessarily visit places with similar or dissimilar income levels, racial demographics, or political views, as reflected in the behavior captured by these LBSNs.

Across all scenarios, geographic distance shows the strongest negative correlation, as seen in Tables 5 and 6. This indicates that the greater the geographic distance between two neighborhoods, the lower the edge weights. Notably, the correlation with geographic distance is stronger in the iNET modeled by Foursquare. A plausible explanation lies in the differing ways users engage with LBSNs. This result aligns with expectations, as several studies [Cheng *et al.*, 2021; González *et al.*, 2008; Rhee *et al.*, 2008; Brockmann *et al.*, 2006] highlight the tendency for users' travel distances — both in LBSNs and similar datasets — to cluster at shorter distances and become increasingly rare over greater distances, as we have demonstrated.

The results for racial composition, though relatively weak, show the second-highest correlation across all scenarios, making it an interesting finding, as it was less expected compared to geographic distance. Moreover, there is supporting evidence for this outcome. As shown in de la Prada and Small [2024], racial differences between areas tend to increase with distance, up to a threshold of 10 km. This insight opens new possibilities for exploring other factors resembling these factors and aims to explain interest in urban areas.

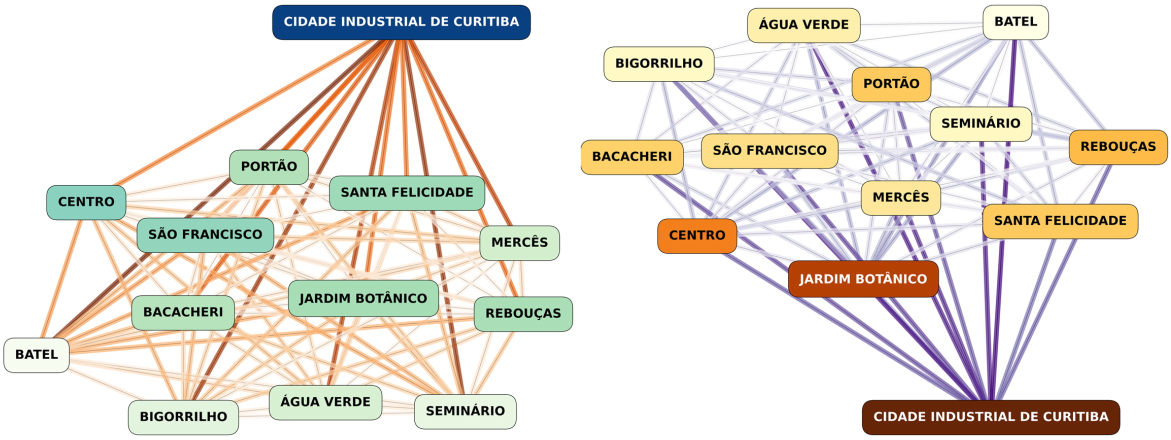
For brevity, we focused the analysis in this section mostly on the Curitiba results. However, our study indicates that similar results concerning the influence of socioeconomic, political, and geographic factors can be found in other cities as well.

5.4 Urban Preference Zones (UPZones)

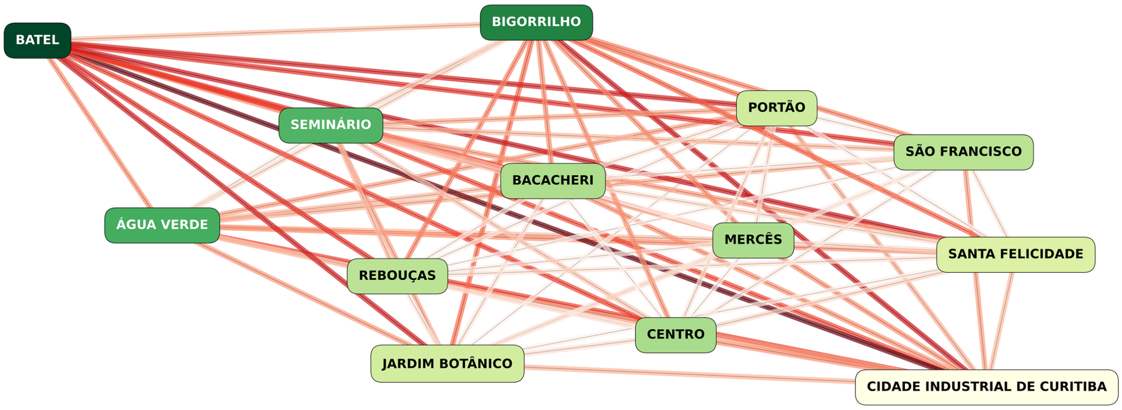
Using the method described in Section 4.5 applied in London, with the h9 cells, we identified 1,760 UPZones on Google Places and 1,023 on Foursquare. Figure 15 illustrates the results of this process, compared with the divisions of the city into boroughs. In this figure, each color represents a distinct UPZone; however, identical colors in non-adjacent areas indicate different zones. The UPZones are delineated by borders, while the significantly larger areas, or boroughs, are marked by gray boundaries. This comparison highlights that constructing urban preference zones can provide more detailed insights into user interest by better capturing the intricacies of the urban landscape.

