Original article

Aggregated data, hidden realities: challenges in the analysis of obesogenic environments

Dados agregados, realidades ocultas: desafios na análise de ambientes obesogênicos

Gustavo Henrique Campos de FARIA [I]

Abstract

The article addresses challenges related to using census data in the analysis of obesogenic environments, highlighting that data aggregation can conceal spatial and social disparities. Using a computational model in the software ArcGIS®, the study investigates obesogenicity in a neighborhood in the city of Belo Horizonte and shows that geospatial techniques can reveal patterns hidden by conventional statistics. The results underline the need for more detailed approaches to promote the understanding of urban dynamics related to public health and contribute to more effective urban planning and health promotion policies.

Keywords*:* urban health; obesity; space syntax.

Resumo

O artigo explora os desafios dos dados censitários na análise de ambientes obesogênicos, destacando como a agregação de dados pode ocultar disparidades espaciais e sociais. Utilizando um modelo computacional no ArcGIS®, o estudo investiga a obesogenicidade em um bairro de Belo Horizonte, demonstrando como técnicas geoespaciais podem revelar padrões encobertos pela estatística convencional. Os resultados apontam a necessidade de abordagens mais detalhadas para compreender as dinâmicas urbanas relacionadas à saúde pública, contribuindo para políticas mais eficazes de planejamento urbano e promoção da saúde.

Palavras-chave: *saúde urbana; obesidade; sintaxe espacial.*

Introduction

The relationship between the built environment and public health has been extensively examined, with particular emphasis on the influence of urban structure on mobility, accessibility, and quality of life. Spatial configuration can either facilitate or hinder healthy choices, thereby impacting collective well-being. Within this context, research on obesogenic environments investigates urban elements that shape dietary habits and physical activity, mapping their distribution to inform public policies aimed at addressing obesity.

Obesity, a multifactorial phenomenon of increasing global relevance, results from the complex interaction between biological, social, cultural, and environmental factors (Downs et al., 2020; Egger and Swinburn, 1997; Matozinhos et al., 2022). Obesogenic environments are characterized by limited access to healthy foods, an abundant supply of processed and ultra-processed products, and the degradation of public spaces that should promote active living in cities (Faria, Souza and Mendes, 2024).

The analysis of such environments requires a detailed understanding of multiple variables, with the main source of data being the Population Census. While it provides a comprehensive overview of the socioeconomic and demographic structure of cities, its data aggregation methodology can conceal intra--urban inequalities and obscure local patterns that are essential for identifying areas of heightened vulnerability to disease. The categorization and spatialization of census data often produce representations that do not accurately reflect the analytical scale of urban dynamics, as the homogenization of large geographic areas dilutes critical contrasts, making it difficult to identify obesogenic sites.

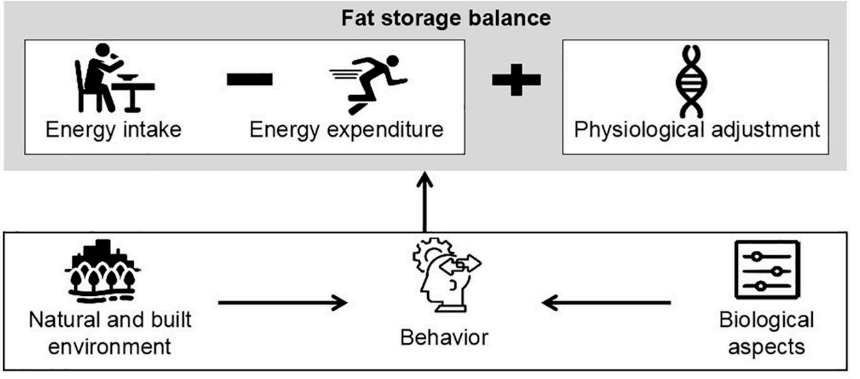
Taking this background into account, the present study investigates the influence of census data granularity on the analysis of obesogenic environments, employing normal distribution curves to demonstrate how varying levels of data aggregation affect the identification of spatial patterns. To this end, a case study was conducted in Santa Tereza neighborhood of Belo Horizonte, Minas Gerais, Brazil, integrating advanced geospatial methods and concepts from the epistemology of complexity to provide a critical analysis of urban planning and its relationship with public health.

Theoretic and conceptual framework

Understanding obesogenic environments requires a complex and interdisciplinary approach that integrates space, social behavior, and public health. The ecological model of obesity (Figure 1), proposed by Egger and Swinburn (1997), reinforces this perspective by conceptualizing obesity as a normal response to an abnormal environment in which spatial factors influence dietary habits and physical activity.

In the urban context, limited access to healthy foods, the widespread availability of processed and ultra-processed products, and the deterioration of public spaces contribute to the growing prevalence of obesity. Furthermore, the street network configuration and the spatial distribution of food establishments determine access to different products and services.

Figure 1 – Ecological model of obesity



Source: translated and edited by the author, in 2022; adapted from Egger and Swinburn (1997).

The configurational analysis of urban space has proven to be an effective tool for understanding these dynamics. Initially grounded in quantitative approaches, this perspective has evolved to incorporate both social and spatial aspects, as exemplified by Space Syntax Theory (Hillier, 2007; Hillier and Hanson, 1984), which employs topological concepts to examine the relationship between spatial organization and social phenomena (Netto, 2016). This theory relies on metrics such as connectivity, choice, and integration[1] to map and analyze patterns of movement and accessibility.

Segment analysis (Turner, 2001), urban network analysis (Sevtsuk, 2014; Sevtsuk and Mekonnen, 2012), and gravity models (Haynes and Fotheringham, 1985; Jensen-Butler, 1972) expand the precision and scope of this approach, enabling the identification of territories more susceptible to obesity by examining the relationship between spatial elements and social forms. These models treat buildings as attractors and investigate their accessibility and interaction with streets and intersections.

Constructing such models requires the acquisition of population and socioeconomic data. In Brazil, the Population Census represents one of the most relevant data sources for this purpose. Conducted by the Brazilian Institute of Geography and Statistics (IBGE), the Census aims primarily at the systematic collection of information on the country's population and territory. Despite its essential role in public policy formulation and scientific research, challenges such as budgetary constraints, methodological changes, and operational difficulties (Cano, Uchôas and Gallina, 2021) affect data quality.

Furthermore, the aggregation of information into large units (census tracts) can mask local inequalities, hindering the identification of critical spatial patterns. With advances in geoprocessing and the integration of databases such as the Multipurpose Technical

Registry (CTM), it has become possible to conduct more detailed analyses of land use and urban infrastructure. Nevertheless, data granularity remains a challenge for both researchers and public administrators. Considering this, the present research combines multiple data sources and analytical techniques – including geostatistical modeling, configurational analysis, and network analysis – to deepen the understanding of the relationship between urban morphology and obesity, while also addressing the challenges associated with the use of Census data.

Methodology

This study investigates the impact of census data aggregation on the analysis of obesogenic environments, assessing how different scales influence the identification of spatial patterns related to obesity. Using Santa Tereza neighborhood in Belo Horizonte as a case study, the research integrates geospatial methods and the epistemology of complexity to conduct a critical analysis of the relationships between urban planning and public health.

