

# Metodologia de caracterização dos padrões de mobilidade dos usuários de transporte público por ônibus através de dados de bilhetagem eletrônica e análise espacial

*Methodology for characterizing the mobility patterns of users of public bus systems through smart card data and spatial analysis*

Renato Goersch Andrade Parente<sup>1</sup>, João Lucas Albuquerque Oliveira<sup>2</sup>, Ivana Maria Feitosa Silva<sup>3</sup>, Francisco Moraes de Oliveira Neto<sup>4</sup>

<sup>1</sup>Universidade Federal do Ceará, Ceará – Brasil, renatogap@hotmail.com

<sup>2</sup>Universidade Federal do Ceará, Ceará – Brasil, joao@det.ufc.br

<sup>3</sup>Universidade Federal do Ceará, Ceará – Brasil, ivsfeitosa@gmail.com

<sup>4</sup>Universidade Federal do Ceará, Ceará – Brasil, moraes@det.ufc.br

## Recebido:

26 de fevereiro de 2022

## Aceito para publicação:

30 de agosto de 2022

## Publicado:

13 de dezembro de 2022

## Editor de área:

Cira Pitombo

## Palavras-chave:

Padrão de mobilidade.  
Bilhetagem Eletrônica.  
Caracterização da Mobilidade.  
Transporte Público.

## Keywords:

Mobility Patterns.  
Smart Card.  
Mobility Characterization.  
Public Transport.

DOI:10.14295/transportes.v30i3.2749

## RESUMO

O sistema de transporte público no Brasil vem perdendo demanda nas últimas décadas. Devido ao importante papel desse sistema em promover a equidade no acesso e a sustentabilidade do sistema de transportes, torna-se essencial conhecer os padrões de mobilidade dos usuários e como estes padrões variam espacialmente e temporalmente. Este artigo propõe um método para caracterizar os padrões de mobilidade no sistema de transporte público através do uso de dados de bilhetagem eletrônica (*Smart Card*) e análise espacial. O método contribui para análise da variação da demanda para diferentes padrões de mobilidade, em cidades brasileiras onde uma parcela considerável da população depende do transporte público nos seus deslocamentos diários. Assim, o método teve como base hipóteses sobre os tipos de padrões, a variabilidade espacial e temporal dos padrões, e de como eles se relacionam espacialmente com a variação da demanda. A aplicação do método para os dados de 2014 e 2018 do sistema de transporte público por ônibus de Fortaleza, mostrou principalmente que os usuários do padrão de uso regular, que moram em zonas periféricas com baixos níveis de acessibilidade, são os que mais vêm deixando o sistema ao longo dos anos.

## ABSTRACT

Considering the importance of the public transport system in promoting equity in access and sustainable transport, it is essential to know how the mobility patterns of users vary spatially and temporally. This paper proposes a method to assess the mobility patterns in the public transportation system using smart card data and spatial analysis. The method contributes to understand the relationship between the mobility patterns and the demand variation in Brazilian cities, where a major part of their population depends on public transportation for daily commuting. Thus, the method was based on hypotheses about the types of patterns, their spatial and temporal variability, and how they relate spatially to the public transport demand. The application of the method to the 2014 and 2018 data of Fortaleza's public bus system showed a great decrease in the number of users who live in peripheral areas, with the lowest levels of accessibility, and who use the system regularly to commute.



## 1. INTRODUÇÃO

Atualmente, o cenário nas cidades brasileiras é de queda expressiva na quantidade de usuários que utilizam o transporte coletivo, principalmente em grandes centros urbanos. Segundo a Associação Nacional das Empresas de Transportes Urbanos (NTU, 2018), em 9 grandes capitais brasileiras, incluindo Fortaleza, entre 1997 e 2017, houve redução acumulada de aproximadamente 35,6% de usuários do transporte público por ônibus. Esta redução da mobilidade pode indicar problemas de exclusão social da população que depende do transporte coletivo, já que a mobilidade pode ser vista como indicador de participação em atividades (Martens, 2019). Este fenômeno pode gerar também externalidades ambientais (Garcia, 2018), caso venha ocorrendo uma migração da demanda para modos menos sustentáveis, como o automóvel particular. Nesse contexto, é necessário tornar o transporte coletivo mais acessível, mais eficiente e de maior qualidade (Gomide, 2003). Para isso, a fim de melhorar os níveis de mobilidade e acessibilidade dos usuários, é imprescindível que os planejadores conheçam os padrões de mobilidade dos usuários do transporte público e os fatores determinantes desses padrões (Kieu *et al.*, 2015; Mahrsi *et al.*, 2014).

A compreensão dos padrões de mobilidade é feita tradicionalmente a partir de pesquisas domiciliares. Estas coletas manuais de dados requerem um elevado recurso de tempo e custo para serem realizadas. Em geral não são feitas com frequência e resultam em pequenas amostras da população (Pelletier *et al.*, 2011), o que não permite captar alterações nas percepções e preferências dos usuários ao longo do tempo devido a mudanças nos sistemas de transportes (Devilleine *et al.*, 2013). Devido a esta limitação de amostragem, as análises que utilizam dados de pesquisas domiciliares tendem a caracterizar a mobilidade dos usuários através de indicadores agregados por zona, assumindo que as zonas são homogêneas considerando as características socioeconômicas dos usuários e por conseguinte de suas viagens. Esta premissa contradiz o fenômeno de heterogeneidade da demanda apontado por alguns autores. Hawkins e Habib (2020) sugerem que seja evitada a atribuição de um único valor de mobilidade à população heterogênea de uma grande região. Wang *et al.* (2019) aconselham que os planos precisam ser capazes de atender às diferentes preferências de todos os grupos de usuários. Portanto, as pesquisas domiciliares não permitem captar a variabilidade para um mesmo indivíduo e heterogeneidade nos padrões de mobilidade entre os indivíduos.

Muitas cidades no mundo vêm adotando sistemas de bilhetagem eletrônica (*Smart Card*) e de localização automática de veículos (*Global Positioning System* - GPS), assim como a representação da rede e programação do sistema a partir de *Especificação Geral de Feeds de Transporte Público* (GTFS). Estes sistemas fornecem várias informações indiretas, quase instantâneas, altamente desagregadas, quase populacionais dos usuários de transporte público, a custo marginal significativamente menor em relação àqueles métodos mais tradicionais e ainda oferecem a oportunidades de análises mais dinâmicas do comportamento dos usuários, o que pode auxiliar nesse contexto de queda expressiva de demanda.

Vários estudos na literatura têm sido desenvolvidos com intuito de investigar padrões de mobilidade a partir de dados de *Smart Card* (Briand *et al.*, 2016; Ma *et al.*, 2013; Pelletier *et al.*, 2011). Em geral, estes estudos visam identificar as diferentes formas de uso do sistema para caracterizar os usuários em diferentes níveis de similaridade tanto no tempo (p. ex., uso frequente em determinados horários) (Agard *et al.*, 2006; Cats e Ferranti, 2022; Morency *et al.*, 2006) como no espaço (p. ex., tendência de utilizar rotas similares e os mesmos pontos de paradas) (Faroqi *et al.*, 2018; Faroqi *et al.*, 2019). Os estudos também buscam interpretar os

padrões conforme algumas características dos usuários presentes nos dados, como o tipo de cartão utilizado pelo usuário - gratuidade, vale transporte, estudante, entre outros (Agard *et al.*, 2006; Morency *et al.*, 2006). Os padrões encontrados também têm sido explicados a partir de fatores socioeconômicas (p. ex., renda média) e do tipo de uso do solo no desembarque dos usuários (Pieroni, 2018; Pieroni *et al.*, 2021; Zhang *et al.*, 2021). Em geral, os estudos aplicam técnicas de clusterização para gerar os grupos de usuários, como o *k-means* (Cats e Ferranti, 2022; Morency *et al.*, 2006; Zhang *et al.*, 2021).

Em sua essência, os métodos buscam distinguir entre padrões que sejam regulares ou típicos de uso do sistema, ou padrões que sejam esporádicos ou atípicos (Kieu *et al.*, 2015; Ma *et al.*, 2013). Com relação a regularidade, alguns autores defendem que a dependência temporal nos padrões de mobilidade (e.g., hora da viagem, modo de viagem, trajeto, destino, etc) devido a sequência de atividades deve ser considerada (Goulet-Langlois *et al.*, 2018; He *et al.*, 2020; He *et al.*, 2021; Viallard *et al.* 2019). Assim, considera-se dados longitudinais sobre os atributos das viagens para considerar a dependência nos padrões de viagens.

Como visto na literatura, o principal intuito de investigar padrões de mobilidade seria identificar a heterogeneidade da demanda em relação à utilização do sistema de transporte público. Esta heterogeneidade pode se referir às diferenças de intensidade ou de regularidade de uso do sistema, que podem variar entre grupos e espacialmente. A intensidade diz respeito à frequência de uso e a regularidade à utilização de forma periódica do sistema.

A maioria dos estudos desenvolvidos têm como foco somente a identificação de padrões sem relacionar com algum tipo de fenômeno no meio urbano. São incipientes os estudos que utilizam dados de *Smart Card* para compreender os comportamentos de viagens em cidades onde uma parcela considerável da população depende do transporte público nos seus deslocamentos diários (Pieroni *et al.*, 2021). Uma exceção é o estudo de Pieroni *et al.* (2021) que estudou o padrão de viagens de populações residentes em assentamentos precários em São Paulo. Especialmente, os dados de *Smart Card* ainda não foram usados para analisar a evolução dos padrões com foco no fenômeno da queda da demanda do sistema de transporte público, que vem ocorrendo em cidades brasileiras. Acredita-se ser importante identificar quem são usuários (padrões de deslocamento) que podem estar ingressando ou deixando o sistema ao longo do tempo, assim como identificar onde estes usuários estão localizados no espaço (endereço ou zona de origem da viagem). Existem também algumas lacunas metodológicas quanto à definição de grupos conforme o fenômeno em análise com consequente dificuldade de interpretação dos grupos gerados.