Figure 16 illustrates how the construction of UPZones from the iNET with $h9$ granularity level captures semantically relevant areas within the city. In this figure, each pair, color and number, corresponds to an UPZone based on Google Places data, allowing us to observe their spatial distribution in London. The eight UPZones with the highest activity levels have been selected (the numbers indicate their ranking). Additionally, the figure features a word cloud that visualizes the most prominent categories in each region (see Section 4.6), highlighting categories that are more frequently represented in each zone.

The figure illustrates that, for example, UPZone 1 encompasses one of the city's most popular areas, known as Soho, renowned for its vibrant nightlife and diverse array of restaurants. UPZone 2 is a celebrated region recognized for its artistic activity, serving as a hub for artists and a venue for various performances. Meanwhile, UPZone 3 represents the area known as Mayfair, famous for its clothing and accessories shops. UPZone 6 captures the South Bank, noted for its cinemas, art galleries, and iconic tourist attractions like the London Eye. The other UPZones also reflect well-known areas visited by users with similar demographic interests. It's worth mentioning that the central area of the city was similarly captured by Foursquare data, although the considerably smaller dataset may affect this comparison. The proposed method effectively identifies urban preference zones within a city, enhancing our understanding of user preferences. By combining this technique with insights about the areas and conducting a detailed analysis, we can uncover regions in the city that do not strictly adhere to predefined geographic boundaries, thereby offering a richer understanding of urban space utilization.

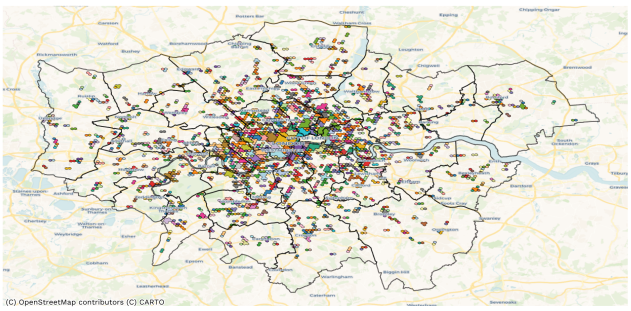


(a) Political Polarization **(b)** Racial Composition

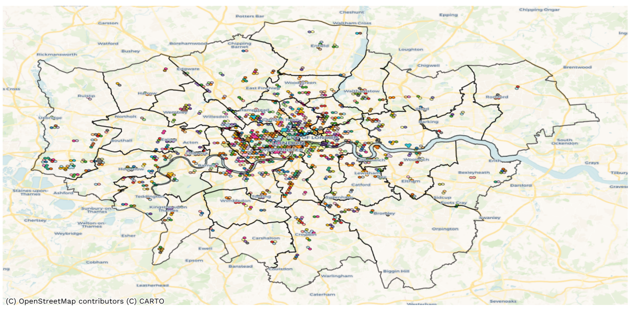


(c) Average Monthly Income

Figure 14. Sub-Graph of Distances Between Socioeconomic Factors



(a) Google Places



(b) Foursquare

Figure 15. Urban Preference Zones (UPZones) in the City of London, UK. Different colors only indicate different zones. For comparison, official delimitations of boroughs are drawn in gray lines on the map