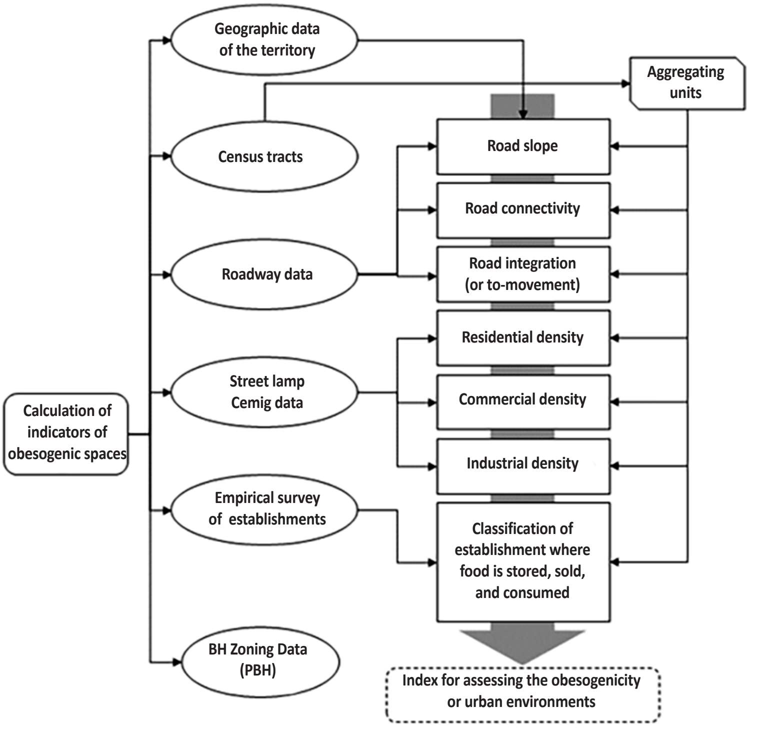
The study develops a computational model[2] within a Geographic Information System (GIS), generating an Obesogenicity Index (IO)[3] to quantify the potential of a given location to foster conditions associated with obesity. This index serves as a tool for identifying high-risk areas and improving approaches in urban planning and public health policies. Furthermore, the proposed methodology provides a foundation for future research, allowing adjustments to variables and the incorporation of new obesogenicity indicators.

Data collection and preparation

Data collection involved both primary and secondary sources. The neighborhood's street network was mapped via OpenStreetMap© (OpenStreetMap Contributors, 2021), while census data from IBGE, information from the Multipurpose Technical Registry (CTM), and data from the Minas Gerais Energy Company (Cemig) complemented the analysis. All datasets underwent cleaning and standardization to ensure interoperability and compatibility with ArcGIS®/ArcMap™ software (version 10.8).

In addition, a field audit identified and georeferenced all establishments that sell, store, or produce food – such as supermarkets, bakeries, butcher shops, and informal vendors – (Faria, Souza and Mendes, 2024). These locations were classified by size, type of consumption, and location, using tools such as Google My Maps. Figure 2 summarizes the data structuring process within the model.

Figure 2 – Methodological synthesis of spatial and geospatial statistical analysis



Source: translated and edited by the author, in 2025; based on Faria (2024).

Model development

The research was conducted in Santa Tereza neighborhood, recognized for its cultural and historical significance, as well as its urban structure characterized by narrow streets and single-family houses (features reminiscent of small-town settings). The analytical model was based on Space Syntax Theory (Hillier, Yang and Turner, 2012; Hillier and Hanson, 1997) and the Road Centre Lines (RCL) technique (Krenz, 2017, 2018), classifying urban streets according to their permeability to movement. Syntactic metrics such as connectivity, depth, choice, and integration (Castro, 2016; Saboya, 2007; Turner, 2001, 2004, 2007) were calculated using depthMapX software (version 0.8.0). To analyze pedestrian movement, impedance data (Network Dataset) were incorporated, considering factors such as street slope, average walking speed, and travel distance (Souza et al., 2016). This model was implemented in ArcGIS®/ArcMap™ software (version 10.8), reflecting the neighborhood's mobility characteristics.

Network analysis: operations

For the network analysis, the adopted methodology examines the relationship between building attraction centrality and food choices, analyzing how food availability contributes to the formation of obesogenic spaces. Establishments were categorized by size – small (+1), medium (+2), and large (+3) – serving as the basis for weighting their attractiveness factors. Food typology and consumption modality were incorporated to relate urban form, spatial distribution, and consumption patterns. To operationalize building attraction centrality, the Urban Network Analysis (UNA) tool was employed (Sevtsuk, 2010; Sevtsuk, Mekonnen and Kalvo, 2013; Sevtsuk and Mekonnen, 2012), identifying areas with a higher potential to attract pedestrians.

In addition, a probabilistic model was developed to assess the impact of different urban scenarios on obesity risk (Faria and Souza, 2024), aiming to test various ways of operationalizing obesogenicity using urban and socioeconomic variables. Inspired by the Drake's equation,[4] the model integrates variables such as population density (DHSO), average income (RM), educational attainment (EM), average age of residents (IDADE), type of food establishment (TE), spaces for physical activity (DEP), and land use patterns (syntactic measures calculated using depthMapX).

The Obesogenicity Index (IO) was initially formulated as a simple product of these variables, referred to as the simplistic scenario (IO_1). For greater accuracy, a weighted version (IO_2) was introduced, in which each factor received a specific weight (w) based on the literature. Despite time and budget constraints, an effort was made to represent the relative influences of the variables in a balanced way, distributed as follows: population density (15%), average income (25%), age (5%), educational attainment (10%), type of food establishment (30%), and presence of spaces for physical activity (15%). Further studies are needed to determine these weights with greater precision. Subsequently, the model was refined through the inclusion of reach centrality (Sevtsuk and Mekonnen, 2012), resulting in the IO_3 equation, which incorporates the spatial accessibility of the analyzed variables.

$$(IO_1) = DHSO \times RM \times IDADE \times EM \times TE \times DEP$$

$$(IO_2) = (DHSO \times w1) \times (RM \times w2) \times (IDADE \times w3) \times (EM \times w4) \times (TE \times w5) \times (DEP \times w6)$$

$$(IO_3) = (DHSO \times w1 \times reach1) \times (RM \times w2 \times reach2) \times (IDADE \times w3 \times reach3) \times (EM \times w4 \times reach4) \times (TE \times w5 \times reach5) \times (DEP \times w6 \times reach6)$$

The selection of variables composing the model was grounded in evidence from the literature on food environments (Downs et al., 2020; Gálvez Espinoza et al., 2017; Mendes, Pessoa and Costa, 2022; Swinburn, Egger and Raza, 1999). Factors such as population density, income, educational attainment, average age, food typology, and the presence of spaces for physical activity have demonstrated significant influence on population behaviors. In the IO_2 scenario, the weights assigned to each variable were defined exploratorily, based on the frequency and impact of these factors in the reviewed literature. Although not presented as absolute or definitive values, these weights provide a starting point for more refined analyses and can be adjusted according to new data or analytical contexts. Chart 1 summarizes the treatment of each variable and its method of incorporation into the model.

Chart 1 – Variable processing for incorporation into the computational model

Variable	Operation
Population Density (DHSO)	DHSO = (Population of the census area) / (Area of the census area). This provides the population density in each spatial unit.
Average Income (RM)	RM = 1 / (Average Income), inverting the value to reflect the inverse relationship with obesogenicity. Areas with higher income contribute less to obesogenicity.
Average Age of Residents (IDADE)	IDADE = ((V32 × 0) + (Σ V_n × μ_n) + (V71 × 97) + (V72 × 100)) / Σ(V_n), where V_n = population in each age group and μ_n = midpoint of each age group.
Educational Attainment (EM)	Direct integration of years of schooling data, reflecting socioeconomic status and its impact on health behaviors.
Food Establishments (TE)	Classification: Residential = 1; Commercial = Small (0–30m²) = 2; Medium (31–70m²) = 3; Large (>70m²) = 4. Applied using VBScript in Field Calculator to convert categories into numerical values.
Spaces for Physical Activity (DEP)	DEP values weighted by walking time: 0–1 min = 0.50; 2 min = 0.55; 3 min = 0.60; 4 min = 0.65; 5 min = 0.70; 6 min = 0.75; 7 min = 0.80; 8 min = 0.85; 9 min = 0.90; 10 min = 0.95; 11+ min = 1.00. Calculated with isochrone curves and automated with VBScript. This inverse relationship ensures that closer facilities have a stronger positive impact on reducing obesogenicity.
Reach (Building Reach Centrality)	Calculated using the UNA tool (Sevtsuk, Mekonnen and Kalvo, 2013; Sevtsuk and Mekonnen, 2012) with a 500 m network radius to determine the number of reachable destinations, reflecting spatial accessibility within the urban environment.