Neste estudo, definiu-se a queda na demanda de transporte público por ônibus em Fortaleza como fenômeno a ser relacionado com os padrões de mobilidade. Neste fenômeno de interesse, acredita-se que características socioeconômicas e demográficas da população, assim como o nível de serviço da oferta do sistema, impactam diretamente na demanda do transporte público (Diab *et al.*, 2020). Além desses fatores, um determinante importante da mobilidade que vem sendo objeto de vários estudos é a acessibilidade, seja local (acesso e difusão do sistema) (Bree *et al.*, 2020) ou territorial (acesso às oportunidades de atividades no meio urbano) (Diab *et al.*, 2020). Embora não se tenha um consenso ou representação teórica de como estes determinantes afetam a demanda por transporte público, suspeita-se que as variações (sejam espaciais ou temporais) nestes fatores implicam em diferentes padrões de mobilidade. Souza (2019) indica, por exemplo, que o fenômeno da segregação espacial da população de menor poder aquisitivo em Fortaleza, que se concentra na periferia distante das áreas centrais com maior concentração

de atividades, está relacionado a variação espacial nos níveis de acessibilidade para este grupo. Assim, espera-se que diferentes padrões de mobilidade ocorram no meio urbano, e estes padrões podem estar relacionados a diversos fatores que expliquem a queda na demanda do transporte público nos últimos anos, como a acessibilidade, as restrições de renda dos usuários, a informalidade, dentre outros.

Suspeita-se que os usuários em Fortaleza com diferentes padrões de mobilidade estão localizados no meio urbano de forma distinta, considerando que os diferentes determinantes da mobilidade variam espacialmente. Além disso, tem-se a hipótese de que a queda da demanda ao longo tempo é distinta conforme os padrões de mobilidade. Estas hipóteses consideram a premissa de que, conforme visto em outros estudos, as diferenças espaciais de acessibilidade às oportunidades pelo transporte público e as restrições individuais (p.ex., renda, acesso a veículo próprio, etc.) podem variar espacialmente e temporalmente.

Este trabalho tem como objetivo propor um método de caracterização da mobilidade dos usuários do transporte público urbano, a partir da identificação de diferentes padrões de mobilidade que podem ocorrer no meio urbano usando dados de *Smart Card*. O trabalho contribui principalmente no método de identificação dos padrões em relação a dois aspectos. O primeiro é na definição e análise de atributos que permita identificar a heterogeneidade da mobilidade dos usuários do sistema de transporte público por ônibus. Já o segundo é na interpretação dos padrões por meio de análises espaciais e temporais, tendo como base o fenômeno da queda da demanda. O método foi aplicado para os dados de *Smart Card* de Fortaleza, para os anos de 2014 e 2018, permitindo analisar a relação entre os padrões encontrados e a queda na demanda que vem ocorrendo nos últimos anos em Fortaleza.

## 2. CONTEXTUALIZAÇÃO E TRATAMENTO DOS DADOS

O objeto de estudo é o Sistema Integrado de Transporte de Fortaleza (SIT-FOR). É um sistema do tipo tronco-alimentador, composto por 269 linhas de ônibus (troncais e alimentadoras), 19 linhas complementares e 7 terminais de integração física. Em 2013, foi implantado o bilhete único (*Smart Card*) permitindo aos usuários realizarem integração tarifária fora dos terminais físicos de integração. O SIT-FOR conta atualmente com sistemas de GPS da frota veicular e toda programação da sua operação é disponibilizada por meio de arquivos do tipo GTFS

Os dados utilizados neste estudo são relativos aos anos de 2014 e 2018 e foram disponibilizados pela Prefeitura Municipal de Fortaleza. Três bases de dados foram utilizadas: Base do Bilhetagem Eletrônica, contendo os registros de transações ou validações em cada dia por tipo de cartão; Base de Cadastro, contendo os registros de cadastro dos usuários do bilhete único; e a Base de GPS da frota veicular, que contém os registros de localização dos veículos da frota a cada 30 segundos. Todas as análises foram feitas para períodos de dias típicos do ano, excluindo finais de semana e feriados. O ano de 2014 foi escolhido por ser o primeiro ano com dados completos de bilhetagem eletrônica após implantação do bilhete único em meados de 2013. No período de 2014 a 2018 ocorreram várias intervenções de transportes com intenção de melhorar a qualidade do serviço, tornando este período de 5 anos de interesse para a investigação da variação da demanda.

Para este estudo, os dados de *Smart Card* do SIT-FOR foram agrupados em 4 tipologias de validação: Estudante, Vale Transporte, Inteira e Gratuidade. Define-se validação como a transação ou pagamento da tarifa por meio do cartão ou dinheiro. O Vale Transporte corresponde ao benefício de transporte concedido ao trabalhador.

A categoria Vale Transporte também inclui o Bilhete Único Avulso que é o cartão que pode ser adquirido por qualquer indivíduo da população que deseje usar o SIT-FOR. A Inteira corresponde as validações pagas em dinheiro, ou seja, sem a utilização do cartão. Dessas categorias de cartões, os usuários de Inteira e com Bilhete Único Avulso pagam a tarifa cheia. Todas as consolidações de dados, bem como as análises, foram produzidas utilizando a linguagem *R* de programação, o software *QGIS* e o *GeoDa*.

Uma etapa importante do método envolve as análises de variação espacial dos padrões de mobilidade (Seção 3.3). Para estas análises, foram usadas as localizações geográficas das residências dos usuários que geraram validações no sistema, com base nos registros de endereços dos usuários registrados na Base de Cadastro do bilhete único. Considerando os registros de cadastro mais recentes em cada ano, 2014 e 2018, e com endereços válidos (com nome do logradouro, número e bairro), a localização geográfica das residências foi encontrada utilizando a ferramenta *Geocoding API* do *Google Maps*, a qual foi escolhida por apresentar um bom desempenho mesmo quando os endereços estão escritos incorretos. A principal desvantagem desta ferramenta é que ela não é ofertada gratuitamente pela *Google*. Alternativas de georreferenciamento podem ser pensadas a partir de bases atualizadas dos logradouros da cidade, mas não foram exploradas neste trabalho, pois o *API do Google* já permite demonstrar a aplicabilidade do método.

A Tabela 1 apresenta os totais de usuários e de validações no SIT-FOR para os anos em análise. Os valores são apresentados somente para os usuários que usam o sistema. Além disso, a Tabela 1 também inclui os totais de validações por Inteira (para usuários que não possuem cartão). Note que os dados da Base de Cadastro estão incompletos, pois não incluem os usuários de Gratuidade, e nem todos os usuários que usam o sistema (Base de Bilhetagem) estão no cadastro (Base de Cadastro) fornecido pela prefeitura de Fortaleza.

**Tabela 1** – Totais de usuários e validações no sistema SIT-FOR em 2014 e 2018

Tipo de cartão	Número de usuários da Bilhetagem	Número de validações	Número de usuários do Cadastro	Número de usuários georreferenciados
<b>2014</b>				
Estudante	328.714	42.306.328	242.393	185.540
Gratuidade	76.217	6.836.767	0	0
Vale Transporte	606.901	119.143.976	197.743	197.177
Inteira	-	96.786.593	0	0
<b>Totais</b>	<b>1.011.832</b>	<b>265.073.664</b>	<b>440.136</b>	<b>382.717</b>
<b>2018</b>				
Estudante	313.640	45.901.822	312.470	251.090
Gratuidade	93.409	15.426.512	0	0
Vale Transporte	466.504	94.084.944	180.275	179.926
Inteira	-	65.992.508	0	0
<b>Totais</b>	<b>873.553</b>	<b>221.405.786</b>	<b>492.745</b>	<b>431.016</b>

Como pode ser observado na Tabela 1, foram georreferenciados (utilizando a ferramenta de *Geocoding API* do *Google Maps*) aproximadamente 88% dos endereços dos usuários do cadastro (sendo todos os endereços válidos localizados), em ambos os anos de análise. Vale ressaltar que somente o código do cartão e endereço residencial foram utilizados da Base de Cadastro, para garantir a confidencialidade dos dados.

Para as análises de variação da demanda por zona foi necessário estimar os endereços ou zonas geográficas das residências dos usuários que usam o sistema, mas que não possuem

registro no cadastro (Tabela 1). Isto foi feito com base nos endereços georreferenciados. Assim, foi necessário verificar se a distribuição espacial dos locais de validação dos usuários do cadastro se assemelha ao padrão espacial das validações da base de bilhetagem eletrônica (Seção 3.3). Para esta análise, os locais de validação foram georreferenciados, combinando os dados de GPS da frota e do *Smart Card*, conforme o método de Braga (2019). Vale ressaltar que no SITFOR não se pode assumir que o local da primeira validação do dia é próximo ao endereço de residência do usuário (Figura 1), pois o sistema de bilhetagem eletrônica é do tipo aberto, ou seja, o usuário pode validar em qualquer ponto do itinerário. Além disso, existe um espaço de acomodação dentro dos veículos entre a porta de embarque e o local de validação, sendo provável, portanto, que vários usuários não validem no momento de embarque, principalmente em horários de maior volume de passageiros.