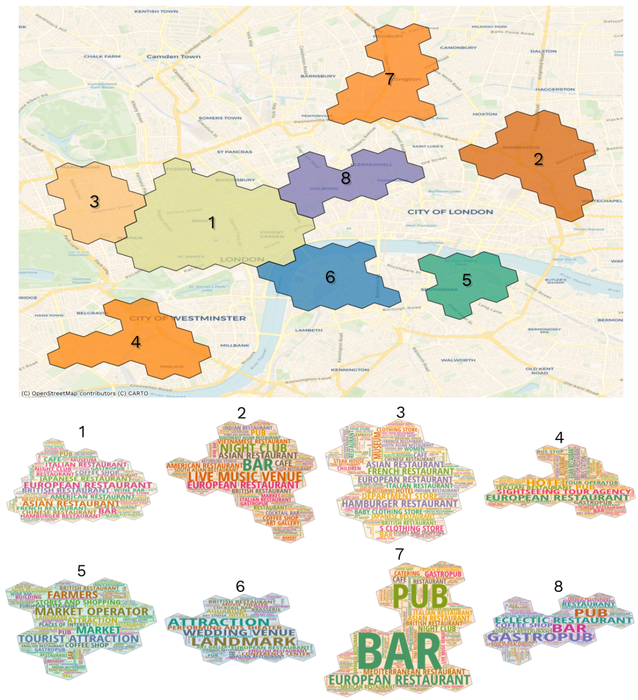


Figure 16. Characterization of the main UPZones found

The modeling of the city's areas of interest was also explored using the cell resulting from $h8$ granularity level as the iNET node, following the same method described earlier. However, as shown in Figure 17, which compares the 8 zones with the highest activity for each resolution (Red - $h8$; Brown - $h9$), it becomes evident that, although both resolutions capture densely connected and nearby cells, the $h8$ resolution forms much larger zones. This results in a loss of detail and subtle insights that the finer $h9$ resolution provides.

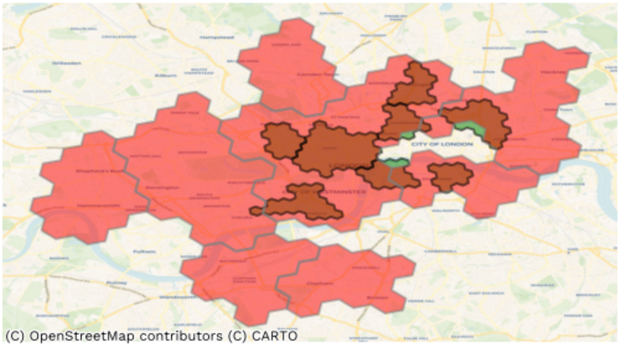


Figure 17. Comparing resolution $h8$ (red) and $h9$ (brown) in modeling UPZones

Using the metrics to compare UPZones defined in Section 4.6, we analyzed the urban scenarios constructed from the Google Places and Foursquare iNETs. This analysis yielded an NMI score of 0.6364. Additionally, the Rand Index obtained from comparing the UPZones from Google Places and Foursquare was 0.7378. These results indicate that when employing UPZones to model users' interests, a notable similarity is revealed despite the differences in the iNETs of the studied LBSNs. Thus, it can be concluded that, regardless of the LBSN utilized, moderately similar UPZones can be derived.

For brevity, we presented results only for London. However, we also did the same experiment on Curitiba, and the message observed is the same. In this way, we have an indication that UPZones can be applied to different cities, offering an enhanced insight into urban space usage.

5.5 UPZ-iNETs: Interest Network of UPZones

The urban preference zones also enable a qualitative analysis of the interest between zones by visualizing which UPZones of the city are most interconnected through user activity. To achieve this, iNETs can be built with UPZones as graph nodes, following the same methodology used to construct the iNETs. This approach allows us to identify zones that share mutual user interests. The resulting network, referred to as UPZ-iNET, is shown in Figure 18 based on Google Places data. To enhance readability and highlight key connections, we retained only the top 0.1% of edges and employed arrows for illustrative purposes. These undirected edges represent the main links between zones, with weights ranging from 203 to $1,456$, corresponding to the number of users who visited both areas. Thicker, darker edges denote stronger connections, reflecting a higher number of shared visitors between zones.

Figure 18 reveals a strong connection, for instance, between the central area and several other regions, including non-central parts of the city. This visualization not only highlights the patterns of user interest between different areas but also supports a qualitative analysis of the zones frequented by the same users. Through this approach, we provide a method for modeling urban areas in any city based solely on LBSN user activity, providing insights into the interaction between these urban spaces.