Source: the author, in 2025.

Although accessibility to services such as supermarkets, schools, or commercial centers is traditionally measured using broader time intervals – such as 5, 10, or 15 minutes – this study opted for 1-minute intervals. This choice aims to capture with greater precision the nuances of pedestrian behavior at a micro-scale, especially in heterogeneous urban contexts with significant physical and topographical barriers, as is the case of Santa Tereza neighborhood.

This methodological decision is embedded in an exploratory and experimental approach based on probabilistic scenarios, seeking to represent with greater sensitivity the complex relationships between space and health. Although it simplifies the interactions among variables – by assuming independence and multiplicative effects – the model constitutes a starting point for studies on obesogenic environments within the fields of architecture and urbanism. In addition to being replicable in other geographic contexts, this framework allows for future adjustments and refinements, such as testing different temporal scales and including new variables and relationships, thereby contributing to the theoretical and methodological advancement of the field.

The application of gaussian curves in the analysis of census data granularity

In the final methodological stage, normalized gaussian distribution curves were constructed for the obesogenicity measures (IO_1, IO_2, and IO_3) and for the building attraction centrality measure, based on simulated data reflecting their spatial distribution in Santa Tereza neighborhood. The process followed three main steps:

1) Data simulation and organization: the measurement data were organized into representative intervals, and the average values of these intervals were used for distribution modeling. All datasets were concatenated into a single DataFrame to facilitate manipulation and analysis.

2) Calculation of gaussian curves: using the norm.pdf function from the scipy.stats library, normalized probability density curves were generated for each measure. Normalization allowed the distributions to be directly compared regardless of scale differences between the indexes.

3) Visualization and comparison of measures: the gaussian curves were plotted using the matplotlib library, enabling graphical comparison of the measures. Distinct colors were assigned to each measure (blue for IO_1, yellow for IO_2, red for IO_3, and green for the gravity measure).

This approach allowed for the identification of patterns in the obesogenicity of the neighborhood, highlighting spatial variations among the measures and suggesting adjustments to the data granularity. The Gaussian curves indicated relevant trends in the distribution of the measures, providing insights into the most appropriate scale for analysis. Charts 2 and 3 present the python codes used for constructing the normal distribution graphs.

Chart 2 – Python code for plotting graphs with IO_1, IO_2, and IO_3 measures

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from scipy.stats import norm

# Dados simulados para as medidas IO1, IO2, IO3
io1_data = {
  'Intervalos': ["0,0029 - 0,0070", "0,0071 - 0,013", "0,014 - 0,017", "0,018 - 0,020", "0,021 - 0,023",
        "0,024 - 0,026", "0,027 - 0,029", "0,030 - 0,036", "0,037 - 0,059", "0,060 - 0,19"],
  'Medidas': [0.005, 0.010, 0.0155, 0.019, 0.022, 0.025, 0.028, 0.033, 0.048, 0.125]
}

io2_data = {
  'Intervalos': ["0,0029 - 0,0070", "0,0071 - 0,013", "0,014 - 0,017", "0,018 - 0,020", "0,021 - 0,023",
        "0,024 - 0,026", "0,027 - 0,029", "0,030 - 0,036", "0,037 - 0,059", "0,060 - 0,19"],
  'Medidas': [0.0045, 0.0095, 0.0145, 0.0195, 0.0215, 0.0255, 0.0285, 0.0335, 0.046, 0.128]
}

io3_data = {
  'Intervalos': ["0 - 0,005", "0,006 - 0,011", "0,012 - 0,017", "0,018 - 0,026", "0,027 - 0,035",
        "0,036 - 0,045", "0,046 - 0,06", "0,061 - 0,08", "0,081 - 0,117", "0,118 - 0,196"],
  'Medidas': [0.0025, 0.0085, 0.0145, 0.022, 0.031, 0.0405, 0.053, 0.0705, 0.099, 0.157]
}

# Convertendo para DataFrames
df_io1 = pd.DataFrame(io1_data)
df_io2 = pd.DataFrame(io2_data)
df_io3 = pd.DataFrame(io3_data)

# Gerando valores x para a curva gaussiana normalizada
x = np.linspace(-3, 3, 1000)

# Criando curvas gaussianas normalizadas
gaussian_io1 = norm.pdf(x, loc=0, scale=1) * max(df_io1['Medidas']) / max(norm.pdf(x, loc=0, scale=1))
gaussian_io2 = norm.pdf(x, loc=0, scale=1) * max(df_io2['Medidas']) / max(norm.pdf(x, loc=0, scale=1))
gaussian_io3 = norm.pdf(x, loc=0, scale=1) * max(df_io3['Medidas']) / max(norm.pdf(x, loc=0, scale=1))

# Plotando as curvas gaussianas normalizadas
plt.figure(figsize=(10, 6))

plt.plot(x, gaussian_io1, label='IO1 - Cenário Simplista', color='blue')
plt.plot(x, gaussian_io2, label='IO2 - Cenário Ponderado', color='yellow')
plt.plot(x, gaussian_io3, label='IO3 - Cenário Alcance Incorporado', color='red')

plt.title('Curvas de Distribuição Gaussiana Normalizada das Medidas IO1, IO2 e IO3')
plt.xlabel('Valores Normalizados')
plt.ylabel('Densidade de Probabilidade')
plt.legend()
plt.grid(True)
plt.tight_layout()

# Exibir gráfico
plt.show()
```

Source: the author, in 2024.

Chart 3 – Python code for plotting graphs with IO_1, IO_2, IO_3 and gravity measures

(To be continued)

```
import matplotlib.pyplot as plt
import pandas as pd
import numpy as np
from scipy.stats import norm

# Dados simulados para as medidas IO1, IO2, IO3
io1_data = {
  'Intervalos': ["0,0029 - 0,0070", "0,0071 - 0,013", "0,014 - 0,017", "0,018 - 0,020", "0,021 - 0,023", "0,024 - 0,026",
"0,027 - 0,029", "0,030 - 0,036", "0,037 - 0,059", "0,060 - 0,19"],
  'Medidas': [0.005, 0.010, 0.0155, 0.019, 0.022, 0.025, 0.028, 0.033, 0.048, 0.125]
}

io2_data = {
  'Intervalos': ["0,0029 - 0,0070", "0,0071 - 0,013", "0,014 - 0,017", "0,018 - 0,020", "0,021 - 0,023", "0,024 - 0,026",
"0,027 - 0,029", "0,030 - 0,036", "0,037 - 0,059", "0,060 - 0,19"],
  'Medidas': [0.0045, 0.0095, 0.0145, 0.0195, 0.0215, 0.0255, 0.0285, 0.0335, 0.046, 0.128]
}

io3_data = {
  'Intervalos': ["0 - 0,005", "0,006 - 0,011", "0,012 - 0,017", "0,018 - 0,026", "0,027 - 0,035", "0,036 - 0,045", "0,046 -
0,06", "0,061 - 0,08", "0,081 - 0,117", "0,118 - 0,196"],
'Medidas': [0.0025, 0.0085, 0.0145, 0.022, 0.031, 0.0405, 0.053, 0.0705, 0.099, 0.157]
}

# Dados simulados para a Medida de Gravidade
gravidade_data = {  'Intervalos': ["0,000000 - 1,824365", "1,824366 - 4,372193", "4,372194 - 8,700763", "8,700764 -
17,602568", "17,602569 - 58,694008"],
  'Medidas': [0.912, 3.098, 6.536, 13.151, 38.148]
}