Um importante requisito do método que será descrito na Seção 3 é que os mesmos usuários sejam rastreados ao longo de cada ano. Como este rastreamento foi feito com base no código do cartão (já que existem vários usuários da bilhetagem sem cadastro) e não no nome ou identidade do usuário, é necessário garantir que não houve mudança de código do cartão ao longo de cada ano. Para amostra de usuários do cadastro, foi observado que a possibilidade de um usuário mudar o código do cartão ao longo de cada ano é mínima. Para 2014 a proporção de usuários do cadastro que mudaram o código do cartão foi de 6,3% e em 2018 foi de 4,4%. Considera-se que estes resultados viabilizam as análises dos padrões espaciais de mobilidade em cada ano, e a comparação destes padrões entre os dois anos de análise (Seção 3.3).

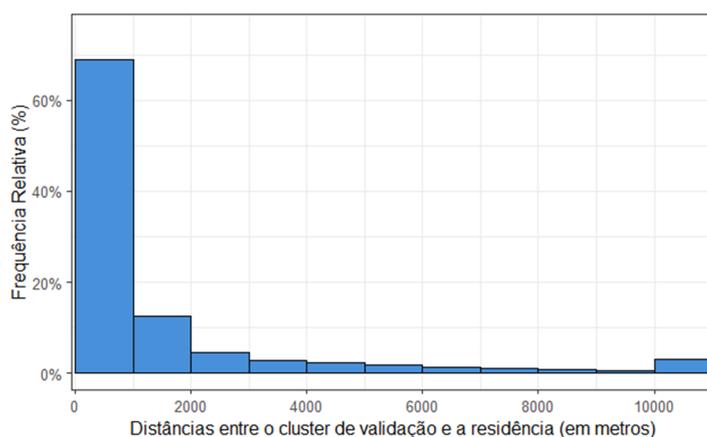


Figura 1. Distâncias do *cluster* da primeira validação do dia ao local de residência, em 2018

Já para a análise de variação temporal dos padrões de mobilidade entre dos dois anos foi necessário verificar as mudanças de código do cartão entre 2014 e 2018. A partir dos dados de cadastro, observando os mesmos nomes ou identidades dos usuários, constatou-se que do total de associações de usuários entre os dois anos, somente 10% deles mudaram de código do cartão. Sendo assim, assumiu-se que associar as duas bases de bilhetagem pelo código do cartão não resultaria em grandes discrepâncias nas análises.

### 3. MÉTODO

Nesta seção, é apresentado um método (Figura 2) para identificação e caracterização dos diferentes padrões de mobilidade a partir de dados de *Smart Card* e GPS da frota. Como já mencionado na introdução, o fenômeno de interesse para análise de caracterização foi a queda

na demanda de transporte público por ônibus em Fortaleza. Este fenômeno de variação da demanda ao longo tempo motivou caracterizar os padrões de mobilidade em dois recortes no tempo, em 2014 e 2018. Os dados e o tratamento necessário para localização dos registros de validações da base da bilhetagem e dos locais de residência para os usuários do cadastro já foram discutidos na Seção 2.

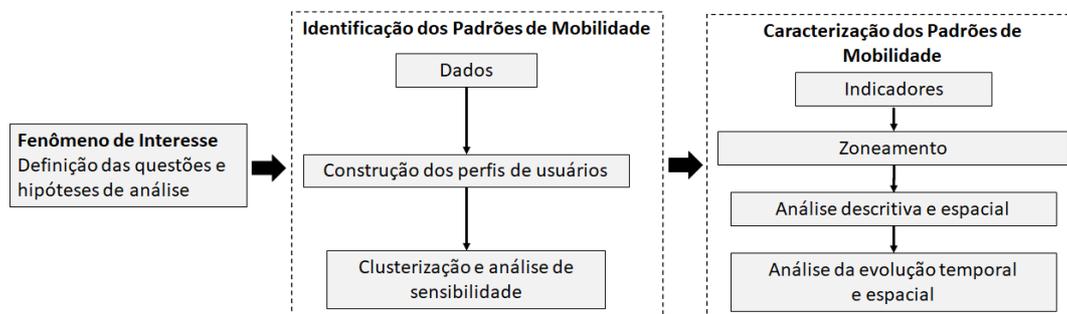


Figura 2. Etapas do método proposto

### 3.1. Construção dos perfis

Na literatura, a identificação de padrões de mobilidade a partir de dados de *Smart Card* é realizada, principalmente, a partir da criação de perfis para cada usuário. Um perfil é construído a partir das características dos usuários, que podem ser características socioeconômicas ou características das viagens. Neste estudo, foram analisados atributos temporais de utilização do sistema que, conforme visto na literatura, podem resultar em diferentes padrões de uso: padrões variando entre usos esporádicos e regulares no sistema.

Conforme os aspectos de intensidade e regularidade dos usos, assim como os dados disponíveis, definiu-se para cada usuário os atributos apresentados no Tabela 2. Para cada usuário, os indicadores de média ou desvio foram calculados para observações de uma amostra de N\_DIAS dias, correspondente aos dias úteis típicos de uso do sistema. No cálculo destes atributos para cada usuário, não foram consideradas as validações que envolvem integração tarifária, quando um indivíduo realiza transferência entre veículos de linhas diferentes sem pagar uma nova tarifa. A integração pode ocorrer tanto dentro dos terminais físicos de integração, sem gerar uma nova validação, como fora dos terminais, quando o intervalo entre validações é menor do que duas horas. As validações fora dos terminais físicos são identificadas na Base de Bilhetagem. Estes registros de integração foram excluídos da análise por estarem associados, em sua maioria, a etapas da mesma viagem.

Espera-se, portanto, que os indicadores envolvendo número de validações capturem a variação da frequência e regularidade das viagens entre os indivíduos, e os indicadores envolvendo intervalo de tempo entre a primeira e a última validação capturem a variação dos tempos alocados pelos indivíduos para realização das atividades. Vale ressaltar que os indicadores temporais (DIST\_TEMP e DIST\_TEMP\_SD) são suficientes para representar o efeito temporal da realização de atividades na regularidade e intensidade de uso do sistema.

Também foram desconsiderados da análise os usuários com frequência média acima de 6 ou com o número de dias menor do que 7 (usuários que raramente utilizam o sistema). A proporção de usuários com frequência média acima de 6 é menor do 1,0% em cada ano. Muitas dessas validações ocorrem de forma simultânea ou quase no mesmo momento, indicando inconsistências ou erros nos dados, ou ainda uso inadequado.

**Tabela 2** – Atributos para identificação dos padrões

Atributo	Definição
FREQ_MED - Frequência média diária de uso (validações/dia)	Média diária de validações sem contar integração - este atributo representa a intensidade de uso diária de cada usuário.
FREQ_MED_SD - Desvio padrão da frequência (validações/dia)	Desvio padrão da frequência diária de uso - esse atributo representa a regularidade da intensidade de uso por dia.
N_DIAS - Número de dias típicos de uso durante o ano (dias)	Quantidade de dias úteis que o usuário utiliza o sistema - este atributo representa a regularidade de uso ao longo do ano.
DIST_TEMP - Média da distância temporal entre a primeira e a última validação (horas)	Média dos intervalos de tempo entre a primeira e a última validação em cada dia - este atributo representa o tempo médio diário que o usuário utiliza para realizar uma ou mais atividades durante o dia.
DIST_TEMP_SD - Desvio padrão da distância temporal entre a primeira e a última validação (horas)	Desvio padrão da variável intervalo de tempo entre a primeira e a última validação em cada dia - este atributo representa a regularidade do tempo estimado que o usuário utiliza para realizar suas atividades durante o dia.

### 3.2. Identificação de padrões

O método de agrupamento (*Clustering*) escolhido para identificar padrões de mobilidade foi o *k-means*. Este é o método de clusterização mais simples e mais utilizado para realizar a identificação de clusters com base em atributos e apresenta um bom desempenho computacional viabilizando a sua utilização para a base de dados utilizada. O *k-means* gera diferentes agrupamentos a partir de uma função que busca minimizar a distância quadrática dos pontos (em que cada ponto é um conjunto de valores de atributos ou perfil do usuário) aos centroides dos clusters (Zaki e Meira, 2014). Esta função de distância é utilizada também como indicador para avaliar o desempenho do algoritmo em identificar grupos homogêneos. Para executar, o *k-means* requer a definição: dos atributos de cada indivíduo (conforme os perfis dos usuários), do número de grupos esperado e de uma medida ou função de similaridade. O número de grupos foi definido com base na análise de variabilidade (*Elbow Method*) dos grupos gerados (variação dentro de cada grupo). Uma análise de variabilidade dos atributos em cada grupo também foi realizada para confirmar os grupos gerados.

A medida de similaridade adotada foi a Euclidiana, em que os atributos são combinados por meio de pesos. Estes pesos foram definidos a partir de uma análise de sensibilidade, considerando a correlação entre os atributos definidos na construção dos perfis de cada usuário, já que diferentes atributos com elevado nível correlação tendem a contribuir de forma similar para geração dos grupos. Essa etapa, muitas vezes ignorada, é importante, uma vez que a alteração dos pesos pode alterar os grupos encontrados.

Como a proposta do método é comparar os padrões entre diferentes anos, foi necessário combinar os perfis dos usuários de 2014 e 2018 a fim de obter um único resultado de identificação dos padrões de mobilidade. Desta forma, assumiu-se que o número de grupos e os pesos na função de similaridade são os mesmos para 2014 e 2018, permitindo a identificação dos mesmos tipos de padrões nos dois anos.