6 Conclusion

Understanding urban behavior is a complex task. In this work, we explored whether two different Location-Based Social Networks (LBSNs), Google Places and Foursquare, could provide comparable insights when modeling users' interests in geographic areas. We also examined how the modeling of LBSN data influences the definition and understanding of urban areas. Through an analysis of the characteristics of two datasets (resulting from Google Places and Foursquare LBSN data) and information collected from Curitiba, London, and several U.S. cities and counties, we found that the resulting graphs — referred to as Interest Networks (iNETs) — effectively capture the dynamics of users' behavior. These iNETs exhibit significant similarities, particularly in connections involving the most central urban areas at the $h6$ granularity level, while greater differences emerge when smaller spatial units are employed for analysis (e.g., $h8$ and $h9$ granularity levels).

Additionally, we investigated whether LBSN users' interest in urban areas could be understood through geographic distance, political polarization, and socioeconomic characteristics from the areas they visit. Our findings indicate that, for the analyzed data, factors such as average income, racial composition, and political polarization of the areas (e.g., neighborhoods in Curitiba) do not sufficiently explain users' preferences. However, we observed that geographic distance plays a limiting role in interactions, as users tend to visit nearby areas.

Another aspect analyzed was the potential to minimize the differences observed between the iNETs derived from smaller areas (e.g., $h8$, $h9$). We proposed a method for defining urban preference zones (UPZones) within a city, emphasizing the capture of densely connected areas. The iNETs formed from these zones, referred to as UPZ-iNETs, exhibited greater similarity across the two LBSNs analyzed.

It is important to highlight that the use of LBSNs must be accompanied by a critical understanding of the limitations and implications of their applicability. For instance, data from Google Places and Foursquare may not fully capture the interest of the entire population, as users of these platforms tend to be younger individuals with access to mobile internet. Nonetheless, these platforms provide valuable insights into the behavior of this demographic, which may or may not reflect broader societal patterns. Additionally, it is important to note that the data used in this study were collected a decade ago, meaning that the results might differ if more recent datasets were analyzed. However, the methodology remains applicable, and since the data sources used are publicly accessible, they offer a solid foundation for further exploration in urban computing research.