# Convertendo para DataFrames
df_io1 = pd.DataFrame(io1_data)
df_io2 = pd.DataFrame(io2_data)
df_io3 = pd.DataFrame(io3_data)
df_gravidade = pd.DataFrame(gravidade_data)

# Normalizing the data for better comparison across IO1, IO2, IO3, and Gravidade
# Concatenating all data into a single DataFrame
all_data = pd.concat([df_io1['Medidas'], df_io2['Medidas'], df_io3['Medidas'], df_gravidade['Medidas']])

# Calculating the overall mean and standard deviation
mean_all = all_data.mean()
std_all = all_data.std()

# Normalizing each dataset
df_io1['Medidas_Norm'] = (df_io1['Medidas'] - mean_all) / std_all
df_io2['Medidas_Norm'] = (df_io2['Medidas'] - mean_all) / std_all
df_io3['Medidas_Norm'] = (df_io3['Medidas'] - mean_all) / std_all
df_gravidade['Medidas_Norm'] = (df_gravidade['Medidas'] - mean_all) / std_all

# Gerando valores x para a curva gaussiana normalizada
x = np.linspace(-3, 3, 1000)

# Creating normalized Gaussian curves
gaussian_io1_norm = norm.pdf(x, loc=0, scale=1) * max(df_io1['Medidas_Norm']) / max(norm.pdf(x, loc=0,
scale=1))
gaussian_io2_norm = norm.pdf(x, loc=0, scale=1) * max(df_io2['Medidas_Norm']) / max(norm.pdf(x, loc=0,
scale=1))
gaussian_io3_norm = norm.pdf(x, loc=0, scale=1) * max(df_io3['Medidas_Norm']) / max(norm.pdf(x, loc=0,
scale=1))
gaussian_gravidade_norm = norm.pdf(x, loc=0, scale=1) * max(df_gravidade['Medidas_Norm']) / max(norm.pdf(x,
loc=0, scale=1
```

Chart 3 – Python code for plotting graphs with IO_1, IO_2, IO_3 and gravity measures

(Conclusion)

```
# Plotting the normalized Gaussian curves
plt.figure(figsize=(10, 6))

plt.plot(x, gaussian_io1_norm, label='IO1 - Cenário Simplista', color='blue')
plt.plot(x, gaussian_io2_norm, label='IO2 - Cenário Ponderado', color='yellow')
plt.plot(x, gaussian_io3_norm, label='IO3 - Cenário Alcance Incorporado', color='red')
plt.plot(x, gaussian_gravidade_norm, label='Medida de Gravidade', color='green')

plt.title('Curvas de Distribuição Gaussiana Normalizada (Comparação Normalizada)')
plt.xlabel('Valores Normalizados')
plt.ylabel('Densidade de Probabilidade')
plt.legend()
plt.grid(True)
plt.tight_layout()