### 3.3. Caracterização dos padrões de mobilidade

Esta etapa teve como objetivos interpretar os tipos grupos gerados, analisar a variabilidade espacial dos padrões e verificar espacialmente a relação entre os padrões gerados com a variação da demanda ao longo dos anos. Inicialmente, os grupos ou padrões gerados foram interpretados de acordo com os tipos usuários (tipo de validação) em cada ano. Para analisar a variação dos padrões entre os anos e espacialmente, definiram-se dois indicadores: (1) *proporção dos grupos*

por zona e (2) diferença de validações em cada grupo entre 2018 e 2014. O *Indicador 1* se propõe a investigar a variabilidade espacial dos diferentes padrões em cada ano. Já o *Indicador 2* foi definido para investigar a hipótese de que a demanda varia no tempo de forma distinta entre os padrões e por localidade.

Nas análises espaciais, adotou-se o zoneamento proposto por Lima (2017), que definiu as zonas de Fortaleza a partir da agregação de setores censitários com características demográficas e socioeconômicas similares para a população usuária do transporte público (Figura 3). A autora propõe uma divisão de Fortaleza em 241 zonas respeitando também as divisões por bairros, em que cinco foram descartadas por se tratar de áreas de parques e/ou corpos de água, além da zona na qual se encontra o Aeroporto Internacional Pinto Martins. A autora também divide Fortaleza em seis macrozonas (Central, Noroeste, Nordeste, Sudeste, Sudoeste e Sul).

Essas macrozonas são propostas através de uma consideração dos principais eixos viários e de aspectos sociais e econômicos dos bairros. A Região Central é composta por um grupo de zonas limitadas pelo primeiro anel viário da cidade, e compreendem o centro histórico, econômico, financeiro e administrativo da capital. As demais regiões, fazendo um contraponto a região central, serão chamadas de regiões periféricas. Sendo essa característica de periferia das zonas mais evidente à medida que estas se distanciam da região central.

Conforme a Seção 2, as zonas das residências dos usuários da Base de Bilhetagem foram identificadas a partir dos dados de cadastro. Tem-se como premissas de que a origem da primeira viagem do dia de cada usuário no transporte público ocorre no local de residência e de que a distribuição espacial das residências dos usuários da Base de Cadastro represente a distribuição espacial das residências da população de usuários do SIT-FOR. Assim, espera-se que a distribuição espacial das validações dos usuários da amostra de cadastro se assemelhe a distribuição espacial das validações de todos os usuários do sistema (incluindo validações do tipo inteira e validações com bilhete único não encontrado na Base de Cadastro), o que pode ser verificado nos mapas de quintis da Figura 3. Para os usuários de validações do tipo inteira, assumiu-se também que estes apresentavam os mesmos padrões de mobilidade observados nos grupos identificados.

Nas análises espaciais com base no *Indicador 1*, utilizou-se as seguintes ferramentas de análise espaciais: Mapas de Quintis e *Lisa Cluster Map*. Estas ferramentas permitem verificar a variação e dependência espacial dos agrupamentos, ou existência de *clusters* espaciais, evidenciando concentrações de um dado tipo de agrupamento.

Já nas análises da relação entre os padrões de mobilidade e a queda da demanda com base no *Indicador 2*, analisou-se inicialmente a dinâmica dos padrões ao longo do tempo, a partir de um diagrama de *Sankey* conforme proposto por Briand et al. (2017). Esta análise permitiu identificar como os grupos mudam ao longo do tempo, quais usuários entraram e deixaram o sistema. Espera-se com esta análise avaliar quais os padrões de mobilidade estão mais associados à queda na demanda ao longo do tempo, e levantar hipóteses sobre as possíveis causas dessa correlação. Em seguida, analisou-se a partir do *Indicador 2* a relação espacial dos padrões de mobilidade e à queda da demanda, a partir de análise de dependência espacial do tipo *Lisa Cluster Map*. Estas análises permitiram avaliar quais os padrões e quais zonas estão mais associadas à queda na demanda no período em análise, permitindo relacionar este fenômeno com características socioeconômicas da população que variam espacialmente.

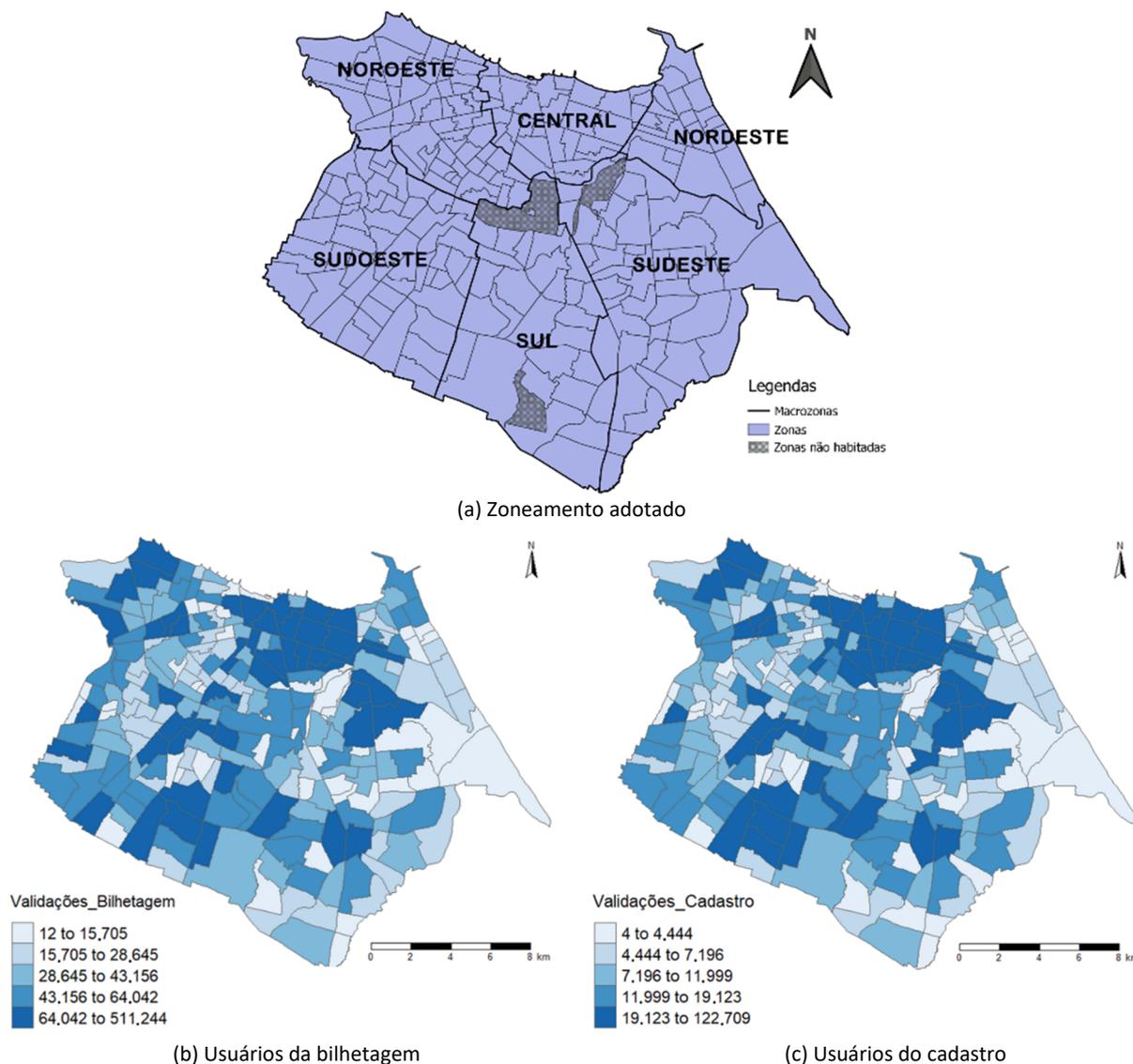


Figura 3. Zoneamento e distribuição espacial das validações dos usuários da bilhetagem e do cadastro

## 4. RESULTADOS

### 4.1. Identificação dos padrões de mobilidade

A análise da variabilidade dos atributos em cada ano evidencia uma heterogeneidade entre os usuários, em especial para o Desvio da frequência diária, a Distância temporal média e o Número de dias de uso durante o ano, que apresentaram coeficientes de variação maiores do que 0,40 nos dois anos (Tabela 3). Essa elevada variação entre os perfis dos usuários corrobora com a hipótese de que diferentes padrões de intensidade e regularidade na mobilidade são esperados para os usuários de transporte público por ônibus de Fortaleza. Além disso, ocorreu uma mudança de 78,8 para 87,0 na média do número de dias de uso entre os dois anos, indicando um aumento na regularidade de uso do sistema.

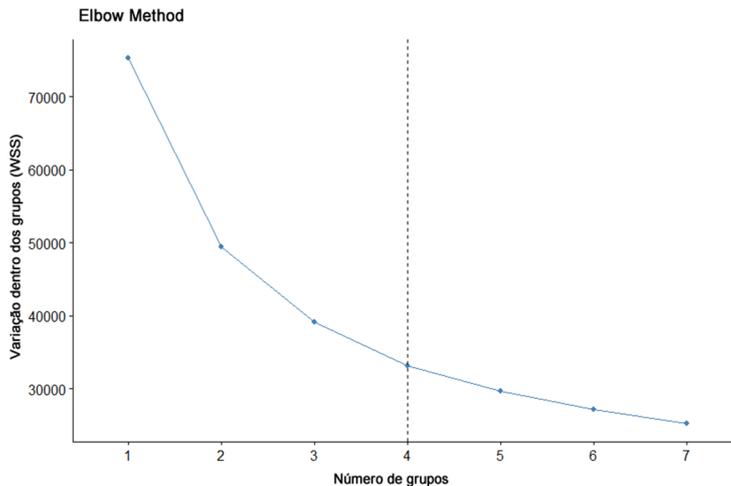
O resultado do método do cotovelo (*Elbow Method*), conforme Figura 4, indicou a existência de 4 agrupamentos nos dados. A Figura 4 apresenta a relação entre a variação dentro dos clusters e o número de clusters, quando o *k-means* é executado para N (quantidade de grupos) de 1 a 7 e para uma função de similaridade com pesos unitários. A definição da quantidade ideal de grupos é realizada de forma visual, identificando em que ponto do gráfico a queda da variação

começa a se estabilizar. Vale ressaltar que comparando os grupos para diferentes aplicações do *k-means*, em que  $N = 3, 4$  e  $5$ , os resultados mostram que um dos grupos para  $N = 3$  se subdivide formando um novo grupo para  $N = 4$ , enquanto de  $N = 4$  para  $N = 5$  ocorre um rearranjo dos grupos, mostrando que realmente parece existir 4 grupos bem definidos nos dados.