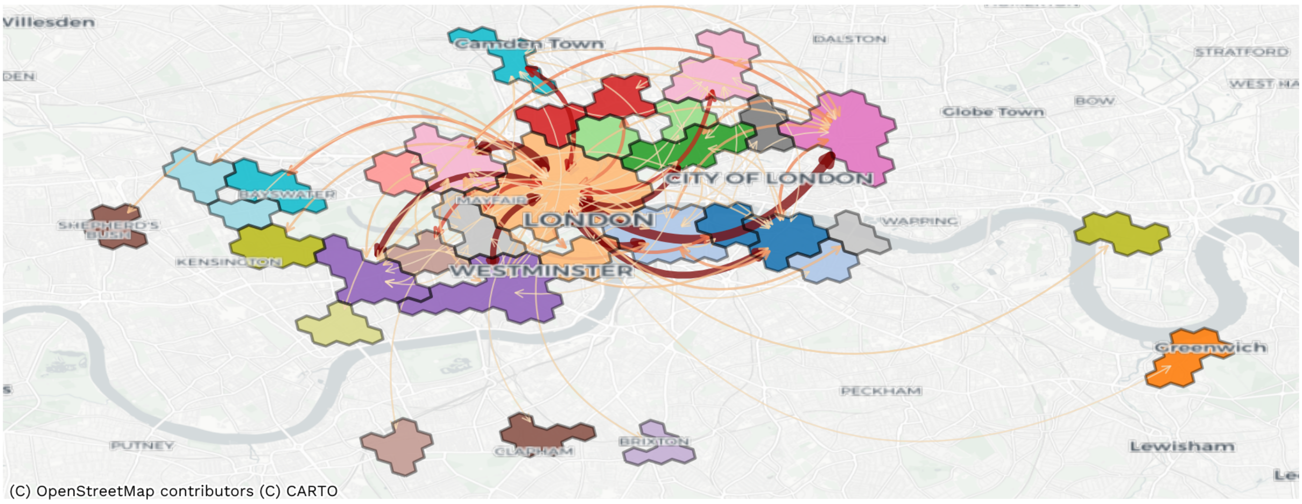


Figure 18. UPZ-iNET: Interest network of urban preference zones

In future research, it would be valuable to analyze more recent datasets to determine whether different LBSNs continue to yield similar insights and to explore how factors influencing iNETs may change over time. This exploration could include examining additional variables such as cultural influences, types of venues, and the content of user reviews, including sentiment analysis and topic modeling. These factors would enrich the analysis and help address questions like whether the places frequented by users in urban environments share cultural similarities, whether individuals are inclined to visit the same types of venues across different areas, and whether certain regions are characterized by specific venue categories. Furthermore, it would be interesting to investigate whether the most interconnected areas within an iNET reflect similar sentiments among users. This comprehensive approach would enhance our understanding of urban behavior and the dynamics of city life concerning venues people commonly visit.

With a better understanding of the factors influencing the interests of the populations in each city, these analyses could be integrated into recommendation systems that highlight the unique characteristics of each neighborhood. Additionally, this type of investigation has the potential to inform public bodies about the most interconnected areas, enabling the development of public policies aimed at combating epidemics or promoting social integration between previously disconnected regions. Furthermore, a deeper exploration of the formation of urban preference zones, combined with validation from residents and experts in various cities, could enhance public policy systems for more effective management of available resources.

Declaration

Acknowledgements

This research was partially supported by the SocialNet project (process 2023/00148-0 of the São Paulo Research Foundation - FAPESP), by the National Council for Scientific and Technological Development - CNPq (processes 313122/2023-7, 314603/ 2023-9, 441444/2023-7, and 444724/2024-9). This research is also part of the INCT of Intelligent Communications Networks and the Internet of Things (ICoNIoT) funded by CNPq (proc. 405940/2022-0) and Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) Finance Code 88887.954253/2024-00.

Funding

This manuscript did not receive any external funding.

Authors' Contributions

GS performed the experiments. GS, FG, MD, and TS helped in the conceptualization of the study and writing of the manuscript. GS is the main contributor and writer of this manuscript. All authors read and approved the final manuscript.

Competing interests

The authors declare that they have no competing interests.

Availability of data and materials

The tool h3-cities used in the study is available at: `https://h3-cities.streamlit.app/`

References

Brockmann, D., Hufnagel, L., and Geisel, T. (2006). The scaling laws of human travel. *Nature*, 439(7075):462–465. DOI: 10.1038/nature04292.

Cheng, Z., Caverlee, J., Lee, K., and Sui, D. (2021). Exploring millions of footprints in location sharing services. *Proceedings of the International AAAI Conference on Web and Social Media*, 5(1):81–88. DOI: 10.1609/icwsm.v5i1.14109.

Cranshaw, J., Schwartz, R., Hong, J., and Sadeh, N. (2012). The Livehoods Project: Utilizing Social Media to Understand the Dynamics of a City. *Proceedings of the International AAAI Conference on Web and Social Media*, 6(1):58–65. DOI: 10.1609/icwsm.v6i1.14278.

de la Prada, A. G. and Small, M. L. (2024). How people are exposed to neighborhoods racially different from their own. *Proceedings of the National Academy of Sciences*, 121(28). DOI: 10.1073/pnas.2401661121.

Ferreira, A. P., Silva, T. H., and Loureiro, A. A. (2020). Uncovering spatiotemporal and semantic aspects of tourists mobility using social sensing. *Computer Communications*, 160:240–252. DOI: 10.1016/j.comcom.2020.06.005.

Ferreira, A. P. G., Silva, T. H., and Loureiro, A. A. F. (2015). Beyond sights: Large scale study of tourists' behavior using foursquare data. In *2015 IEEE International Conference on Data Mining Workshop (ICDMW)*, pages 1117–1124. DOI: 10.1109/ICDMW.2015.234.

Gao, S., Janowicz, K., and Couclelis, H. (2017). Extracting urban functional regions from points of interest and human activities on location-based social networks. *Transactions in GIS*, 21(3):446–467. DOI: 10.1111/tgis.12289.

González, M. C., Hidalgo, C. A., and Barabási, A.-L. (2008). Understanding individual human mobility patterns. *Nature*, 453(7196):779–782. DOI: 10.1038/nature06958.