# Exibir gráfico
plt.show()
```

Source: the author, in 2024.

Results and discussion

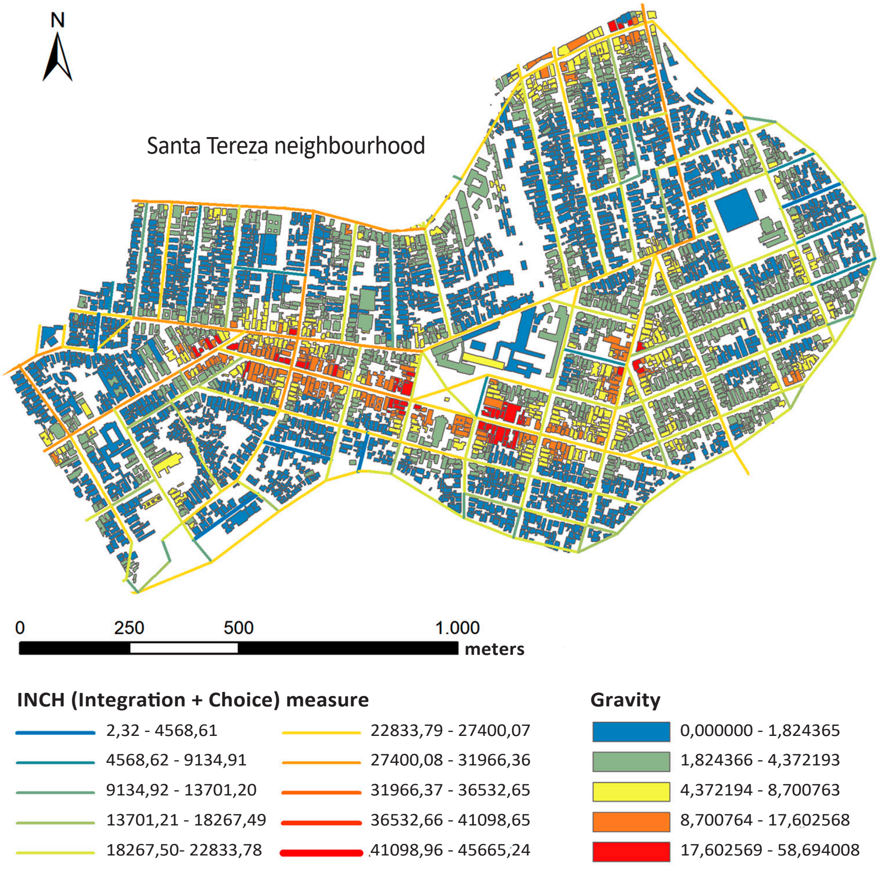
This section presents and analyzes the results of applying the computational model for measuring obesogenicity. Maps of the different operationalizations (building attraction centrality and obesogenicity scenarios – measures IO_1, IO_2, and IO_3) are displayed, revealing spatial patterns. The similarity between IO_1 and IO_2 measures is discussed through a correlation test, along with an analysis of gaussian distribution curves that aid in understanding the data structure and its relationship with the built environment. Finally, the findings are contextualized in relation to the study's objectives, highlighting methodological and urbanistic implications.

General results

The result of building attraction centrality, calculated using the UNA tool and overlaid with the combined representation of integration measure and choice ratio (INCH)[5] of streets, calculated with depthMapX software, is presented in Figure 3. This integrated visualization highlights how land use patterns articulate with the spatial configuration of the street system, influencing urban flow and accessibility potentials in different segments of the urban grid.

Areas of lower attractiveness (cold tones) are concentrated along the neighborhood's periphery, while areas of higher attraction (warm

Figure 3 – Building attraction centrality overlaid with the combined representation of integration measure + choice ratio (INCH) of the streets

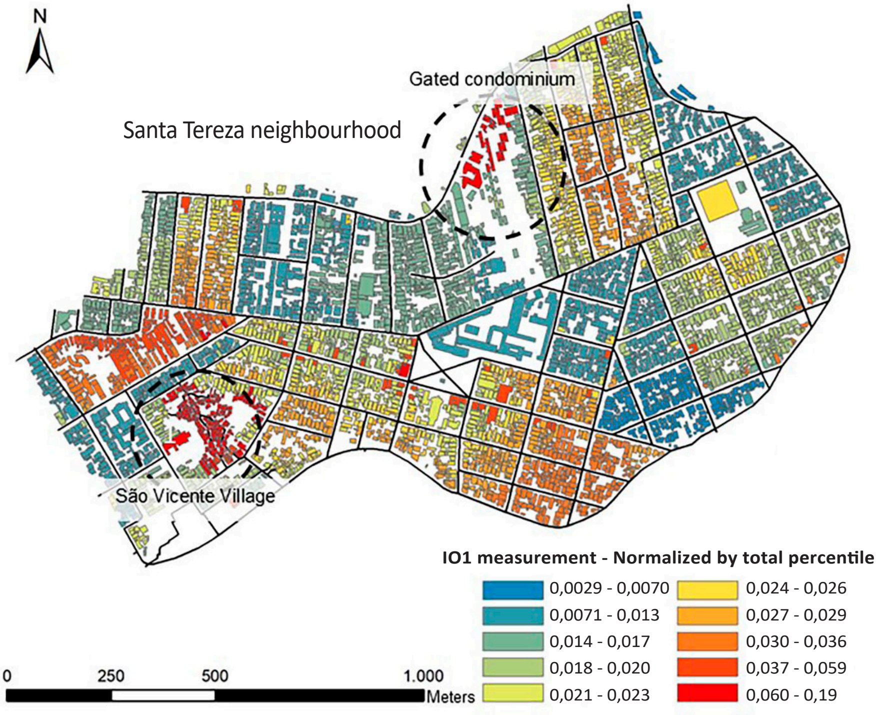


Source: the author, in 2024.

tones) are located along main traffic corridors and central regions. This pattern, aligned with the natural movement theory (Hillier et al., 1993), demonstrates that more integrated streets attract greater pedestrian flow and commercial activities. The observed distribution suggests that urban configuration impacts the concentration of food establishments – predominantly those offering ultra-processed foods – in the most accessible areas. This pattern reinforces the notion that spatial accessibility can play a significant role in shaping dietary habits, highlighting the need for public policies that encourage diversification of the food supply, with an emphasis on fresh products.

Figure 4 presents the map of the simplistic scenario (IO_1), which displays the spatial distribution of obesogenicity in Santa Tereza neighborhood, emphasizing the combined influence of factors such as population density, income, age, education, food availability, and spaces for physical activity.

Figure 4 – Map of obesogenicity measure distribution resulting from the operationalization of simplistic scenario (IO_1) in Santa Tereza neighborhood, classified using the Natural Breaks (Jenks) method

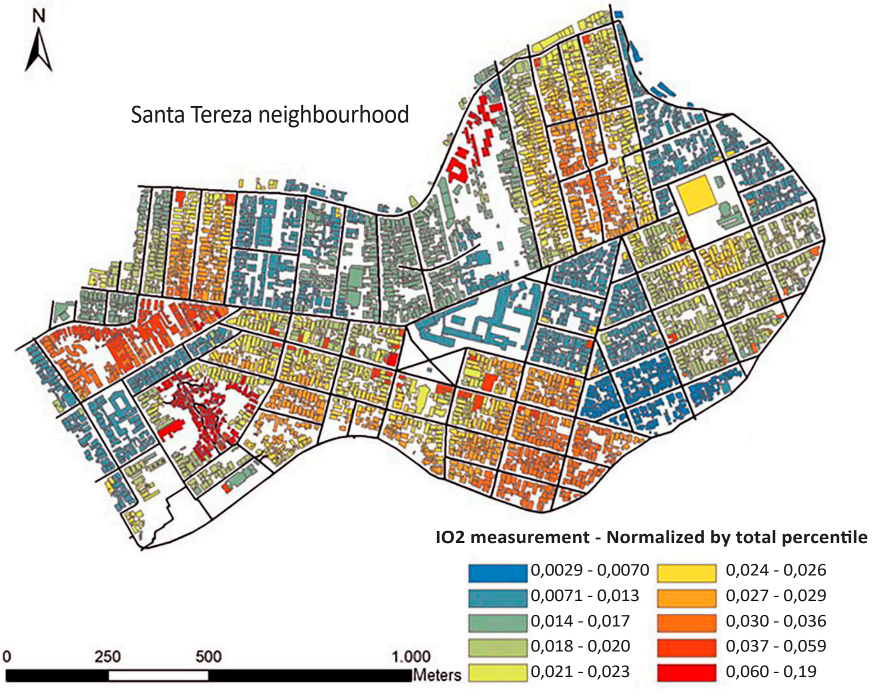


Source: the author, in 2024.

Areas depicted in red and orange, representing higher IO_1 values, are concentrated in the northern and western parts of the neighborhood, especially in São Vicente Village and around a gated condominium, where urban densification is more intense. Conversely, areas shown in blue and green, with lower IO_1 values, correspond to zones of low population density. This pattern suggests that more connected and denser areas tend to attract a greater variety of establishments, including less healthy options, contributing to the formation of obesogenic environments. The distribution of IO_1 is diffuse and gradual, indicating that obesogenicity is not confined to isolated points but is a dynamic phenomenon related to urban structure and social behavior, reinforcing the need for urban planning that considers these dynamics to mitigate their impacts.

The weighted scenario (IO_2) was operationalized by assigning weights to variables based on literature evidence. Visual analysis of the resulting map (Figure 5) reveals similarities with the IO_1 map, raising hypotheses about data normalization. A sensitivity test was conducted to investigate whether normalization – using the Natural Breaks (Jenks) method – was masking differences between the measures. After comparing normalization with the sample quantile method, no significant divergences were observed.

Figure 5 – Map of obesogenicity measure distribution resulting from the operationalization of weighted scenario (IO_2) in Santa Tereza neighborhood, classified using the Natural Breaks (Jenks) method

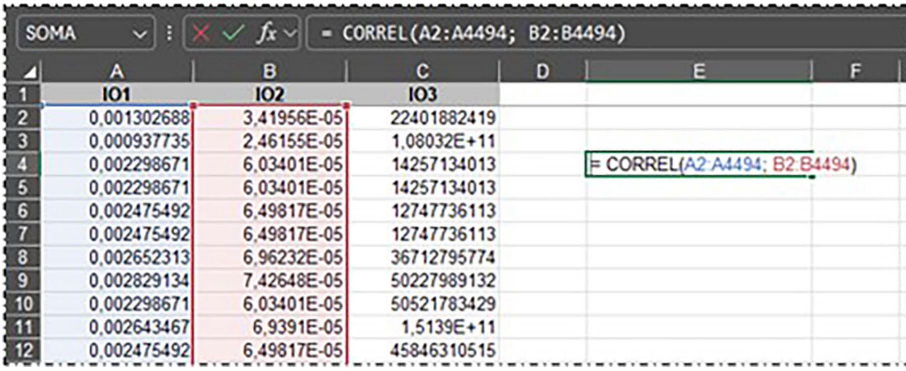


Source: the author, in 2024.

With the hypothesis regarding normalization discarded, the investigation focused on how the mathematical differences between IO_1 and IO_2 could be subtle or visually imperceptible, even with distinct weights. A correlation test between the IO_1 and IO_2 measures, conducted using the "CORREL" function in Excel® (Microsoft®, 2024) (Figure 6), revealed a perfect positive correlation (value of 1), indicating that the spatial variations of the two measures are exactly proportional. In other words, where IO_1 increases, IO_2 also increases in the same proportion, and where IO_1 decreases, IO_2 follows the same pattern.

This result suggests that the aggregation of census data, which uses large spatial units, may have "flattened" the variations in the variables. This uniformity in visual representations and correlation could reflect the aggregated data's inability to effectively distinguish small spatial variations that might exist if the data were collected or analyzed at a finer scale. Therefore, data aggregation limits the capacity for more precise and detailed analysis, concealing disparities or local patterns essential for understanding the interactions of urban phenomena, which may lead to limited conclusions.

Figure 6 – Illustration of the correlation test between IO_1 and IO_2 in Excel® software



SOMA = CORREL(A2:A4494; B2:B4494)

	A	B	C	D	E	F
1	IO1	IO2	IO3			
2	0,001302688	3,41956E-05	22401882419			
3	0,000937735	2,46155E-05	1,08032E+11			
4	0,002298671	6,03401E-05	14257134013		= CORREL(A2:A4494; B2:B4494)	
5	0,002298671	6,03401E-05	14257134013			
6	0,002475492	6,49817E-05	12747736113			
7	0,002475492	6,49817E-05	12747736113			
8	0,002652313	6,96232E-05	36712795774			
9	0,002829134	7,42648E-05	50227989132			