**Tabela 3** – Medidas de tendência central e variação dos atributos entre os usuários

Atributo	2014				2018			
	Média	Mediana	Desvio Padrão	Coefficiente de Variação	Média	Mediana	Desvio Padrão	Coefficiente de Variação
FREQ_MED	2,01	1,96	0,60	0,30	1,96	1,93	0,55	0,28
FREQ_MED_SD	0,79	0,75	0,32	0,40	0,74	0,69	0,31	0,41
DIST_TEMP	4,96	4,73	2,97	0,60	4,83	4,45	2,92	0,60
DIST_TEMP_SD	3,65	3,75	1,39	0,38	3,58	3,62	1,33	0,37
N_DIAS	78,8	54,0	68,7	0,87	87,0	67,0	67,9	0,78

A análise de sensibilidade dos pesos da função de similaridade foi feita variando os pesos dos atributos em torno de pesos pré-definidos com base na correlação entre os dos atributos. Assim diferentes combinações de atributos foram testadas, sendo adotada a combinação que resultou em padrões com menor variabilidade entre grupos para os dois anos de análise. A aplicação do *Elbow Method* com estes novos pesos apontou também a existência de  $N = 4$  clusters. A aplicação conjunta da definição do número de clusters e pesos da função de similaridade permite uma melhor definição da contribuição dos atributos, o que não é feito quando se normaliza os valores dos atributos assumindo pesos constantes.



**Figura 4.** *Elbow Method*

## 4.2. Caracterização dos padrões de mobilidade

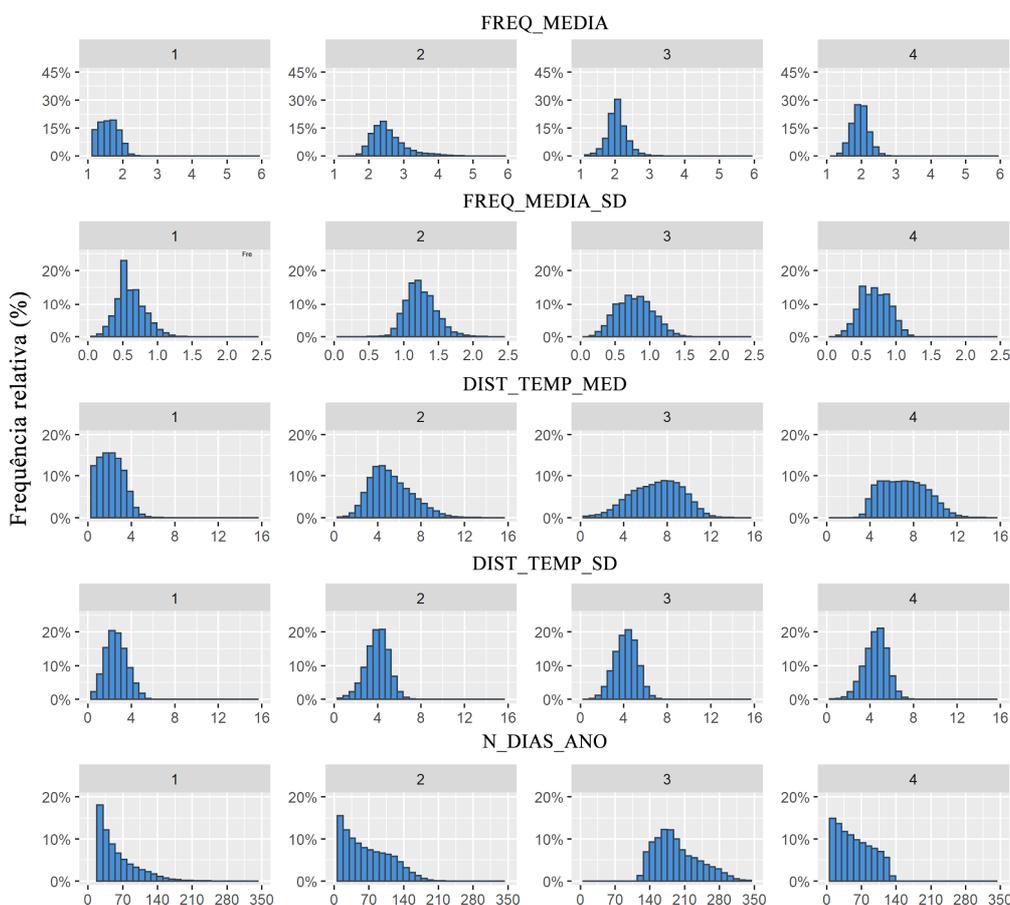
### 4.2.1. Análise descritiva e espacial dos padrões identificados

Definidos os atributos, a quantidade de grupos esperados e os pesos de cada atributo foi possível executar o *k-means* para identificar 4 padrões. A Tabela 4 apresenta composição e as estatísticas descritivas dos quatro grupos encontrados por tipo de cartão. A Figura 5 apresenta as distribuições de frequências relativas para cada atributo e padrão em 2018.

**Tabela 4** – Síntese das estatísticas descritivas por tipo de cartão para cada padrão

Padrão	Tipo de Cartão	Quantidade de cartões (2014 + 2018)	Proporção dentro do padrão	Média das frequências	Média das distâncias temporais entre validações	Média do número de dias
Padrão 1	Estudante	357.666	51,2%	1,6	2,3	41,8
	Gratuidade	123.918	17,7%	1,5	1,3	39,0
	Vale transporte	217.081	31,1%	1,5	1,8	48,6
	<b>Total</b>	<b>698.665</b>		<b>1,6</b>	<b>2,0</b>	<b>43,4</b>
Padrão 2	Estudante	49.879	16,2%	2,4	5,1	63,2
	Gratuidade	32.647	10,6%	3,3	4,2	83,3
	Vale transporte	226.089	73,3%	2,8	5,4	65,3
	<b>Total</b>	<b>308.615</b>		<b>2,8</b>	<b>5,2</b>	<b>66,9</b>
Padrão 3	Estudante	106.515	24,9%	2,2	6,2	175,3
	Gratuidade	9.232	2,2%	2,3	5,3	192,9
	Vale transporte	312.027	72,9%	2,1	7,4	186,2
	<b>Total</b>	<b>427.774</b>		<b>2,1</b>	<b>7,1</b>	<b>183,6</b>
Padrão 4	Estudante	128.294	28,5%	2,0	6,1	61,5
	Gratuidade	3.829	0,9%	1,9	5,7	67,6
	Vale transporte	318.208	70,7%	2,0	7,6	56,2
	<b>Total</b>	<b>450.331</b>		<b>2,0</b>	<b>7,2</b>	<b>57,8</b>

Os resultados da Tabela 4 e Figura 5 evidenciam a existência de um grupo regular (Padrão 3), um grupo esporádico (Padrão 1), e dois grupos não-regulars (Padrões 2 e 4) de uso do sistema. Possivelmente o Padrão 3 é composto por usuários que dependem do transporte público por ônibus para realizar seus deslocamentos diários, especialmente com propósito trabalho, já que neste grupo aproximadamente 73% dos cartões foram do tipo Vale Transporte.



**Figura 5.** Distribuição de frequências relativas dos atributos em 2018

O padrão de usuários esporádico (Padrão 1) é composto majoritariamente (51%) por usuários do tipo Estudante. Nota-se que estes usuários parecem estar usando o sistema para realizar atividades de curta duração (distância temporal entre validações com média de 2,0 horas), que ocorrem poucas vezes por semana (média do número de dias igual a 43,4 dias), e com baixa intensidade de uso do sistema (frequência de 1,6 validações por dia). Observa-se também um percentual expressivo de gratuidades neste Padrão 1, corroborando com caráter esporádico deste grupo, já que viagens do tipo de gratuidade ocorrem com menor regularidade.

O Padrão 2 é o grupo com maior intensidade (2,8 validações por dia). É interessante notar que o intervalo temporal médio entre validações foi de 5,2 horas, sendo menor do que o tempo esperado, de pelo menos 6 horas de duração, de uma atividade de trabalho formal. Portanto, já que proximamente 73% deste grupo é composto por cartões do tipo Vale Transporte, acredita-se que os usuários deste padrão, em geral, se deslocam por motivo de trabalho informal ou por outros motivos. O número de dias de uso ao longo do ano (67 dias) mostra também que esses usuários não são dependentes do sistema, ou que a atividade não ocorre todos os dias da semana.

O Padrão 4 apresenta considerável similaridade com o Padrão 3, conforme Figura 5 e Tabela 4, com exceção no número de dias de uso. Ambos os padrões apresentam uma intensidade de uso em torno de 2,0 validações por dia e um intervalo médio entre validações próximo a 7,0 horas. A composição dos dois grupos por tipo de cartão é também similar (Tabela 4). Tem-se como hipótese de que os usuários do Padrão 4 não dependem somente do sistema por ônibus, e realizam viagens para acessar atividades do tipo trabalho formal ou educação.