He, R., Kang, W.-C., and McAuley, J. (2017). Translation-based recommendation. In *Proceedings of the Eleventh ACM Conference on Recommender Systems*, RecSys '17. ACM. DOI: 10.1145/3109859.3109882.

Hu, Y., Gao, S., Janowicz, K., Yu, B., Li, W., and Prasad, S. (2015). Extracting and understanding urban areas of interest using geotagged photos. *Computers, Environment and Urban Systems*, 54:240–254. DOI: 10.1016/j.compenvurbsys.2015.09.001.

Huang, P. and Butts, C. T. (2023). Rooted america: Immobility and segregation of the intercounty migration network. *American Sociological Review*, 88(6):1031–1065. DOI: 10.1177/00031224231212679.

Hubert, L. and Arabie, P. (1985). Comparing partitions. *Journal of Classification*, 2(1):193–218. DOI: 10.1007/BF01908075.

Jordahl, K., den Bossche, J. V., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., Tratner, J., Perry, M., Badaracco, A. G., Farmer, C., Hjelle, G. A., Snow, A. D., Cochran, M., Gillies, S., Culbertson, L., Bartos, M., Eubank, N., maxalbert, Bilogur, A., Rey, S., Ren, C., Arribas-Bel, D., Wasser, L., Wolf, L. J., Journois, M., Wilson, J., Greenhall, A., Holdgraf, C., Filipe, and Leblanc, F. (2020). geopandas/geopandas: v0.8.1. DOI: 10.5281/zenodo.3946761.

Kuhn, H. W. (1955). The Hungarian method for the assignment problem. *Naval Research Logistics Quarterly*, 2(1-2):83–97. DOI: 10.1002/nav.3800020109.

Ladeira, L., Souza, A., Filho, G. R., Silva, T. H., and Villas, L. (2019). Serviço de sugestão de rotas seguras para veículos. In *Anais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 608–621, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/sbrc.2019.7390.

Liu, X., Andris, C., and Desmarais, B. A. (2019). Migration and political polarization in the u.s.: An analysis of the county-level migration network. *PLOS ONE*, 14(11):e0225405. DOI: 10.1371/journal.pone.0225405.

Martí, P., Serrano-Estrada, L., and Nolasco-Cirugeda, A. (2019). Social media data: Challenges, opportunities and limitations in urban studies. *Computers, Environment and Urban Systems*, 74:161–174. DOI: 10.1016/j.compenvurbsys.2018.11.001.

Miao, R., Wang, Y., and Li, S. (2021). Analyzing urban spatial patterns and functional zones using sina weibo poi data: A case study of beijing. *Sustainability*, 13(2). DOI: 10.3390/su13020647.

Nolasco-Cirugeda, A. and García-Mayor, C. (2022). Social dynamics in cities: Analysis through lbsn data. *Procedia Computer Science*, 207:877–886. DOI: 10.1016/j.procs.2022.09.143.

Pafka, E. (2022). Multi-scalar urban densities: from the metropolitan to the street level. *URBAN DESIGN International*, 27(1):53–63. DOI: 10.1057/s41289-020-00112-y.

Pasricha, R. and McAuley, J. (2018). Translation-based factorization machines for sequential recommendation. In *Proceedings of the 12th ACM Conference on Recommender Systems*, RecSys '18. ACM. DOI: 10.1145/3240323.3240356.

Rhee, I., Shin, M., Hong, S., Lee, K., and Chong, S. (2008). On the levy-walk nature of human mobility. In *IEEE INFOCOM 2008 - The 27th Conference on Computer Communications*. IEEE. DOI: 10.1109/infocom.2008.145.

Rogov, M. and Rozenblat, C. (2018). Urban Resilience Discourse Analysis: Towards a Multi-Level Approach to Cities. *Sustainability*, 10(12):4431. DOI: 10.3390/su10124431.

Santala, V., Miczevski, S., de Brito, S. A., Baldykowski, A. L., Gadda, T., Kozievitch, N., and Silva, T. H. (2017). Making sense of the city: Exploring the use of social media data for urban planning and place branding. In *Anais do I Workshop de Computação Urbana*, Porto Alegre, RS, Brasil. SBC. Available at:`https://sol.sbc.org.br/index.php/courb/article/view/2577`.

Santin, P., Gubert, F. R., Fonseca, M., Munaretto, A., and Silva, T. H. (2020). Characterization of public transit mobility patterns of different economic classes. *Sustainability*, 12(22). DOI: 10.3390/su12229603.