10	0,002298671	6,03401E-05	50521783429			
11	0,002643467	6,9391E-05	1,5139E+11			
12	0,002475492	6,49817E-05	45846310515			

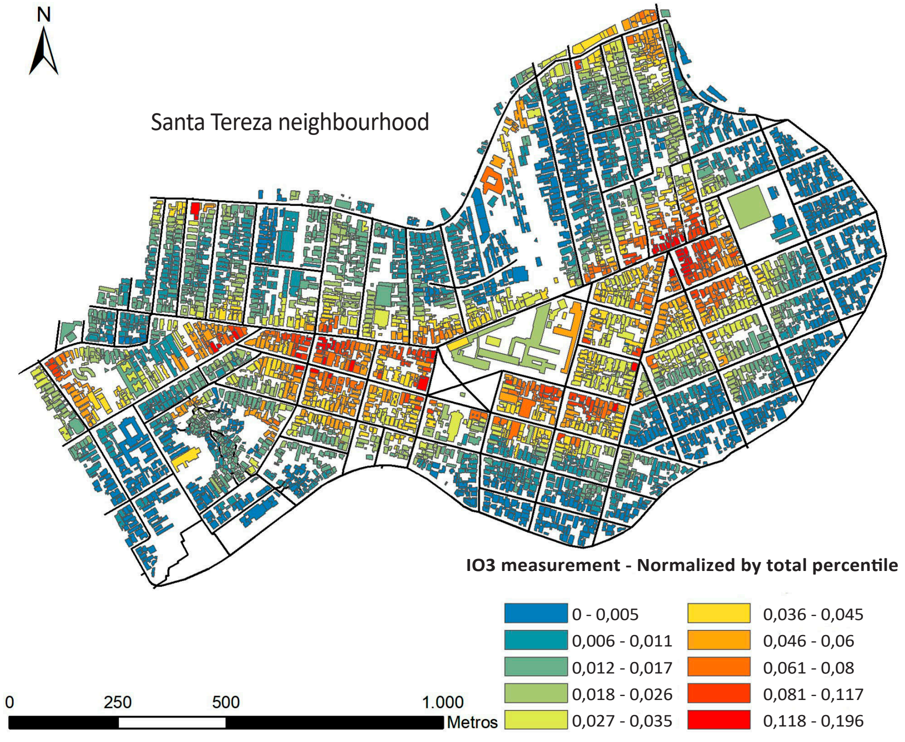
Source: screenshot, edited by the author, in 2024.

In the scenario incorporating reach centrality (IO_3), the introduction of this network analysis metric revealed a heterogeneous spatial distribution of obesogenicity (Figure 7). Areas with higher reach, indicated by warm colors, are concentrated in the central and northern regions of the Santa Tereza neighborhood, characterized by historical and cultural aspects. The concentration of high IO_3 values along the main traffic corridors of the neighborhood, with higher densities of establishments, reinforces that more integrated streets attract commercial activities, directly influencing the obesogenic environment. Furthermore, it can be stated that the main traffic corridors act as catalysts in the development of food distribution and consumption environments.

When compared to the IO_1 and IO_2 scenarios, the introduction of reach centrality in the model reveals more complex nuances, with regions that previously exhibited moderate obesogenicity values now showing higher levels. This indicates that the spatial connectivity relationship and the interaction among neighboring buildings play a significant role in shaping the obesogenic environment. Furthermore, the reach measure demonstrates how the concentration of buildings in dense and well-connected areas increases obesogenic potential, highlighting the interplay between environmental, social, and infrastructural factors in the development of obesity-prone spaces.

The data analysis in Santa Tereza neighborhood reveals how various socioeconomic and urban factors shape local obesogenicity. Areas with high population density, low income, and low educational attainment exhibit higher obesogenicity indexes (IO_1 and IO_2), whereas regions with better access to spaces for physical activity and higher educational levels show lower values. Overlaying geospatial variables with obesogenicity measures enables the identification of these patterns, emphasizing the direct influence of the built environment in forming obesogenic spaces.

Figure 7– Map of obesogenicity measure distribution resulting from the operationalization of incorporated reach centrality scenario (IO_3) in Santa Tereza neighborhood, classified using the Natural Breaks (Jenks) method



Source: the author, in 2024.

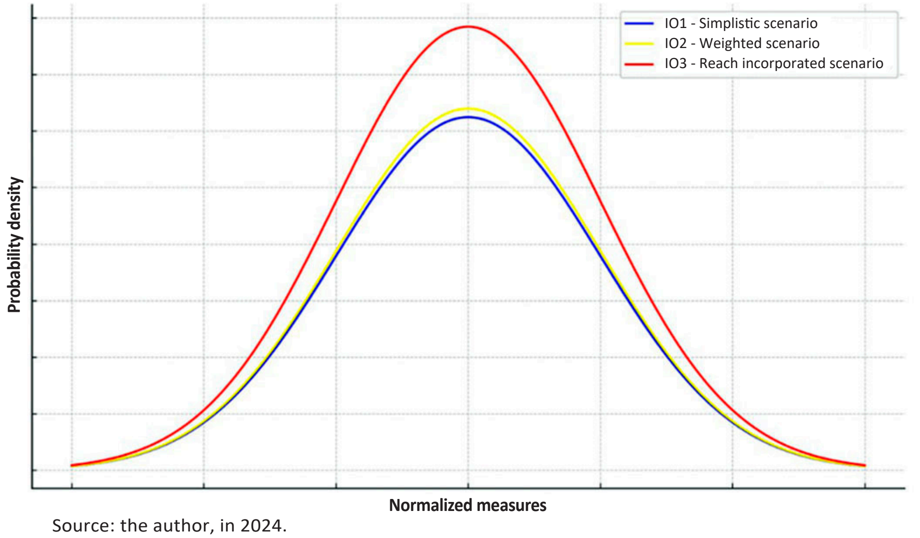
However, the use of aggregated data, such as census data, tends to smooth finer and more localized variations, making it difficult to perceive subtle spatial differences that could be relevant. The combination of data into larger units can mask important nuances, rendering results less sensitive to characteristics specific to smaller scales. In the IO_3 scenario, which accounts for accessibility to physical activity spaces, this smoothing effect becomes even more apparent, as proximity to such locations may partially offset the negative impacts of high population density and unfavorable socioeconomic conditions. This reveals the complexity of interactions among urban and social factors, suggesting that urban planning must be sensitive to these variabilities to promote healthier and more resilient environments.

Evaluating the distribution curves

The Gaussian, or normal, distribution curve is a graphical representation illustrating the distribution of continuous data, with the highest concentration of values around the mean and a decreasing probability of extreme values as they deviate from the mean (Batanero, Tauber, and Sánchez, 2001; Bittencourt and Viali, 2006). In the context of the obesogenicity measures (IO_1, IO_2, and IO_3) for Santa Tereza neighborhood, the normalized curve was used to compare the variation and concentration of the indexes calculated in each scenario, allowing observation of how different weighting approaches and spatial variables impact obesogenicity. Normalization enables the distributions to be comparable despite differences in absolute values across scenarios. Figure 8 presents normalized Gaussian distribution curves for the IO_1, IO_2, and IO_3 measures.

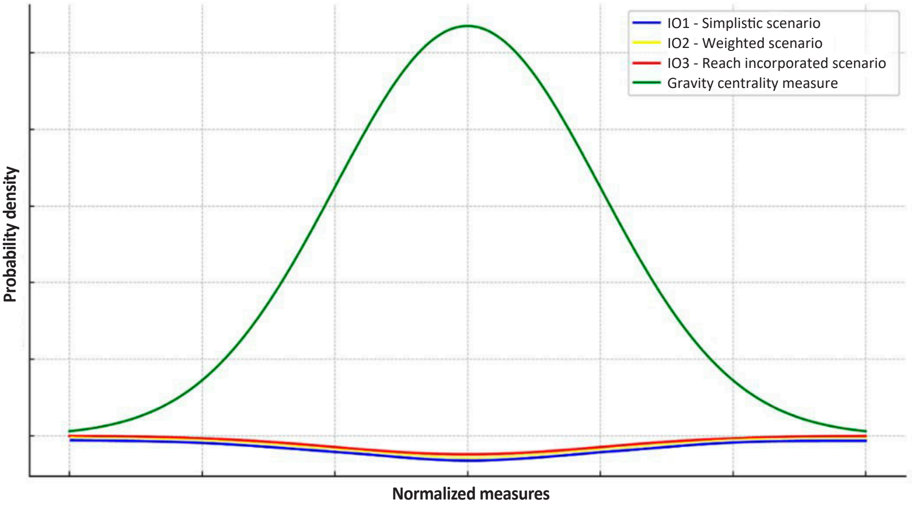
In contrast, the comparative analysis of the normalized curves for the obesogenicity measures (IO_1, IO_2, and IO_3) and the building attraction centrality measure (Figure 9) reveals fundamental differences in how these methodologies capture the spatial structure of the neighborhood. On one hand, the gravity measure curve shows a high concentration around the central value, indicating that few areas stand out as poles while the majority maintain moderate values. This characteristic suggests that the building attraction centrality methodology is effective in identifying well-defined urban center, making it especially useful for territorial planning. However, this

Figure 8 – Gaussian distribution curves of IO_1, IO_2 and IO_3 measurements



Source: the author, in 2024.

Figure 9 – Gaussian distribution curves of the IO_1, IO_2, IO_3, and gravity centrality measurements



Source: the author, in 2024.

concentration also implies lower sensitivity to more subtle spatial variations, which may limit its applicability in studies focused on diffuse factors such as obesogenicity.

On the other hand, the curves for obesogenicity measures show greater dispersion, reflecting the influence of multiple spatial and social variables. This dispersion indicates that the IO measures are more sensitive to variations in the urban environment, providing a detailed view of the distribution of obesity risk factors. Furthermore, the absence of a perfectly gaussian distribution suggests that phenomena related to obesogenicity operate in multiple metastable states, capturing particularities not revealed by the building attraction centrality methodology. Thus, while the gravity measure highlights areas of central attraction, obesogenicity indexes offer a more diffuse and complex reading of urban space, making the approaches complementary in analyzing territorial dynamics and public health.