A Tabela 5 apresenta o número de usuários em cada grupo e o quanto estes usuários contribuem para a receita (número de validações) do sistema. Vale ressaltar que as validações do tipo Inteira (pagas à dinheiro) não foram consideradas nesta análise. Um primeiro fato que pode ser notado nesta tabela é que a maior parte dos usuários do sistema é do tipo esporádico ou não-regular (mais de 70% em 2014 e mais de 80% em 2018), e que o grupo que mais gera validações (58% em 2014 e 43% em 2018) é do tipo regular. Contudo, observa-se que entre 2014 e 2018 ocorreu uma perda de usuários, principalmente do tipo regular (Padrão 3), que podem ter perdido o emprego ou que podem estar migrando para outros modos de transporte, como o automóvel particular ou motocicleta. Esta queda no número de usuários ocorreu principalmente para os usuários do tipo Vale Transporte (redução de aproximadamente 81000 usuários no Padrão 3 e de mais de 48000 no Padrão 2), ou seja, possivelmente são usuários que realizam viagens por motivo trabalho. Nota-se também um acréscimo de viagens de usuários esporádicos e não regulares (Padrões 1, 2 e 4), especialmente por motivo outros ou educação. Ou seja, enquanto houve um crescimento de viagens/usuários que possivelmente não dependem do transporte público, ocorreu uma acentuada queda (de aproximadamente 30%) de usuários e viagens do tipo regular.

**Tabela 5** – Totais de usuários e validações em cada cluster

Padrão	Usuários (2014)	Validações (2014) (×10 <sup>6</sup> )	Usuários (2018)	Validações (2018) (×10 <sup>6</sup> )
Padrão 1	355.022 (35%)	20,2 (12%)	343.643 (39%) ↓	27,8 (18%) ↑
Padrão 2	173.710 (17%)	28,9 (17%)	134.905 (15%) ↓	29,8 (19%) ↑
Padrão 3	262.018 (26%)	97,9 (58%)	165.756 (19%) ↓	67,0 (43%) ↓
Padrão 4	221.082 (22%)	21,3 (13%)	229.249 (26%) ↑	30,9 (20%) ↑
<b>Total</b>	<b>1.011.832</b>	<b>168,3</b>	<b>873.553</b> ↓	<b>155,4</b> ↓

↑ Acréscimo, ↓ Redução

A Figura 6 apresenta as distribuições espaciais das proporções de usuários (*Indicador 1*) em cada grupo por zona para o ano de 2018. O padrão de usuários regulares (Padrão 3) e o padrão de não-regulares com atividade formal ou educação (Padrão 4) apresentam uma maior proporção de usuários nas zonas periféricas, enquanto os usuários esporádicos (Padrão 1) estão proporcionalmente mais presentes nas zonas centrais. Já o Padrão 2 de usuários não-regulares e de atividade informal apresentou um cluster *Low-Low* (menor concentração) bem definido na região sudoeste, e uma concentração *High-High* na região nordeste.

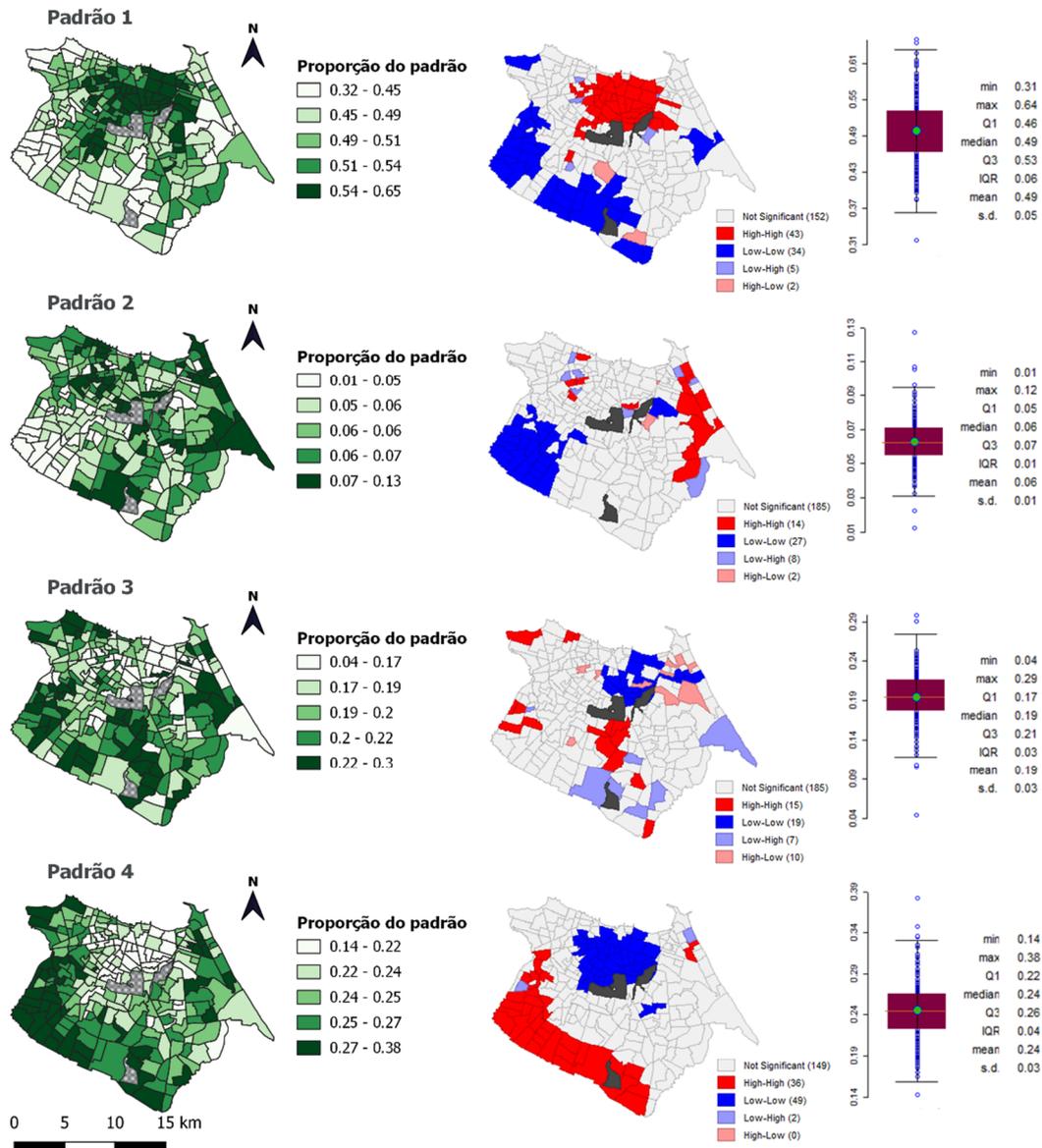


Figura 6. Mapa de quintis; Lisa Map e Boxplot para o Indicador 1 em 2018

Estes resultados reforçam, portanto, a hipótese de que os usuários regulares que usam o sistema para atividade trabalho (Padrão 3), residem proporcionalmente em zonas mais periféricas, com menor acesso às oportunidades. Além disso, evidencia que usuários que usam sistema de forma esporádica para viagens não pendulares (Padrão 1) residem proporcionalmente em zonas centrais da cidade, as quais apresentam os melhores níveis de acesso às atividades.

#### 4.2.2. Análise temporal dos padrões identificados

A Figura 7 apresenta o diagrama de *Sankey* que associa os códigos dos cartões do bilhete único em cada padrão entre os anos de 2014 e 2018. Pode-se notar no diagrama que neste período de 5 anos ocorreu um acréscimo de quase 65% novos cartões (Grupo x) em 2018, enquanto aproximadamente 70% dos cartões em 2014 deixaram de ser usados no sistema (Grupo y). Em geral, ocorreu uma mudança considerável no total de cartões em todos os padrões, sendo os padrões 1 e 3 os mais estáveis, com menor alteração ao longo do tempo. Os Padrões 2 e 4 apresentaram as maiores proporções de cartões que deixaram de ser utilizados, possivelmente usuários que não dependem somente do transporte público, e que passaram a usar outros modos de transportes.

Apesar do Padrão 1 ser responsável pela maior quantidade de cartões deixando o sistema, também é o padrão que recebeu mais cartões no ano de 2018. O padrão de usuários regulares (Padrão 3) é o que mais reduziu de tamanho entre o período analisado, ou seja, perdeu mais cartões do que ganhou. Os resultados, portanto, indicam uma dinâmica nos padrões de mobilidade ao longo dos anos, indicando que algum fator externo pode estar afetando os usuários que usam o sistema de forma regular. Vale ressaltar, conforme Seção 2, que alguns usuários podem ter mudado de código do cartão entre os anos de 2014 e 2018.