Santos, G., Gubert, F., Delgado, M., and Silva, T. (2024). Redes de interesse: comparando o google places e foursquare na captura da escolha de usuários por áreas urbanas. In *Anais do VIII Workshop de Computação Urbana*, pages 99–112, Porto Alegre, RS, Brasil. SBC. DOI: 10.5753/courb.2024.3248.

Senefonte, H. C. M., Delgado, M. R., Lüders, R., and Silva, T. H. (2022). Predictour: Predicting mobility patterns of tourists based on social media user's profiles. *IEEE Access*, 10:9257–9270. DOI: .

Shouji Du, Shihong Du, B. L. X. Z. and Zheng, Z. (2020). Large-scale urban functional zone mapping by integrating remote sensing images and open social data. *GIScience & Remote Sensing*, 57(3):411–430. DOI: 10.1080/15481603.2020.1724707.

Silva, T. H., de Melo, P. O. S. V., Almeida, J. M., and Loureiro, A. A. F. (2017a). Uma fotografia do instagram: Caracterização e aplicação. Available at:`http://143.54.25.88/index.php/RB-RESD/article/view/74`.

Silva, T. H., de Melo, P. O. V., Almeida, J. M., Musolesi, M., and Loureiro, A. A. (2017b). A large-scale study of cultural differences using urban data about eating and drinking preferences. *Information Systems*, 72(Supplement C):95–116. DOI: 10.1016/j.is.2017.10.002.

Silva, T. H. and Fox, M. S. (2024). Integrating social media data: Venues, groups and activities. *Expert Systems with Applications*, 243:122902. DOI: https://doi.org/10.1016/j.eswa.2023.122902.

Silva, T. H. and Silver, D. (2024). Using graph neural networks to predict local culture. *Environment and Planning B: Urban Analytics and City Science*. DOI: 10.1177/23998083241262053.

Silva, T. H., Vaz de Melo, P. O. S., Almeida, J. M., Salles, J., and Loureiro, A. A. F. (2013). A comparison of foursquare and instagram to the study of city dynamics and urban social behavior. In *Proc. ACM SIGKDD Int. Workshop on Urban Computing (UrbComp'13)*, Chicago, USA. DOI: 10.1145/2505821.2505836.

Silva, T. H., Viana, A. C., Benevenuto, F., Villas, L., Salles, J., Loureiro, A., and Quercia, D. (2019). Urban computing leveraging location-based social network data: A survey. *ACM Computing Surveys*, 52(1):1–39. DOI: 10.1145/3301284.

Silver, D. and Silva, T. H. (2023). Complex causal structures of neighbourhood change: Evidence from a functionalist model and yelp data. *Cities*, 133:104130. DOI: 10.1016/j.cities.2022.104130.

Skora, L. E., Senefonte, H. C., Delgado, M. R., Lüders, R., and Silva, T. H. (2022). Comparing global tourism flows measured by official census and social sensing. *Online Social Networks and Media*, 29:100204. DOI: 10.1016/j.osnem.2022.100204.

Traag, V. A., Waltman, L., and van Eck, N. J. (2019). From Louvain to Leiden: guaranteeing well-connected communities. *Scientific Reports*, 9:5233. DOI: 10.1038/s41598-019-41695-z.

Veiga, D. A. M., Frizzo, G. B., and Silva, T. H. (2019). Cross-cultural study of tourists mobility using social media. In *Proceedings of the 25th Brazillian Symposium on Multimedia and the Web*, WebMedia '19, page 313–316, New York, NY, USA. Association for Computing Machinery. DOI: 10.1145/3323503.3360620.

Wu, D. Q., Tan, J., Guo, F., Li, H., Chen, S., and Jiang, S. (2020). Multi-Scale Identification of Urban Landscape Structure Based on Two-Dimensional Wavelet Analysis: The Case of Metropolitan Beijing, China. *Ecological Complexity*, 43:100832. DOI: 10.1016/j.ecocom.2020.100832.

Ye, C., Zhang, F., Mu, L., Gao, Y., and Liu, Y. (2021). Urban function recognition by integrating social media and street-level imagery. *Environment and Planning B: Urban Analytics and City Science*, 48(6):1430–1444. DOI: 10.1177/2399808320935467.