However, the granularity of census data imposes limitations on this analysis, as the aggregation of information by census tract smooths out finer spatial variations. In the simplistic scenario (IO_1), this characteristic results in a more homogeneous distribution of the indexes, where spatial differences appear less pronounced. As the scenarios evolve, incorporating weightings (IO_2) and factors such as accessibility (IO_3), variations become more evident but remain subject to the scale of the data used. Since census tracts aggregate a heterogeneous group of individuals into a single statistical unit, there is a risk of averages masking internal inequalities within the territory. This means that spatial micro-variations, essential for understanding obesogenicity patterns, may be diluted, reducing the perception of significant contrasts between different areas. Although normalization of the curves allows for comparisons between methodologies, the inherent limitations of the census scale indicate the need for complementary approaches, such as the use of data at finer resolutions or methods that minimize aggregation effects, ensuring a more accurate analysis of the relationship between the urban environment and obesogenicity.

Final considerations

The analysis developed throughout this study revealed intrinsic limitations in the use of aggregated data, particularly regarding the preservation of social nuances and heterogeneities. The ability to synthesize large volumes of information through statistical aggregation is undoubtedly a powerful tool for policy formulation and trend analysis. However, this same process can inadvertently perpetuate structural inequalities by masking critical disparities among population subgroups and territories. The dilution of socioeconomic and territorial disparities within aggregated data is not a trivial side effect but a phenomenon that can reinforce biased narratives and perpetuate preexisting asymmetries. This effect is especially critical in urban contexts, where inequalities are dynamic and spatially segmented. Thus, the exclusion of more detailed layers of information may result in inaccurate diagnosis and ineffective policies that fail to address local specificities.

Considering this, the present study underscores the need for hybrid methodologies that combine quantitative and qualitative analyses, incorporating more granular approaches to mitigate aggregation biases. The use of microdata, spatial analysis techniques, and the inclusion of intersectional variables can provide a more comprehensive and equitable understanding of social phenomena. Therefore, improving aggregated data analysis techniques and integrating new methodologies are fundamental steps to ensure more accurate diagnosis and more effective and equitable public policies. The challenge lies in developing approaches that reconcile the need for statistical synthesis with the preservation of diversity and the specificities of the populations analyzed.

Regarding the model proposed in this research, it moves away from a deterministic view and seeks, within the complexity of the city, a more flexible and open understanding of urban dynamics. Rather than predicting future patterns or behaviors, the model reveals the complex interactions between space and society, challenging the logic of absolute predictions and providing a tool to observe and interpret the city in its richness of possibilities and uncertainties. Thus, the model not only seeks to illuminate aspects of urban reality but also positions itself as an invitation to continuous exploration, acknowledging the limitations of knowledge and embracing the inherent complexity of urban phenomena. It does not offer definitive answers but provides a deeper perspective on the relationship between space and health, suggesting an integrated approach that considers both spatial specificities and social interactions that shape urban environments and their dynamics.

[I] https://orcid.org/0000-0002-1568-4498
Universidade Federal de Minas Gerais, Escola de Arquitetura, Programa de Pós-Graduação em Arquitetura e Urbanismo. Belo Horizonte, MG/Brasil.
guscamfar@live.com

Notes

(1) To learn more about Space Syntax metrics and relations, see: (Hillier, 2008; Hillier et al., 1993, 2012; Hillier & Hanson, 1997)

(2) The model is the result of the doctoral thesis "Modelo de Espaços Obesogênicos em Ambientes Urbanos" defended in 2024 by the Graduate Program in Architecture and Urbanism (NPGAU) at the Federal University of Minas Gerais (UFMG). To learn more about the construction of the model, see Faria (2024).

(3) Whenever the work mentions the obesogenicity index or measure (OI), it will be referring to the relative index of the potential of a given location to promote behaviors and conditions that favor the development of clinical conditions of obesity.

(4) Drake's equation is used to estimate the number of extraterrestrial civilizations in the Milky Way (Ciência Todo Dia, 2021).

(5) INCH, a Space Syntax metric, combines integration and choice measures to assess accessibility and movement potential within a network. The integration measure indicates how connected an element is to the rest, while the choice ratio represents how often a route can be used as the shortest path between different points.

References

BATANERO, C.; TAUBER, L.; SÁNCHEZ, V. (2001). Significado y comprensión de la distribución normal en un curso introductorio de análisis de datos. *Quadrante*, v. 10, n. 1, pp. 59-92.

BITTENCOURT, H. R.; VIALI, L. (2006). Contribuições para o ensino da distribuição normal ou curva de Gauss em cursos de Graduação. In: III SEMINÁRIO INTERNACIONAL DE PESQUISA EM EDUCAÇÃO MATEMÁTICA.

CANO, E. C.; UCHÔAS, A. A.; GALLINA, F. (2021). Os desafios do IBGE e a realização do Censo Demográfico 2022. *Geofronter*, n. 7. DOI: https://doi.org/10.61389/geofronter.v7.6553.

CASTRO, A. (2016). *Sintaxe espacial e a análise angular de segmentos, parte 3: calculando medidas sintáticas*. Disponível em: https://aredeurbana.com.br/2016/06/26/sintaxe-espacial-e-a-analise-angular-de-segmentos-parte-3-calculando-medidas-sintaticas/. Acesso em: 10 ago 2021.

CIÊNCIA TODO DIA (2021). A Equação de Drake: existe vida fora da Terra? *YouTube*. Disponível em: https://www.youtube.com/watch?v=wbqlz0znhrq. Acesso em: 10 jul 2022.