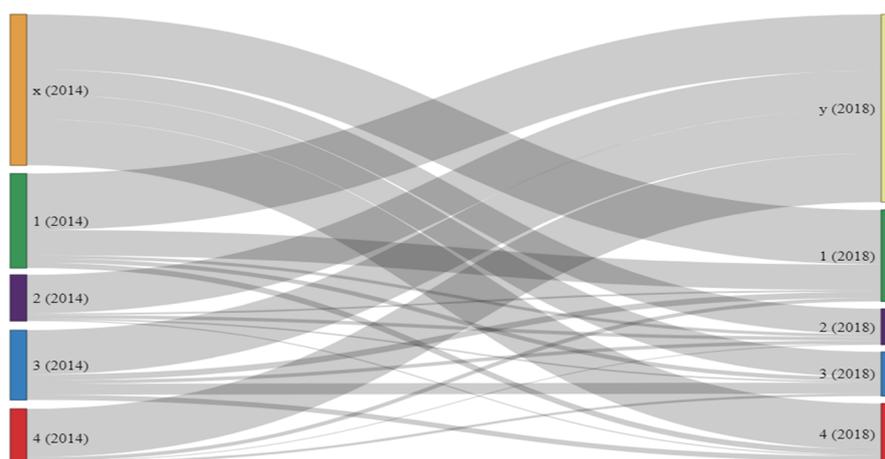


Figura 7. Diagrama de *Sankey* para a quantidade de usuários em cada padrão (2014-2018)

#### 4.2.3. Análise da variação espacial

A Figura 8 mostra a distribuição espacial do indicador de diferença de validações (*Indicador 2*) para todos os padrões. Em geral observa-se uma queda na produção de validações, principalmente nas zonas periféricas. Algumas zonas categorizadas como *High-Low* (zonas com menor redução, ou até acréscimo de validações, circundadas por zonas com maior queda na demanda) foram identificadas na região sul, bem distantes das zonas centrais. Este resultado pode estar relacionado às mudanças no uso do solo oriundas de programas habitacionais, como é o caso do Residencial Cidade Jardim II lançado em 2018 pela Prefeitura de Fortaleza e o Governo do Estado do Ceará.

Na Figura 8 é possível visualizar que o padrão dos usuários regulares (Padrão 3) é o que mais sofre redução de validações ao longo dos cinco anos de análise. As concentrações do tipo *low-low* para este padrão estão localizadas nas zonas periféricas das regiões Noroeste, Sudoeste, e Sul. Estas zonas concentram uma parcela considerável da população de baixa renda (Souza,

2019), e apresentam os mais baixos níveis de acessibilidade pelo transporte público. Por outro lado, pode-se observar um acréscimo de validações nos padrões de usuários esporádicos (Padrão 1) e não regulares (Padrão 4), em algumas zonas na região central ou adjacentes às zonas centrais ou aos corredores importantes de transporte público da cidade. Portanto, tem ocorrido uma redução expressiva de viagens de usuário regulares localizados em zonas periféricas da cidade, ao mesmo tempo em que vem ocorrendo um aumento de validações geradas por usuários não-regulares do sistema ou que utilizam o sistema de forma esporádica.

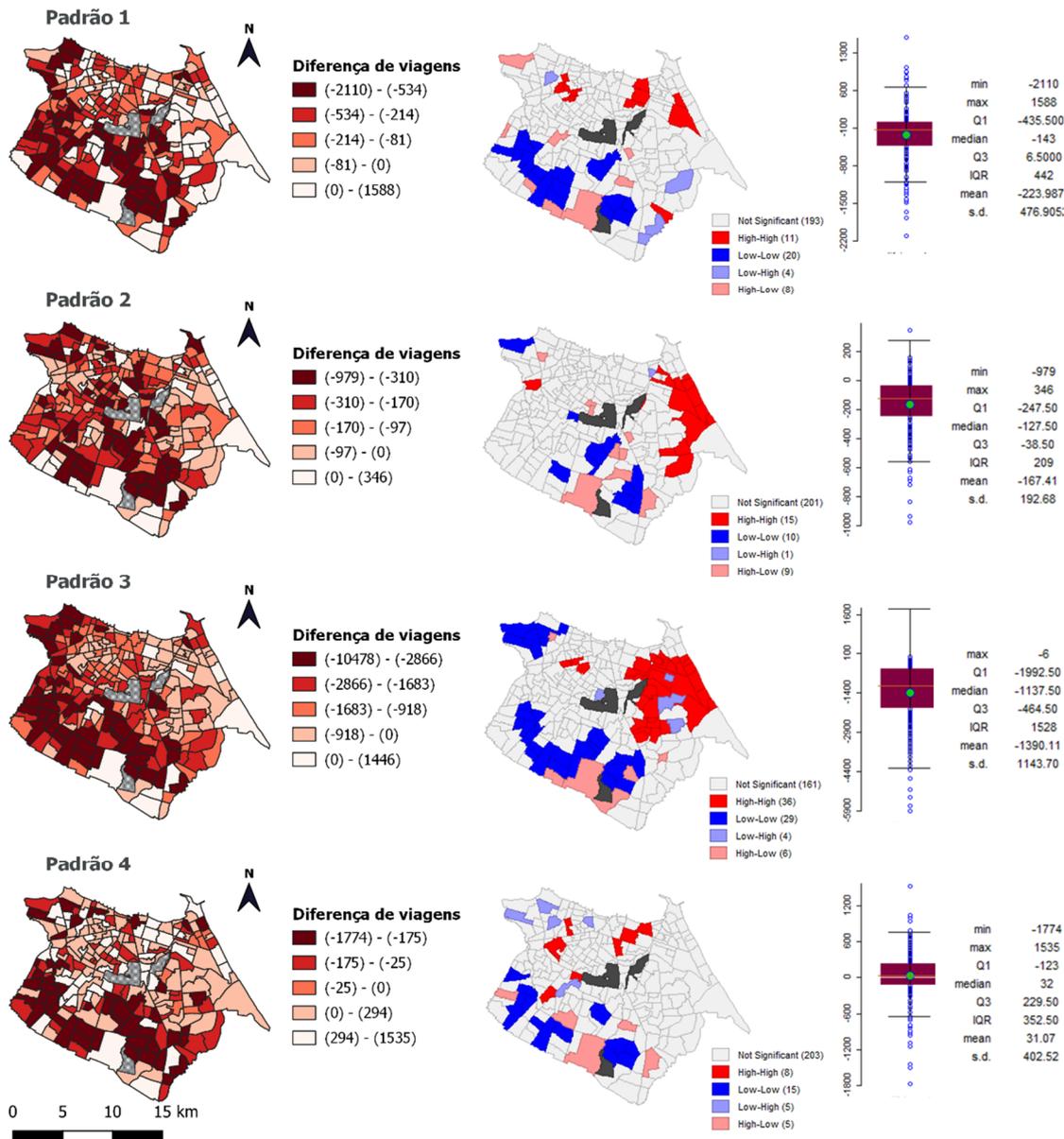


Figura 8. Mapa de quintis; Lisa Map e Boxplot para o Indicador 2

### 5. CONSIDERAÇÕES FINAIS

Este trabalho apresentou um método para caracterizar os padrões de mobilidade dos usuários de transporte público por ônibus através da utilização de dados de bilhetagem eletrônica e GPS da frota veicular. O caráter contínuo e quase populacional desses dados, diferentemente das bases de dados comumente utilizadas no planejamento de transportes, permite a geração de

atributos e indicadores que permitem identificar a heterogeneidade na mobilidade e as mudanças nos padrões de mobilidade ao longo do tempo. No método proposto, a identificação dos padrões é feita por meio de técnicas de *clusterização*, combinando atributos (espaciais ou temporais) que representem a mobilidade dos usuários e permitam analisar a variação da demanda ao longo dos anos (perfis construídos para 2 anos de dados).

Diferentemente dos estudos anteriores, no método proposto, sugere-se identificar e analisar padrões de mobilidade relacionando-os com algum fenômeno de interesse, que no presente caso foi a redução na quantidade de viagens que vem ocorrendo nos últimos anos no SIT-FOR. A queda na demanda é um problema não só para as empresas operadoras e gestores do sistema, mas também pode ser um problema para a sociedade em geral, caso os usuários estejam migrando para modos menos sustentáveis. Assim, identificar as características de mobilidade dos diferentes segmentos de usuários, que podem estar deixando de utilizar o sistema, é essencial para apoiar decisões táticas ou estratégicas que busquem atrair mais passageiros para o sistema.

O método foi aplicado para a base de bilhetagem e GPS do SIT-FOR para analisar a variação da demanda entre os anos de 2014 e 2018. Os resultados indicaram que existe uma heterogeneidade na mobilidade em Fortaleza, podendo coexistir no meio urbano 4 padrões, variando de esporádico até regulares de uso do sistema. Estes diferentes padrões de mobilidade podem estar relacionados ao propósito das viagens, ao nível de acessibilidade pelo transporte público, e ao acesso dos indivíduos a outros modos de transportes.

As análises de caracterização dos padrões revelaram três fatos principais: i) os usuários do padrão regular (que tem o menor número de usuários e que mais gera receita para o sistema) vem deixando o sistema ao longo do tempo, com uma consequente redução nas validações (em 2014 eles contribuíam com quase 60% das validações, já em 2018 este percentual caiu para 43%); ii) os usuários do padrão regular estão mais concentrados nas zonas periféricas de Fortaleza, distantes das zonas de emprego, enquanto aqueles usuários do tipo esporádico estão em regiões centrais com os melhores níveis de acessibilidade; e iii) ocorreu uma redução significativa nas validações entre 2014 e 2018 nas zonas periféricas e para os usuários do grupo regular. Estes resultados reforçam a hipótese de que a queda na demanda está relacionada principalmente a fatores que podem alterar o padrão de mobilidade regular, como a mudança de trabalho, o desemprego, e o acesso a outros modos de transportes (p. ex., motocicleta e automóvel).

Como limitações do método, pode-se citar primeiro que muitos usuários do sistema não usavam o bilhete único durante os anos de análise, dificultando a localização das suas zonas de residência. Acredita-se que essa limitação será amenizada nos próximos anos, pois as empresas que operam o sistema estão removendo gradativamente a opção de pagamento por dinheiro, fazendo com que os usuários adotem a utilização de cartões eletrônicos. Segundo, o método proposto requer uma Base de Cadastro atualizada do sistema. Contudo, o cadastro usado neste estudo não contém o registro de todos os usuários do sistema em 2014 e 2018, necessitando de um método de localização das zonas de residência dos usuários sem endereço cadastrado. Os dados também limitam a identificação e interpretação dos padrões já que muitos atributos das viagens (p. ex., distâncias e propósitos das viagens) e dos usuários (p. ex., renda e posse de veículo) não estão presentes em dados de bilhetagem eletrônica.