DOWNS, S. M.; AHMED, S.; FANZO, J.; HERFORTH, A. (2020). Food environment typology: advancing an environments toward sustainable diets. *Foods*, v. 9, n. 4, p. 532. DOI: https://doi.org/10.3390/foods9040532.

EGGER, G.; SWINBURN, B. (1997). An "ecological" approach to the obesity pandemic. *BMJ*, v. 315, n. 7106, pp. 477-480. DOI: https://doi.org/10.1136/bmj.315.7106.477.

FARIA, G. H. C. de; SOUZA, R. C. F. de (2024). Desenvolvimento de um índice para estimar a distribuição de espaços obesogênicos em ambientes urbanos: uma abordagem inspirada na Equação de Drake. *Anais.* Seminário Internacional PROJETAR: Projetar para quem? Projetar para quê? Projetar como? DOI: https://doi.org/10.29327/projetar2023.657636.

FARIA, G. H. C. de; SOUZA, R. C. F. de; MENDES, L. L. (2024). Modelos configuracionais para a avaliação da obesogenicidade em entornos urbanos. *Gestão & Tecnologia de Projetos*, v. 19, n. 1, pp. 5-28. DOI: https://doi.org/10.11606/gtp.v19i1.209856.

FARIA, G. H. C. de F. (2024). *Modelo de espaços obesogênicos em ambientes urbanos*. Tese de doutorado. Belo Horizonte, Universidade Federal de Minas Gerais.

GÁLVEZ ESPINOZA, P.; EGAÑA, D.; MASFERRER, D.; CERDA, R. (2017). Propuesta de un modelo conceptual para el estudio de los ambientes alimentarios en Chile. *Revista Panamericana de Salud Pública*, pp. 1-9. DOI: https://doi.org/10.26633/rpsp.2017.169.

HAYNES, K. E.; FOTHERINGHAM, A. S. (1985). *Gravity and Spatial Interaction Models* (Web Book). WVU Research Repository.

HILLIER, B. (2007). *Space is the machine: a configurational theory of architecture* (Edição eletrônica). UCL Space Syntax.

______ (2008). *Using DepthMap for urban analysis: a simple guide on what to do once you have an analysable map in the system*. The Barttlet School of Graduate Studies.

HILLIER, B.; HANSON, J. (1984). *The social logic of space*. University Press. https://doi.org/10.1017/CBO9780511597237.

______ (1997). The reasoning art: or, the need for an analytical theory of architecture. 1st International Space Syntax Symposium, 01.1-01.5.

HILLIER, B.; PENN, A.; HANSON, J.; GRAJEWSKI, T.; XU, J. (1993). Natural movement: or, configuration and attraction in urban pedestrian movement. *Environment and Planning B: Planning and Design*, v. 20, n. 1, pp. 29-66. DOI: https://doi.org/10.1068/b200029.

HILLIER, B.; YANG, T.; TURNER, A. (2012). Normalising least angle choice in Depthmap and how it opens up new perspectives on the global and local analysis of city space. *The Journal of Space Syntax,* v. 3, n. 2, pp. 155-193.

JENSEN-BUTLER, C. (1972). Gravity models as planning tools: a review of theoretical and operational problems. *Geografiska Annaler*. Series B, Human Geography, v. 54, n. 1, p. 68. DOI: https://doi.org/10.2307/490586.

KRENZ, K. (2017). Employing volunteered geographic information in space syntax analysis. 11th International Space Syntax Symposium, v. 150, n. 26, pp. 1-150.

______ (2018). *Network centralities in polycentric urban regions: methods for the measurement of spatial metrics*. Tese de doutorado. Bartlett Faculty of the Built Environment.

MATOZINHOS, F. P.; GOMES, C. S.; GARDONE, D. S.; CARMO, A. S. do; SANTOS, L. C. dos; MENDES, L. L.; PESSOA, M. C. (2022). "Ambiente obesogênico". In: MENDES, L. L.; PESSOA, M. C.; COSTA, B. V. de L. (org.) *Ambiente alimentar: saúde e nutrição*. Rio de Janeiro, Rubio, pp. 135-146.

MICROSOFT® (2024). Microsoft Excel (Versão 2407 Build 16.0.17830.20166).

NETTO, V. (2016). 'What is space syntax not?' Reflections on space syntax as sociospatial theory. *Urban Design International*, v. 21, n. 1, pp. 25-40. DOI: https://doi.org/10.1057/udi.2015.21.

OPEN STREET MAP CONTRIBUTORS (2021). Open Street Map database [PostgreSQL via API] (Available under the Open Database Licence from: openstreetmap.org). OpenStreetMap Foundation. Disponível em: https://planet.openstreetmap.org. Acesso em: 20 jul 2022.

SABOYA, R. (2007). *Sintaxe espacial.* Disponível em: https://urbanidades.arq.br/2007/09/03/sintaxe-espacial/. Acesso em: 10 ago 2021.

SEVTSUK, A. (2010). *Path and Place: A Study of Urban Geometry and Retail Activity in Cambridge and Somerville, MA.* Tese de doutorado. Massachusetts, Institute of Technology.

______ (2014). "Analysis and Planning of Urban Networks". In: Encyclopedia of Social Network Analysis and Mining. Nova York, Springer, pp. 25-37. DOI: https://doi.org/10.1007/978-1-4614-6170-8_43.

SEVTSUK, A.; MEKONNEN, M. (2012). Urban network analysis. A new toolbox for ArcGIS. *Revue internationale de géomatique,* v. 22, n. 2, pp. 287-305. DOI: https://doi.org/10.3166/rig.22.287-305.

SEVTSUK, A.; MEKONNEN, M.; KALVO, R. (2013). Urban Network analysis: toolbox for ArcGIS 10 / 10.1 / 10.2.

SOUZA, R. C. F. de; OLIVEIRA, V. B. de; PEREIRA, D. B.; COSTA, H. S. de M.; CAIAFFA, W. T. (2016). Viver próximo à saúde em Belo Horizonte. *Cadernos Metrópole.* São Paulo, v. 18, n. 36, pp. 326-344. DOI: https://doi.org/10.1590/2236-9996.2016-3601.

SWINBURN, B.; EGGER, G.; RAZA, F. (1999). Dissecting obesogenic environments: The development and application of a framework for identifying and prioritizing environmental interventions for obesity. *Preventive Medicine*, v. 29, n. 6 I, pp. 563-570. DOI: https://doi.org/10.1006/pmed.1999.0585.

TURNER, A. (2001). Depthmap: A Program to Perform Visibility Graph Analysis. In: 3rd INTERNATIONAL SYMPOSIUM ON SPACE SYNTAX, v. 30, n. 11, pp. 1-30,.

______ (2004). *Depthmap 4: A Researcher's Handbook*. Bartlett School of Graduate Studies.

______ (2007). From Axial to road-centre lines: a new representation for space syntax and a new model of route choice for transport network analysis. *Environment and Planning B: Planning and Design*, v. 34, n. 3, pp. 539-555. DOI: https://doi.org/10.1068/b32067

Authorship contribution

Gustavo Henrique Campos de Faria: formal analysis; conceptualization; investigation; methodology; writing; visualization.

Data Availability Statement

The dataset supporting the results of this study is not publicly available.

Editors: Lucia Bógus and Luiz César de Queiroz Ribeiro
Organizers of this issue: Suzana Pasternak and Luís Felipe Aires Magalhães

Translation: this article was translated from Portuguese to English by the author himself.

Received: March 28, 2025
Approved: July 30, 2025