Por fim, foi observado uma dinâmica inesperada (uma elevada taxa de renovação dos cartões) dos cartões em cada padrão de mobilidade ao longo do tempo que pode ser melhor anali-

sada considerando os anos intermediários entre 2014 e 2018, e relacionando com fatores externos e internos do sistema de transportes que sofreram mudanças ao longo do tempo. Vale ressaltar que estes resultados podem ter sido afetados por mudanças de código do cartão ao longo tempo.

Recomenda-se para estudos futuros o desenvolvimento e aplicação de métodos de reconstrução das viagens a partir das localizações das validações. Em fortaleza, a identificação dos locais de embarque, transferência e desembarque, assim como do motivo da viagem, são tarefas desafiadoras, pois no SIT-FOR os usuários podem validar em qualquer ponto dos seus itinerários. Assim, estudos futuros devem abordar essas questões para sistemas como o de Fortaleza, permitindo assim gerar mais atributos das viagens para a análise de padrões de mobilidade.

## AGRADECIMENTOS

Os autores agradecem ao CNPq pela concessão de bolsas de produtividade e de mestrado e a Secretaria de Conservação e Serviços Públicos da Prefeitura Municipal de Fortaleza (PMF) que forneceu a base de dados.

## REFERÊNCIAS

- Agard, B., Morency, C. and Trépanier, M. (2006) 'Mining public transport user behaviour from smart card data', *IFAC Proceedings Volúmes*, v. 39, n. 3, p. 399–404. DOI: 10.3182/20060517-3-FR-2903.00211.
- Bree, S., Fuller, D. and Diab, E. (2020) 'Access to transit? Validating local transit accessibility measures using transit ridership', *Transportation Research Part A: Policy and Practice*, v. 141, p. 430–442. DOI: 10.1016/j.tra.2020.09.019.
- Briand, A. S., Côme, E., El Mahrsi, M. K. and Oukhellou, L. (2016) 'A mixture model clustering approach for temporal passenger pattern characterization in public transport', *International Journal of Data Science and Analytics*, v. 1, p. 37–50. DOI: 10.1007/s41060-015-0002-x.
- Briand, A. S., Côme, E., Trépanier, M. and Oukhellou, L. (2017) 'Analyzing year-to-year changes in public transport passenger behaviour using smart card data', *Transportation Research Part C: Emerging Technologies*, v. 79, p. 274–289. DOI: 10.1016/j.trc.2017.03.021.
- Cats, O. and Ferranti, F. (2022) 'Unravelling individual mobility temporal patterns using longitudinal smart card data', *Research in Transportation Business & Management*, v. 43. DOI: 10.1016/j.rtbm.2022.100816.
- Devillaine, F., Munizaga, M. and Trépanier, M. (2012) 'Detection of activities of public transport users by analyzing smart card data', *Transportation Research Record: Journal of the Transportation Research Board*, v. 2276, n. 1, p. 48–55. DOI: 10.3141/2276-06.
- Diab, E., DeWeese, J., Chaloux, N. and El-Geneidy, A. (2020) 'Adjusting the service? Understanding the factors affecting bus ridership over time at the route level in Montréal, Canada', *Transportation*, v. 48, p. 2765–2786. DOI: 10.1007/s11116-020-10147-3.
- Faroqi, H., Mesbah, M. and Kim, J. (2019) 'Comparing sequential with combined spatiotemporal clustering of passenger trips in the public transit network using smart card data', *Mathematical Problems in Engineering*, v. 2019, p. 1–16. DOI: 10.1155/2019/5070794.
- Faroqi, H., Mesbah, M., Kim, J. and Tavassoli, A. (2018) 'A model for measuring activity similarity between public transit passengers using smart card data', *Travel Behavior and Society*, v. 13, p. 11–25. DOI: 10.1016/j.tbs.2018.05.004.
- Garcia, C. S. H. F., Macário, M. do R. M. R., Menezes, E. D. de A. G. and Loureiro, C. F. G. (2018) 'Strategic Assessment of Lisbon's Accessibility and Mobility Problems from an Equity Perspective', *Networks and Spatial Economics*, v. 18, p. 415–439. DOI: 10.1007/s11067-018-9391-4.
- Goulet-Langlois, G., Koutsopoulos, H. N., Zhao, Z. and Zhao, J. (2018) 'Measuring Regularity of Individual Travel Patterns', *IEEE Transactions on Intelligent Transportation Systems*, v. 19, n. 5, p. 1583–1592. DOI: 10.1109/TITS.2017.2728704.
- Gomide, A. A. (2003) 'Transporte urbano e inclusão social: elementos para políticas públicas', *Texto para discussão No 960*, Instituto de pesquisa econômica aplicada, Brasília.
- Hawkins, J. and Habib, K. N. (2020) 'Heterogeneity in marginal value of urban mobility: evidence from a large-scale household travel survey in the Greater Toronto and Hamilton Area', *Transportation*, v. 47, p. 3091–3108. DOI: 10.1007/s11116-019-10041-7.
- He, L., Agard, B. and Trépanier, M. (2020) 'A classification of public transit users with smart card data based on time series distance metrics and a hierarchical clustering method', *Transportmetrica A: Transport Science*, v. 16, n. 1, p. 56–75. DOI: 10.1080/23249935.2018.1479722.
- He, L., Trépanier, M. and Agard, B. (2021) 'Space-time classification of public transit smart card users' activity locations from smart card data', *Public Transport*, v. 13, p. 579–595. DOI: 10.1007/s12469-021-00274-0.
- Kieu, L. M., Bhaskar, A. and Chung, E. (2015) 'Passenger Segmentation Using Smart Card Data', *IEEE Transactions on Intelligent Transportation Systems*, v. 16, n. 3, p. 1537–1548. DOI: 10.1109/TITS.2014.2368998.
- Lima, L. S. (2017) *Espraiamento urbano por atossegregação e seus impactos na acessibilidade urbana de Fortaleza*. Dissertação de Mestrado, Programa de Pós-Graduação em Engenharia de Transportes, Universidade Federal do Ceará.

- Ma, X., Wu, Y. J., Wang, Y., Chen, F. and Liu, J. (2013) 'Mining smart card data for transit riders' travel patterns', *Transportation Research Part C: Emerging Technologies*, v. 36, p. 1–12. DOI: 10.1016/j.trc.2013.07.010.
- Mahrsi, M. K. El, Côme, E., Baro, J. and Oukhellou, L. (2014) 'Understanding Passenger Patterns in Public Transit Through Smart Card and Socioeconomic Data', *The 3rd International Workshop on Urban Computing (UrbComp 2014)*, New York, NY, USA.
- Martens, K. (2019) 'Why accessibility measurement is not merely an option, but an absolute necessity', in Silva, C., Pinto, N. and Bertolini, L. (ed.) *Designing accessibility instruments: lessons on their usability for integrated land use and transport planning practices*. New York, NY: Routledge, p. 37–51.
- Morency, C., Trépanier, M. and Agard, B. (2006) 'Analysing the variability of transit users behaviour with smart card data', in *2006 IEEE Conference on Intelligent Transportation Systems*, Proceedings, ITSC, Toronto, ON, Canada. DOI: 10.1109/ITSC.2006.1706716.
- NTU (2018) Associação Nacional das Empresas de Transportes Urbanos, *Anuário NTU 2017-2018*, Brasília.
- Pelletier, M. P., Trépanier, M. and Morency, C. (2011) 'Smart card data use in public transit: A literature review', *Transportation Research Part C: Emerging Technologies*, v. 19, n. 4, p. 557–568. DOI: 10.1016/j.trc.2010.12.003.
- Pieroni, C. B. V. (2018) *Analysis of travel patterns from precarious settlements transit users in São Paulo through smart card data mining*. Dissertação de Mestrado, Programa de Pós-graduação em Engenharia de Transportes, Universidade de São Paulo.
- Pieroni, C., Giannotti, M., Alves, B. B. and Arbex, R. (2021) 'Big data for big issues: Revealing travel patterns of low-income population based on smart card data mining in a global south unequal city', *Journal of Transport Geography*, v. 96, p. 103203. DOI: 10.1016/j.jtrangeo.2021.103203.
- Sousa, F. F. L. de M. (2019) *Diagnóstico estratégico das desigualdades socioespaciais na acessibilidade ao trabalho em Fortaleza*. Tese de Doutorado, Programa de Pós-Graduação em Engenharia de Transportes, Universidade Federal do Ceará.
- Viallard, A., Trépanier, M. and Morency, C. (2019) 'Assessing the Evolution of Transit User Behavior from Smart Card Data', *Transportation Research Record: Journal of the Transportation Research Board*, v. 2673, n. 4, p. 184–194. DOI: 10.1177/0361198119834561.
- Wang, Y., Wang, Y. and Choudhury, C. (2020) 'Modelling heterogeneity in behavioral response to peak-avoidance policy utilizing naturalistic data of Beijing subway travelers', *Transportation Research Part F: Traffic Psychology and Behavior*, v. 73, p. 92–106. DOI: 10.1016/j.trf.2020.06.016.
- Zaki M. J. and Meira, W. JR (2014) *Data mining and analysis: fundamental concepts and algorithms*. 2<sup>nd</sup> edition. Cambridge University Press, Cambridge.
- Zhang, S., Yang, Y., Zhen, F., Lobsang, T. and Li, Z. (2021) 'Understanding the travel behaviors and activity patterns of the vulnerable population using smart card data: An activity space-based approach', *Journal of Transport Geography*, v. 90. DOI: 10.1016/j.jtrangeo.2020.102938.